

XI Workshop de Investigadores en Ciencias de la Computación



San Juan, 7 y 8 de Mayo de 2009

**Aval Académico Facultad de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de San Juan
Res. N° 646-08-CDFCFN**



**Red de Universidades Nacionales
con Carreras en Informática**



**Departamento e Instituto de Informática
Facultad de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de San Juan**

XI Workshop de Investigadores de Ciencias de la Computación.

1a ed. - San Juan: Univ. Nacional de San Juan, 2009.

CD-ROM.

ISBN 978-950-605-570-7

1. Programación de Computadoras.

CDD 005.1



Índice

Arquitectura, Redes y Sistemas Operativos	
Desarrollo de un administrador de memoria segmentada para un sistema operativo didáctico	1
Detección de Intrusiones mediante el uso de Redes Neuronales	6
UTN VoIP Test Bed (Voice over Internet Protocol)	12
Análisis experimental del comportamiento del protocolo 802.11e en enlaces troncales.	17
Clúster Modular Autocontenido	21
Dispositivo de almacenamiento estándar para solución embebida	24
Análisis de compresión de un sistema de archivos sobre memoria Flash usando OLPC	29
Evaluación y simulación del rendimiento de los protocolos de ruteo para MANET bajo restricciones de QoS	35
Seguridad y Privacidad en la Plataforma Android	37
Conversor Digital-Resolver para digitalización de Servosistemas Inductivos	42
Agentes y Sistemas Inteligentes	
Aplicación de técnicas de inteligencia artificial para la Navegación de robots móviles utilizando el simulador Player/stage	46
Procesamiento de Ontologías Inconsistentes en la Web Semántica Usando Argumentación Rebatible: Reporte de Progreso	50
Proyecto NNGen: Aplicación de controladores neuronales a la navegación autónoma	55
Algoritmos Metaheurísticos para Optimización y Aplicación a Problemas NP Completos	60
Nuevos desarrollos para sistemas adaptativos inteligentes	65
Desarrollo de Agentes Inteligentes Basados en Argumentación para Robótica Móvil	70
Tópicos avanzados en categorización de textos	75
Integrando capacidades de alto nivel en la toma de decisiones de agentes inteligentes	80
Validación de firmas ológrafas	85
Manejo de Restricciones y Mejora en los Procesos de Búsqueda en Algoritmos Evolutivos aplicados a Problemas Dinámicos y Scheduling	90
Programación Orientada a Agentes en el Marco de Lenguajes Multiparadigmas	94
Metaheurísticas poblacionales aplicadas a la resolución de problemas complejos	98
Sistemas Inteligentes. Aplicaciones en Minería de Datos, Procesamiento de Señales y Redes de Computadoras	103
Generador de motores de Scheduling en dominios industriales	108
Extensión de algoritmos ACO usando conceptos de Tabu Search	113
Modelos Formales de Diálogo para Sistemas Multi-Agente	118
Dinámica de bases de conocimiento compartidas y distribuidas	123
Propuesta para optimización de sistemas productivos modelados con simulación por eventos discretos	128
Una plataforma de Servicios Web	133
Técnicas de minería de datos y metaheurísticas aplicadas a la explotación eficiente de energía Eólica	136
Propuesta para resolver el problema de predicción de complejos de proteínas usando agentes inteligentes que aprenden usando técnicas de sistemas inmunológicos artificiales	141
Inteligencia Artificial en Tiempo Real: Razonamiento Rebatible en Juegos Estratégicos Digitales	146
Seguridad y gestión del ambiente basado en un Sistema Multi-agentes	150
Complejidad de la Programación en Lógica Rebatible	153
Aprendizaje de Estructuras de Independencia de Modelos Probabilísticos Gráficos	157
Optimización de funciones mono-objetivo con y sin restricciones, y funciones Multi-objetivo a través de heurísticas Bio-inspiradas	163
Computación Gráfica, Imágenes y Visualización	
Optimización de funciones mono-objetivo con y sin restricciones, y funciones Multi-objetivo a través de heurísticas Bio-inspiradas.	168

Modelización de fenómenos físicos en aplicaciones de Computación gráfica en tiempo real	172
Realidad Aumentada en un entorno gráfico de alta performance	177
Sistema automático no destructivo para la clasificación de la calidad de frutas basado en visión por computadora	182
Medición de diámetro pupilar ocular	187
Interfaces No Convencionales. Su impacto en las interacciones	192
Recuperando el detalle y evaluando la calidad visual en GPU de modelos topográficos simplificados	197
Tratamiento de imágenes digitales y video. Visión y Reconstrucción 3D.	202
Reconocimiento Estadístico de Patrones Máquinas de Soporte Vectorial y Series Temporales	207
Extracción de Información a partir de Datos no estructurados no textuales	211
Ingeniería de Software y Bases de Datos	
SFP Tool: una Herramienta para Medir Puntos Función	216
Tics aplicadas al registro de transacciones originadas en puestos de trabajo móviles	218
Ingeniería Inversa Aplicada a Software Numérico: Modelos Climáticos	223
Mejora en los procesos de identificación de Servicios, obtención de requisitos, integración de Información y desarrollo de software	228
Hacia un modelo de trazas desacoplado de los modelos de requisitos	232
Procedimientos de la explotación de información para la identificación de datos faltantes, con ruido e Inconsistentes	236
Servicios: exploraciones en soa y web.	239
Estereotipos de clases entidad y clusteriG de objetos de sistemas de gestion	244
Transformación de Procesos de Desarrollo de Software Tipo SPEM a Procesos Workflow.	249
Una Propuesta de Caso de Estudio: SmallRUP	
Un Cálculo de Patrones Arquitectónicos	254
Transformaciones Genéricas para la Implementación de Web Services en Diferentes Plataformas	259
Un marco para la Medición de Modelos Workflow	264
Calidad en los Modelos de Procesos de Negocio	269
Entornos para usar BPM en aplicaciones JAVA: un análisis comparativo	273
Tríptico de la ingeniería del software. Análisis Desarrollo y aplicaciones de metodología raise	278
Imputación de datos con redes neuronales	281
Diseño de un Profile para el Modelado de Aplicaciones Paralelas y Concurrentes	286
Relevamiento de los Pilares de Gobernabilidad Electrónica en Sitios Web Municipales	291
Construcción de una ontología para gramáticas formales y máquinas abstractas utilizando Protégé para la elicitación de requerimientos	296
Evaluación de Atributos de Sitios de Gobierno Electrónico	301
Storyboard basados en escenarios futuros	306
Desarrollo de un Framework para la Gestión de la Calidad en Ingeniería Reversa	310
Desarrollo de un Framework para la Gestión de la Calidad en Ingeniería Reversa	314
Comprensión de Programas	318
Proceso de agregación para estudios experimentales de informática	323
Metodología y herramientas de costeo en ingeniería del conocimiento	327
Integración de sistemas basados en conocimiento y de descubrimiento. Aplicación a la toma de decisiones estratégicas operacionales	330
Clasificación automática para la prevención del estrés de los suelos y la fatiga de soja en el noroeste argentino	333
Clasificación de Componentes SIG	336
Las Ubicuas Dependencias de Comparación de Conjuntos de Valores: Conceptualización, Análisis y Modelado	341
Especificación Formal en RSL de una Infraestructura abierta y estándar de Servicios Web para Sistemas de Información Geográfica	347
Herramienta para consultas complejas orientada a usuarios finales	350

Geometría Computacional y Bases de Datos Espacio-Temporales	356
Análisis e Indexación de Datos no Convencionales	361
Índices para Bases de Datos de Texto	366
Selección de Centros para Índices en Espacios Métricos	371
Identificación de requerimientos de software en las pequeña y Mediana empresa (pymes) de la provincia de Córdoba	376
Modelos y métricas para evaluar calidad de Software	382
Calidad total en un modelo integrado de proceso software y ciclo de vida	388
Modelado de aplicaciones sensibles al contexto	393
Bases de Datos Métrico-Temporales	398
Participación de los usuarios en el proceso de desarrollo de software	403
Desarrollo de herramientas de evaluación para los procesos de gestión de PYMEs de software	407
Aplicativo de soporte para la valoración de modelos y estándares de calidad para el proceso de software.	409
TICs. Aplicaciones en E-Government y Procesos Productivos	414
Calidad en el desarrollo de Sistemas de Software	419
Sistemas de Software Distribuido	423
Transformando datos en conocimiento: nuevas Prácticas	429
Enseñanza de Ingeniería de Software en un Escenario Distribuido y Colaborativo	434
Sistema de información geográfica aplicado a Turismo y patrimonio histórico y cultural	438
Modelo de casos de uso - un eje para el proceso de Desarrollo de software	442
Prometeo: Una Herramienta Para El Aprovechamiento De Metadatos De Base De Datos Relacionales	448
Gestión de los Sistemas de Información (GSI)	453
Proyecto de desarrollo de un sistema de control de flotas de vehículos	457
Estructuras de Datos Métricas para la Recuperación de Información Multimedia en la Web.	462
Metodología para especificación de requisitos en proyectos de explotación de información	467
Métodos de educación en ingeniería de sistemas Inteligentes	470
E-citizen e e-learning. Utilizando las tics para reducir la brecha digital y mejorar los servicios al Ciudadano.	472
Procesamiento Distribuido y Paralelo	
Gestión de la degradación de las comunicaciones en sistemas de vídeo bajo demanda a gran escala en Internet	476
Programación Paralela en Sistemas Híbridos	481
Diseño y clasificación de los Sistemas de Descubrimiento de recursos Grid	486
Aplicaciones del cómputo de altas prestaciones	491
Aplicaciones de Búsquedas por Similitud en un Portal GRID Orientado a Aspectos	496
Sistemas Paralelos sobre Arquitecturas Distribuidas. Cluster / Multicluster / Grid / Cloud Computing.	501
Procesamiento distribuido y paralelo. Fundamentos y aplicaciones.	506
Metodología de Sincronización de Relojes para Instrumentación	512
Sistemas de cómputo de altas prestaciones con alta disponibilidad: evaluación de la performance en diferentes configuraciones	517
Uso de threads para la ejecución en paralelo sobre una malla computacional	521
Seguridad en Entornos Virtuales	526
Nuevas Tecnologías en desarrollo de Sistemas de Software	
AP-SIG: un SIG con funciones específicas para Agricultura de Precisión	531
Un modelo de validación automático para la definición y mantenimiento de procesos de desarrollo de software	536
Verificación estática de confidencialidad en un sistema de múltiples niveles de seguridad basado en Java bytecode	541
Desarrollo y Evaluación de Ontologías en Áreas de la Informática Teórica y Aplicada	546

Consumo de Web Service desde dispositivos móviles heterogéneos	551
Integración de Bases de Datos Heterogéneas para Servicios Web en Gobierno Electrónico	556
Computación Evolutiva y Aprendizaje Automático para la Inferencia, Modelado y Simulación de Redes Regulatorias de Genes	563
Análisis para el Desarrollo Multiparadigma	568
Aprendizaje y compartición de conocimientos entre sistemas inteligentes autónomos	573
Tecnología Informática Aplicada en Educación	
Gestión del conocimiento aplicada al peritaje informático	577
Gestión del conocimiento aplicado a la educación profesional continua en gestión de proyectos informáticos	580
Uso del lenguaje unificado en la modelización de sistemas operativos	584
Blended-learning como herramienta para disminuir la deserción de alumnos	589
Personalización en sistemas de enseñanza virtual	594
Sistemas de gestión de aprendizaje de código abierto	598
Capacitando Comunidades Marginales a través de un Medio Masivo de Comunicación	603
Simulación y métodos numéricos en ciencias de la Computación	608
Recuperación personalizada de recursos educativos	613
Definición de los principales actores y sus roles para los cursos de EaD en la Facultad de Ciencias Económicas de la UNPSJB	618
Repositorio de Objetos de Aprendizaje	622
Caracterización de problemas de aprendizaje basada en explotación de información	627
Herramientas didácticas para el dictado de informática teórica	630
Modelos predictivos y técnicas de minería de datos para la identificación de factores asociados al rendimiento académico de alumnos universitarios	635
Sitio web educativo con enlaces conceptuales y aplicativos de diversas materias de las carreras de informática	640
Línea de investigación: Las TICs y la didáctica en la enseñanza de la química en cursos universitarios iniciales	645
Línea de investigación: Sistemas Tutores Inteligentes orientados a la enseñanza para la comprensión	650
Entornos para el aprendizaje virtual: producción de contenido estándar para carreras informáticas y su gestión mediante repositorios de objetos de aprendizaje	655
"E-LEARNING"	660
TICs y Estrategias de Formación de Recursos Humanos en Informática	665
Software educativo para el tratamiento de algunos temas de cálculo numérico: primera etapa de evaluación	670
Diagnostico adaptativo del estudiante en sistemas tutoriales inteligentes	674
La telepresencia, la teleoperación y la generación de competencias en el Marco de Sistemas CIM (Computer Integrated Manufacturing).	680
Calidad y Semántica de Objetos de Aprendizaje	684
Desarrollo de competencias para la gestión de información y construcción de conocimientos: TICS y Nuevos Ambientes Educativos	689
Líneas de investigación Programa Permanente de Educación a Distancia de la Facultad de Ciencias Exactas, Físicas y Naturales de la UNSJ	695
Avances en el Proyecto Software para Aprendizaje y Trabajo Colaborativo	701
Estudio de mundos virtuales para mejorar el Aprendizaje	705
Calidad en el Diseño y Producción de Materiales de Educación a Distancia	710
Estrategia para una gestión de calidad de cátedras Universitarias.	715
Intervención didáctica para educación a distancia: Reconfiguración de Medios educativos según trama disciplinar.	721
Educación Basada en la Web	727
Contenidos Didácticos y SCORM para E-learning en carreras de grado	732
Digitalización de Contenidos Accesibles por la WEB - DICA	739
Software Educativo para el aprendizaje de Clasificación de Sistemas de Información	745

Desarrollo de un administrador de memoria segmentada para un sistema operativo didáctico

Nicanor Casas Graciela De Luca Martín Cortina Hugo Ryckeboer
ncasas@unlam.edu.ar gdeluca@unlam.edu.ar mcortina@unlam.edu.ar h_ryckeboer@yahoo.com.ar

Universidad Nacional de la Matanza

Departamento de Ingeniería e Investigaciones Tecnológicas

Dirección: *Florencio Varela 1703* - Código Postal: 1754 - Teléfono: 4480-8900/8835

CONTEXTO

Uno de los objetivos del desarrollo del sistema operativo SODIUM es permitir a los alumnos de las materias de sistemas operativos observar y sacar conclusiones del funcionamiento de los distintos algoritmos que componen un sistema operativo, el tratamiento de los eventos que ocurren, la forma de atención de estos, que injerencia tienen en los tiempos de ejecución de los procesos, en la eficiencia de la administración y la gestión de los recursos. Este administrador de memoria segmentada nos permite continuar con la línea de administradores de memoria, mostrando la evolución de la administración en cuanto al reparto, mejor aprovechamiento de la memoria en áreas compartidas por varios procesos, disminución de la fragmentación externa y aumento de la complejidad del manejo del hardware disponible como forma de asignación del recurso. Todo esto debe realizarse con el mismo conjunto de programas fuente de prueba, compilarlos y ejecutarlos en las mismas condiciones para luego realizar las comparaciones de los diferentes sistemas de administración que posee el SODIUM.

Nuestro Sistema Operativo utiliza actualmente el formato "binario plano" obtenido desde un archivo objeto ELF por lo tanto los procesos de usuario que ejecutarán y serán ubicados en memoria utilizarán formato

estandarizado ELF para sus archivos ejecutables, por lo que necesitamos adaptarlos para la segmentación utilizando las secciones y segmentos que este provee.

RESUMEN

La finalidad de crear un administrador de memoria segmentada sin ningún tipo de solapamiento en el sistema operativo SODIUM, es la de generar un administrador clásico para que los alumnos de Sistemas Operativos puedan comprobar el estado de asignación de la memoria, los distintos segmentos en los que queda dividido el proceso, su localización en memoria, la forma en que se traducen las direcciones, como se realiza la asignación dinámica de memoria para el proceso, la asignación de memoria compartida y los mecanismos de protección de las distintas áreas de memoria comparando la eficiencia en la asignación del recurso con la de otros administradores. Se presentaron dificultades que son generadas por la disponibilidad de las diferentes arquitecturas y los compiladores existentes en el mercado, ya que la mayoría trabaja por defecto con paginación o segmentación-paginada. Se tomó un formato estandarizado de archivos ejecutables, como es el ELF y se estudió la adaptación a las necesidades del administrador de memoria, estableciendo relaciones entre las diferentes direcciones ya sean lógicas, físicas o lineales y las distintas secciones o segmentos.

Palabras Claves: Segmentación, paginación, ELF, asignación, reasignación, GDT, LDT, memoria compartida, dirección virtual, dirección física, dirección lineal.

1.- INTRODUCCIÓN

Nos hemos planteado la necesidad de que un sistema operativo didáctico conste de más de un administrador de memoria, para poder estudiarlos y realizar comparaciones de comportamiento de cada uno. Habiendo realizado ya administradores de asignación contigua simple, particiones fijas y particiones variables, la segmentación [TAN03] resulta de la evolución normal de las particiones variables agregando eficiencia en la asignación del recurso, complejidad en la administración y generación de programas para trabajar con memoria segmentada.

Esto nos permite el apoyo a una materia específica, a través de la visualización de las secciones de los procesos [TISELF], cuales de ellas están en el archivo ejecutable en el disco y cuales son creadas por el administrador de memoria, las tablas que utiliza el sistema para la traducción de direcciones [INTEL] y para la asignación de memoria del sistema. La entrega de información sobre los eventos que ocurren dentro del sistema, la localización de los distintos segmentos de cada proceso en memoria, aumentado o disminuyendo su tamaño para mejorar la fragmentación externa y la posibilidad de cambiar la configuración para realizar comparaciones. Este administrador está desarrollado por los alumnos, lo cual los lleva a un conocimiento de:

- las particularidades de los compiladores [VHA06]
- los formatos de los ejecutables que el administrador de memoria debe conocer para utilizar
- la arquitectura en la cual implementamos nuestro administrador y el entorno que la contiene.

También se realiza una transferencia de conocimientos, a través de interfaces para cada funcionalidad, lo que permite que el alumno conciba y escriba módulos alternativos.

Existe una diferencia fundamental entre los sistemas operativos de uso general (tales como Linux o Windows) y el SODIUM ya que este trata intercambiar los distintos administradores de memoria para poder estudiar y sacar conclusiones respecto a la asignación del recurso, traducción de direcciones, protección, siempre respetando los algoritmos de administración de memoria publicados en los libros más afamados de la materia Sistemas Operativos. La diferencia es debida a que los sistemas operativos de uso general trabajan con un único sistema de administración de memoria [CAR01] [RUS00], actualmente de paginación por demanda, no permitiendo cambios de configuración ni una modificación sencilla para poder utilizar dos algoritmos o más, manteniendo un conjunto de procesos de prueba que permita realizar las estadísticas y comparaciones pertinentes.

2. LINEAS DE INVESTIGACION y DESARROLLO

Si bien la arquitectura Intel x86 provee estructuras de datos que pueden ser utilizadas para la administración de memoria con segmentación tales como la GDT (Global Descriptor Table) o LDT (Local Descriptor Table) y un conjunto de registros tales como CS, DS, ES, FS; GS y SS, donde CS es un selector de segmento de código, SS un selector de segmento de stack y los restantes son selectores de segmentos de datos, para la segmentación no son suficientes ya que vamos a utilizar varias áreas de código ubicadas en segmentos diferentes debido a que necesitamos por lo menos un área de memoria o segmento para cada una de las bibliotecas dinámicas que se necesiten, para luego poder comparar el estado de asignación de memoria utilizando en la compilación de los procesos bibliotecas estáticas ó dinámicas. También re-

sultarán insuficientes los registros de datos ya que serán necesarios un registro de datos para las variables globales (inicializadas o no -BSS-), un registro para el heap o memoria de asignación dinámica del proceso y además si trabajamos con threads, para cada uno que se ejecute será necesaria un área de datos del thread , no así un área de código del thread ya que este se comparte con el código del proceso. Estos inconvenientes que presenta la arquitectura se solucionan con la tabla de relocación y los registros de relocación del ejecutable ELF. Fig 1“Asignación de memoria del proceso”

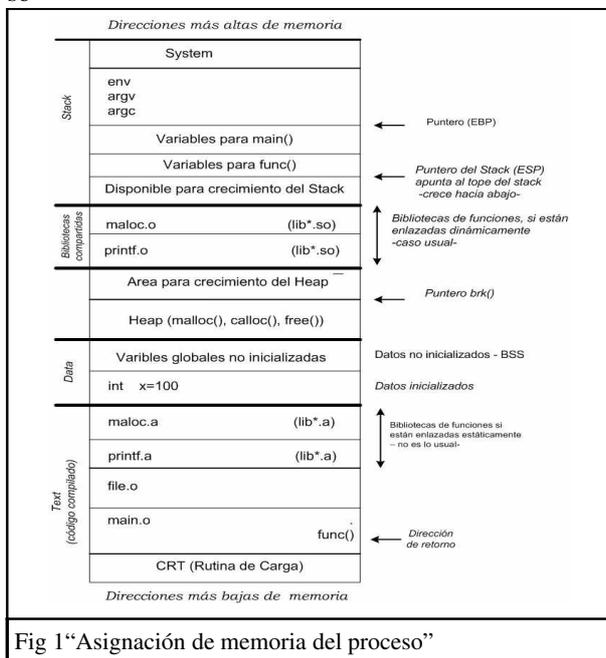


Fig 1“Asignación de memoria del proceso”

En segundo lugar tenemos también, la traducción de las direcciones virtuales Fig 2 “ Direcciones Virtuales a Direcciones físicas”, que se realizan utilizando una LDT para todos los procesos y en el futuro podrán ser una LDT por proceso. Esto se decidió así ya que la LDT posee una cantidad de entradas suficientes para la cantidad de procesos que podemos generar para estudio. El uso de una LDT por proceso tiene la ventaja que podemos identificar fácilmente todos las entradas del proceso en la tabla y los datos de cada segmento pero no agrega mejoras sustanciales a los objetivos del SODIUM .

Tercero, la asignación y recuperación de memoria se realizará mediante dos métodos dife-

rentes para poder realizar comparaciones. El primero mediante una lista doblemente enlazada de huecos libres y ocupados permitiendo utilizar los algoritmos First Fit, Best Fit, Worst Fit y Next Fit que ya están incorporados al SODIUM para las particiones variables y el segundo con el algoritmo Buddy System, con modificaciones para tamaños de memoria no potencias de dos.

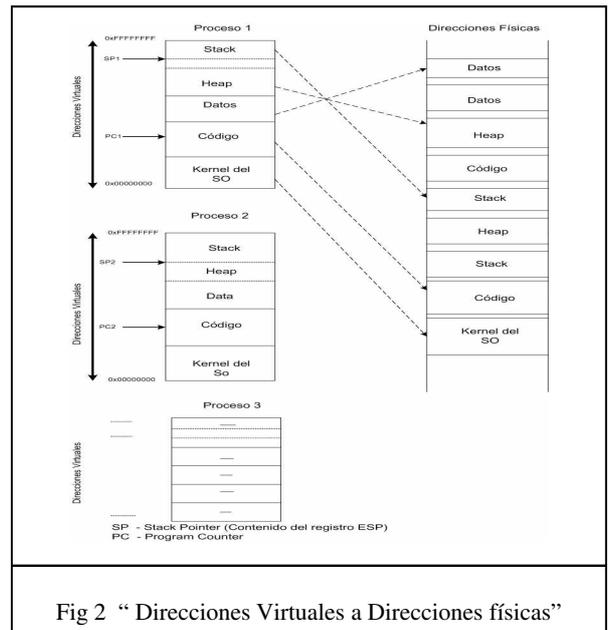


Fig 2 “ Direcciones Virtuales a Direcciones físicas”

3.RESULTADOS OBTENIDOS/ ESPERADOS

Partiendo de los sistemas de administración ya existentes en el SODIUM, se ha utilizado la misma estructura cambiando los parámetros de lista enlazada para asignar la cantidad de huecos de memoria necesaria para la segmentación, se estudió la forma de generar distintos segmentos para el proceso tal como se ve en los libros de Sistemas Operativos que permiten al programador disminuir el tamaño de las funciones de los procesos para aprovechar los huecos pequeños que se encuentran libres para su uso (con las formas anteriores de asignación que ya habíamos incorporado al SODIUM teníamos únicamente un segmento de datos que contenía los datos y el stack, y otro segmento de código o un único segmento que contenía el código los datos y el stack).

Para poder ver las distintas características de los administradores, tales como el estado de la memoria libre u ocupada, las estadísticas de asignación de memoria a los procesos y las estadísticas del sistema se desarrolló un entorno gráfico.

Se está trabajando actualmente sobre el header de programa del ejecutable ELF y sobre el header de sección para generar en la compilación archivos ejecutables compatibles con el modelo segmentado.

Queda aún por generar:

- bibliotecas dinámicas para C en SODIUM para poder generar una mayor cantidad de segmentos
- áreas de memoria compartida para la comunicación de procesos
- administradores para las áreas del heap.
- Crear varios threads por proceso, con sus áreas de datos propias.

4. FORMACION DE RECURSOS HUMANOS

Se realizó:

- la primera transferencia de los conocimientos obtenidos a los alumnos que cursan Sistemas Operativos, ya que realizaron el análisis de la arquitectura y las distintos formatos de ejecutables conjuntamente con el análisis del SODIUM e intervinieron en el desarrollo de los administradores.
- Transferencia de conocimientos a los alumnos de Sistemas de Computación II de la Universidad de La Matanza y a los alumnos de Sistemas Operativos de la Universidad Tecnológica Nacional, Regional Buenos Aires.
- Publicación de los avances en la investigación en dos congresos internacionales.

- Se prevé continuar con las publicaciones en otros congresos internacionales

Se está estudiando:

- realizar convenios de colaboración con otras universidades nacionales estatales y privadas de las cuales recibimos ofrecimientos de colaboración, con el objetivo de intercambiar conocimientos y ampliar los alcances del sistema.

En esta línea de investigación tenemos:

- dos trabajos de la Maestría en informática en curso.
- colaboran también en el desarrollo un becario, un alumno de Ingeniería electrónica y un alumno que cursa los primeros años de la carrera de Ingeniería en informática.
- Recibimos la colaboración además de alumnos que ya han cursado la materia y han desarrollado parte del SODIUM en los años anteriores.

5. BIBLIOGRAFÍA

Publicaciones

- [ANG98] Angulo José M. y Funke Enrique – Microprocesadores avanzados 386 y 486 – Introducción al Pentium y Pentium – Pro Editorial Paraninfo – Cuarta Edición
- [BRE00] BRE00- Brey Barry B. – Los Microprocesadores Intel – Editorial Prentice Hall – Quinta Edición.
- [CAR01] Card Rémy, Dumas Eric, Mével Franck - Programación Linux 2.0 API del sistema y funcionamiento del núcleo – Enrolles y Ediciones Gestión 2000 S.A.
- [DEI90] Deitel Harvey M. – Introducción a los Sistemas Operativos - Addison-Wesley Iberoamericana – Segunda Edición

- [INTEL] Manual de microprocesadores 386 y 486 y Pentium.
- [KNUT] Donald E. Knuth. Fundamental Algorithms, The Art of Computer Programming Vol. 1, Addison Wesley, Second Edition,
- [MCD07] MacDougall, Mauro, Solaris Internals (Solaris 10 and opensolaris kernel architecture) –Editorial Prentice Hall – Second Edition
- [MIL94] Milenkovic Milan – Sistemas Operativos Conceptos y diseño – Mc Graw Hill – Segunda edición
- [SIL97] Silverschatz, Avi; Galvin, Peter – Operating System Concepts – Addison-Wesley Longman – Fifth Edition
- [SMC00] Standard Microsystems Corporation – Application note 6.12
- [STA98] Stallings Willams – Operating Systems Internals and design principles – Prentice Hall International – Third Edition
- [TAN97] Tanenbaum Andrew S., Woodhull Albert S. – Operating Systems Design and Implementation – Prentice Hall – Second Edition
- [TAN03] Tanenbaum Andrew S.– Sistemas Operativos Modernos – Pearson Education – Segunda Edición
- [TISELF] Executable and Linking Format (ELF) Specification – Tool Interface Standard (TIS) Committee
- [TUR03] Turley James L. – Advanced 80386 programming techniques – Osborne McGraw Hill
- [VHA06] William Von Hagen – The Definitive Guide to GCC – Apress – Segunda Edición
- [RUS00] Russinovich, Salomon – WINDOWS INTERNALS Windows Server 2003, Windows XP, Windows 2000 – Microsoft Press

Revistas , Publicaciones e Internet

- [01] A Study of Replacement Algorithms for Virtual Storage Computers (1966) by L A Belady IBM Systems Journal
<http://www.research.ibm.com/journal/sj/052/belady.pdf>
- [02] Evaluating models of memory allocation (1994) by Benjamin Zorn, Dirk Grunwald -ACM- Transactions on Modeling and Computer Simulation
<ftp://ftp.cs.colorado.edu/pub/cs/techreports/zorn/CU-CS-603-92.ps.Z>
- [03] Microsoft “Memory allocation for long-running server applications” (1998) Per-Åke Larson and Murali Krishnan
ftp://ftp.research.microsoft.com/users/palarson/ismm98_lk.ps
- [04] The compiler, assembler, linker, loader and process address space programming tutorial - hacking the process of building programs using C language -notes and illustration.
www.tenouk.com/ModuleY.html
- [05] Jasmin Blanchette, Mark Summerfield, C++ GUI Programming with Qt 3
- [06]The Memory Fragmentation Problem: Solved? Mark S. Johnstone,Paul R. Wilson The University of Texas at Austin
<http://grothoff.org/christian/teaching/2007/4705/ismm98.pdf>
- [07] BAYS, C.: "A Comparison of Next-Fit, First-Fit, and Best-Fit", Commun. of the ACM, vol. 20, pp.# 191-192, marzo de 1977

Detección de Intrusiones mediante el uso de Redes Neuronales

Britos J. Daniel¹², Arias. Silvia¹³, Vargas Laura¹⁴

1. Facultad de Ciencias Exactas, Físicas y Naturales. Universidad Nacional de Córdoba. Laboratorio de Redes y Comunicaciones (LARyC).
2. Departamento Universitario de Informática. Universidad Nacional de Córdoba.
3. Departamento Tecnologías de Información. Instituto Universitario Aeronáutico.
4. Facultad Regional Córdoba. Universidad Tecnológica Nacional.

dbritos@gmail.com, edith.edit@gmail.com, lauramonicavargas@gmail.com

Resumen:

Con el crecimiento explosivo de Internet y particularmente de las aplicaciones de comercio electrónico, los ataques a las redes se han vuelto más comunes y sofisticados. Las redes demandan medidas de protección más elaboradas para garantizar que éstas operen seguras y dar continuidad a los servicios críticos, estas medidas incluyen métodos de detección y repuesta en tiempo real a los intentos de intrusión. Este trabajo usa modelos estadísticos y clasificadores multivariantes para detectar perfiles de tráfico anómalos, utilizando redes neuronales. El análisis estadístico basa su cálculo en el álgebra de las funciones de densidad de probabilidad (PDF). La red neuronal integra esta información en una sola salida reportando el estado de la red, la que alimenta un software que configura al firewall, para producir un rechazo de la amenaza en tiempo real.

Palabras claves: Detección de Intrusiones, Seguridad Integral, Alerta Temprana, Inundación de Paquetes, Denegación de Servicios, Redes de Datos.

Contexto

El presente trabajo de investigación se ocupa de los sistemas IDS del tipo NID, está avalada y subsidiado por la SECyT – UNC 05/M067.

Estado del Arte

Los intentos de vulnerar los sistemas de computadoras crecen día a día y el uso de métodos avanzados de encriptación no es suficiente para proteger los sistemas informáticos. Esto hace que sea necesario proveer a las redes con sistemas de protección bien planeados y políticas integrales de defensa contra los ataques.

Para ese rol juegan un papel importante los sistemas de alerta temprana o sistemas de detección de intrusiones (IDS Intrusion Detection Systems), como amplia contramedida de defensa. Un antecedente en este tema es el trabajo de Papavassiliou [1] que propone el método de detección de intrusiones estadística, usando como herramienta la estadística de Kolmogorov-Smirnov y redes neuronales para modelar y detectar ataques.

Introducción

Ataques a Computadoras

Existen básicamente dos tipos de ellos:

- Ataques Pasivos
- Ataques Activos

Ataques Pasivos

Estos son los de escucha sin autorización o de monitoreo de tráfico. Los objetivos de estos ataques consisten en obtener la mayor cantidad de información del mensaje transmitido y del oponente. Las distintas modalidades de ataques pasivos son las siguientes:

Descarga de contenidos del mensaje: Están incluidos dentro de este tipo de ataque la escucha de una conversación telefónica, la lectura de un mensaje de correo electrónico o la información confidencial capturada por un oponente.

Análisis de tráfico: Este es un ataque muy sutil. Se supone que hay medios de envíos de mensajes confidenciales, que no permiten al atacante poder acceder al contenido del mensaje. El atacante tiene sólo la posibilidad de observar la transmisión de los mensajes y obtener de éstos, por ejemplo datos tales como: la frecuencia de emisión de los mensajes y la longitud del mensaje. Esta información puede ser de mucha ayuda para inferir la naturaleza de la comunicación.

Los ataques pasivos son muy difíciles de detectar y reconocer, porque ellos son un medio de reconocimiento previo a la realización de ataques activos.

Ataques Activos

Los ataques activos involucran y comprometen los pilares básicos de las prácticas de seguridad: la confidencialidad, la integridad y la disponibilidad (CIA Confidentiality, Integrity and Availability). Los ataques activos son:

Denegación de Servicio: El efecto de este ataque es impedir la posibilidad de acceso a toda persona a un determinado servidor.

Enmascarado ("Masquerade"): En este caso el atacante se representa él mismo como un legítimo usuario con el objeto de robar, alterar o destruir recursos informáticos.

Reinterpretar ("Replay"): Este ataque es llevado a cabo mediante una captura pasiva de datos, para que luego sean retransmitidos y con ello producir efectos no autorizados.

Modificación de contenidos del mensaje: La información original es alterada de tal forma que permita obtener un resultado no autorizado.

Sistema de detección de intrusiones

La detección de intrusión en redes IDS es un componente vital en la defensa contra ataques a redes y es abordado desde diferentes perspectivas como puede verse en el trabajo de Bai[2]. En él se implementan dos métodos principales de detección: La detección de intrusiones de redes NID (Network Intrusion Detection), y la detección de intrusiones de servidores HID (Host Intrusion Detection).

Detección de intrusiones de redes

Los sistemas NID están relacionados con el tráfico de información entre servidores y clientes. Típicamente referidos como espías de paquetes (packet-sniffers), estos dispositivos interceptan paquetes que viajan por los medios de comunicación y transportan datos encapsulados en diferentes protocolos tales como, "Frame Relay" o ATM (Modo de Transferencia Asíncrona, Asynchronous Transfer Mode). Algunos dispositivos NID comparan el paquete con una base de datos de firmas de ataques conocidos y huellas digitales de paquetes maliciosos, mientras que otros analizan la actividad de paquetes buscando un comportamiento anómalo que pueda ser malicioso. En cualquiera de ambos casos un dispositivo IDS debe ser visto principalmente como una defensa perimetral de la red.

Los dispositivos NID en el pasado y por la complejidad de la tarea que realizan, han sido incapaces de operar en los siguiente ambientes:

- Redes conmutadas
- Redes encriptadas
- Redes de alta velocidad

Recientemente esta limitación ha sido superada por la potencia de procesamiento y cálculo de los microprocesadores modernos. Los conmutadores de redes ya vienen equipados con dispositivos IDS, capaces inclusive de realizar "packetsniffing" (espías de paquetes) a velocidades de gigabit. Las técnicas de NID pueden desagregarse en dos tendencias principales y complementarias entre sí: Detección de mal uso y Detección de anomalías.

La detección de mal uso como lo explica Vigna [3] modela ataques conocidos y realiza búsquedas de la ocurrencia de esos patrones.

Los sistemas de detección de anomalías, como lo señala Valdes [4] alertan de intrusiones mediante la observación de las desviaciones del comportamiento típico del tráfico de la red.

Detección de intrusiones de servidores

La detección de intrusiones basada en servidores, está diseñada para responder a ataques sobre un determinado servidor. Se basan en la supervisión de las acciones de los usuarios y de los archivos del servidor. En auditoria de los registros de actividad de los servidores y estado de los archivos del sistema, existen técnicas robustas que ofrecen administración de políticas de auditoria, análisis estadístico y soporte de evidencias, los que proveen medidas de control de la actividad de los servidores. La detección de intrusión en servidores sirve tanto para detectar ataques externos como internos.

Materiales y Métodos

En el análisis de la información se utilizan programas específicos y herramientas GNU. Los programas se desarrollaron en lenguaje C, en un ambiente Linux, bajo normas(CMM/ISO/IEC 12207), para asegurar calidad y capacitar en las prácticas adecuadas de ingeniería de software. Se utilizan simuladores de redes como el ns2 de la universidad de Berkeley. Para la simulación de las redes neuronales se utilizarán programas específicos y herramientas tales como Stuttgart Neural Network Simulator. El sistema de detección de intrusiones propuesto en el siguiente trabajo de investigación utiliza modelos estadísticos y una red neuronal clasificadora tipo, para detectar las condiciones anómalas de la red. El análisis estadístico basa su cálculo en el álgebra de las funciones de densidad de probabilidad (pdf), en lugar de usar muestras aisladas o sus promedios solamente. Éstas funciones pdf serán procesadas estadísticamente antes de ser ingresados a la red neuronal. la red neuronal integrará esta información en una sola salida reportando el estado de la red. En el presente trabajo se propone la implementación y las pruebas de funcionamiento en un ambiente real, utilizando el firewall (Pared cortafuegos) de la Universidad Nacional de Córdoba.

Resultados

Síntesis de Diseño e Implementación:

Módulo de captura: Este módulo recolecta datos del comportamiento de la red, midiendo seis parámetros: Tráfico de Protocolo Internet (IP Internet Protocol) en bps (bits por segundo), Tráfico IP en pps (paquetes por segundo), Tráfico de Protocolo de Datagrama de

Usuario (UDP, User Datagram Protocol) en bps, Tráfico UDP en pps, longitud de paquetes IP en bytes y longitud de paquetes UDP en bytes.

Preprocesador de eventos: Recibe reportes del módulo de captura y calcula la función de distribución de probabilidad (PDF) para los seis parámetros en estudio.

Procesador estadístico: Mantiene un modelo de referencia de la actividad normal de red, lo compara con los reportes del preprocesador de eventos para cada uno de los seis parámetros y forma un vector estímulo que combina todos ellos para alimentar la red neuronal.

Red neuronal clasificadora de ataques: Procesa el vector estímulo del procesador estadístico para decidir si el estado de la red es normal o se encuentra bajo ataque.

Visualizador de estado: Obtiene la información de salida de la red neuronal y la representa gráficamente en una curva en función del tiempo. Esta curva podrá estar disponible a través de una página web.

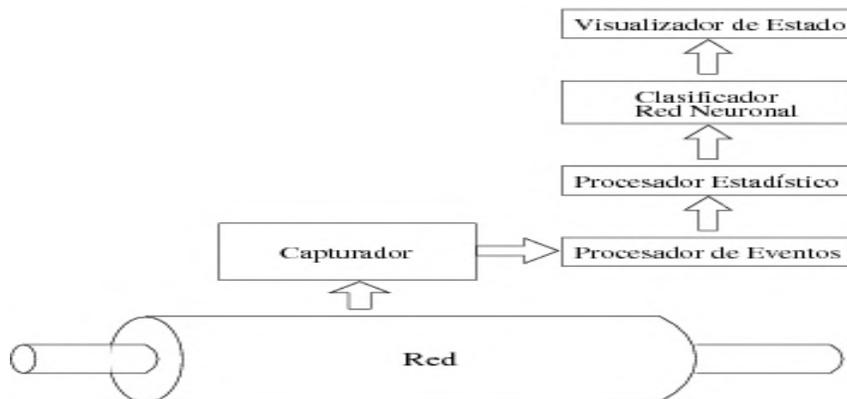


Diagrama de Bloques del IDS

Todo el sistema completo genera una salida indicando el nivel de ataque con una variable real entre los valores 0 y 1, donde el valor 1 representa una intrusión con absoluta certeza y una salida con valor 0 indica normalidad. Esta salida es actualizada con una frecuencia determinada.

Obtención de los parámetros de entrada y salida

Los patrones utilizados en el entrenamiento es un conjunto de entradas salidas conocidas. Este conjunto esta integrado tanto de patrones de tráfico normal de la red como de patrones de tráfico bajo condición de ataque. Para obtener los patrones de entrada de la red neuronal (vector k-dimensional cuyos componentes son los valores de similitud) se utiliza un programa que incluye solamente el bloque de captura. Con las muestras de entrada, se procedió al entrenamiento de la red propiamente dicha. Para ello se utilizó un programa que lee los datos del archivo de las muestras de la red y las procesa estadísticamente, generando los patrones de entrada para la red neuronal. Los patrones de salida, son conocidos, mediante el registro de los períodos de tiempo en los cuales se realizaron los ataques. Se asignó 1 para condición de tráfico normal y -1 para la condición de ataque. Una vez obtenidos los patrones de entrada y salida, se realizó el entrenamiento de la red. Los pesos

calculados de la red se almacenan en un archivo, que se utilizará en el próximo paso. El error de aprendizaje resultante fue de 0,56 %. Este error es el porcentaje de salidas de la red neuronal con falsos positivos y falsos negativos en relación al total de salidas. Se observa la convergencia del error cuadrático medio de la red neuronal, a medida que transcurren las iteraciones. Es importante notar que el error cuadrático medio disminuye muy rápidamente luego de sólo algunas iteraciones, alcanzando niveles de convergencia satisfactorios dentro de las 15 iteraciones. Estas características son especialmente deseables para sistemas de detección de intrusiones en redes (NID), los cuales necesitan supervisión en tiempo real y entrenamiento en línea. Una vez realizado el entrenamiento, utilizamos el segundo conjunto de muestras de tráfico de red y un programa para verificar el correcto aprendizaje de la red neuronal y analizar su generalización. El error de generalización obtenido fue del 1,97 %.

Conclusiones

En primer lugar, se vio que la red neuronal puede entrenarse para distinguir entre condiciones normales y en condiciones de ataque de la red. Se obtuvo un error de aprendizaje de 0,56 % y un error de generalización de 1,97 %, utilizando una configuración sencilla de la red neuronal (configuración back-propagation con sólo dos neuronas en la capa oculta). Este resultado se ha logrado por el uso del módulo procesador estadístico que colabora en forma eficiente en la detección de intrusiones.

Como se puede advertir, para la detección de inundaciones UDP, el aporte realizado por los parámetros relacionados con el protocolo IP es pequeño en relación al aporte de los parámetros del protocolo UDP al momento de llevar a cabo la detección. Por esta razón, se podría omitir la utilización de los parámetros del protocolo IP.

Trabajos Futuros

El sistema se podría adaptar para detectar otros tipos de intrusiones, tales como inundaciones TCP-SYN e ICMP (Internet Control Message Protocol, Protocolo Internet de Mensajes de Control). En el caso de inundaciones TCP-SYN, se debería incluir el uso de parámetros relacionados al protocolo TCP mientras que las inundaciones de paquetes ICMP se detectarían a través de los parámetros del protocolo IP. Todo ello sería factible siempre y cuando la red neuronal pueda aprender los distintos tipos de ataque.

Se debe tener en cuenta que en este caso, no tenemos control sobre el tráfico de fondo de la red. Por el contrario sí se tuviese, a través de programas de simulación y de modelado de tráfico de red, se podría analizar el rendimiento del sistema para distintos tráficos de fondo y distintos niveles de ataque para mejorar su desempeño.

Con el sistema IDS implementado, se detectaron ataques en la red, se dejó para una etapa posterior la elaboración de ataques más complejos y elaborar acciones defensivas a través de la construcción automática de reglas al firewall.

Publicaciones

“Statistical Intrusion Detection in Data Networks”. IEE Latin America Transactions.
Page(s): 373-380. Digital Object Identifier:10.1109/TLA.2007.4378531

Formación de Recursos Humanos

Obtención del grado de Maestría en Ciencias de la Ingeniería Mención Telecomunicaciones. En la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de Córdoba., por parte de uno de los integrantes del equipo de investigación.

Agradecimientos

A la Secretaría de Ciencia y Técnica de la Universidad Nacional de Córdoba que avala el proyecto.

Bibliografía

- [1] C. Manikopoulos and S. Papavassiliou, "Network intrusion and fault detection: a statistical anomaly approach", *Communications Magazine*, vol. 40, pp. 76-82, Oct. 2002.
- [2] Y. Bai and H. Kobayashi, "Intrusion detection systems: technology and development", *Advanced Information Networking and Applications*, 2003. AINA 2003.17th International Conference on, vol. Issue, 27-29, pp. 710 - 715, March 2003.
- [3] G. Vigna and A. Kemmerer, "Netstat: a network-based intrusion detection approach", in *Computer Security Applications Conference*, vol. 1, Dec. 1998, pp. 25-34.
- [4] A. Valdes and D. Anderson, "Statistical methods for computer usage anomaly detection using nides", Tech. rep.: SRI International, Jan 1995.
- [5] Y.Ñong et al., "Statistical process control for computer intrusion detection", *DARPA Information Survivability Conference Exposition II*, 2001. DISCEX '01. Proceedings, vol. Volume 1, pp. 3 - 14, June 2001.
- [6] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY: Oxford Univ. Press, 1995.
- [7] P. Ramasubramanian and A. Kannan, "Intelligent multi-agent based backpropagation neural network forecasting model for statistical database anomaly prevention system", *Intelligent Sensing and Information Processing*, 2004. Proceedings of International Conference on, pp. 108 - 113, 2004.
- [8] S. C. Lee and D. Heinbuch, "Training a neural-network based intrusion detector to recognize novel attacks", *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, vol. Volume 31, pp. 294 - 299, July 2001.

UTN VoIP Test Bed (Voice over Internet Protocol)

*Patricia Clérigo, Gustavo Mercado, Armando Lima y Gerardo Gosetto
{pclerigo, gmercado, gdgosetto}@frm.utn.edu.ar*

*gridTICs – Grupo UTN de Investigación y Desarrollo en TICs
Departamento de Electrónica
Facultad Regional Mendoza-Universidad Tecnológica Nacional
<http://www.gridtics.frm.utn.edu.ar>*

CONTEXTO

El proyecto UTN VoIP Test Bed se inscribe en el área de proyectos de desarrollo y transferencia en la temática Redes de Datos que desarrolla el Grupo UTN de I&D en TICS (gridTICs), de la Facultad Regional Mendoza.

El gridTICs desarrolla e investiga en temáticas de redes de datos tales como IPv6, Telemedicina, Diseño y análisis de redes LAN, Wireless Sensor Networks, Embedded Networking y Sistemas Reconfigurables; de los cuales se han presentado trabajos en los principales congresos de computación e informática del país.

RESUMEN

La tecnología Voice over IP permite transmitir paquetes de voz usando los protocolos de Internet. Esto es el basamento de la convergencia de video, voz y datos en una sola red y bajo el mismo protocolo; metodología que promete confiabilidad, accesibilidad y por sobre todo bajos costos. La UTN VoIP es un servicio de Telefonía que usa los protocolos VoIP de Internet y es montado sobre la Red Universitaria Tecnológica II. Cuando esté implementada conectará todas las Facultades Regionales y la Unidades Académicas de la UTN, convirtiéndose en la red de VoIP académica más amplia del país. En este trabajo se presenta el diseño y desarrollo de un test bed a escala de la UTN VoIP. Los elementos presentes, tales como: Server/Proxy SIP, Gateways, soft-phones e ip-phones, se arreglan y conectan para simular las distintas redes de Regionales. Se elije el protocolo SIP (Session Initiation Protocol) para la señalización y se prueban distintos codex (Codification/Decodification) para la conversión de la voz analógica en digital. El test bed también soportará autenticación, contabilidad, control de gestión, seguridad y calidad de servicio y se verificará el plan de numeración diseñado para la UTN VoIP

1. INTRODUCCION

La telefonía IP, también conocida por VoIP (Voice over IP), puede ser definida como cualquier

aplicación telefónica usada en una red de datos que utiliza el protocolo “Internet Protocol” (IP) [1].

La utilización de voz sobre IP racionaliza el uso de la infraestructura de comunicaciones, posibilitando la convergencia de datos y voz por la red Internet, con disminución de costos tradicionales de telefonía, principalmente cuando las llamadas son de larga distancia. La tecnología VoIP permite extender el acceso telefónico a localizaciones donde existe una red de datos, mas no existe la disponibilidad de ramales telefónicos, además de incorporar soporte para movilidad del usuario. El soporte de movilidad permitirá que los usuarios de una institución puedan ser localizados en cualquier parte del globo y, cada vez que utilicen el servicio, bastando para tal tener acceso a Internet (banda ancha) y un teléfono IP configurado.

Ventajas adicionales de VoIP en redes académicas de la telefonía sobre una red de datos se basan en la flexibilidad de Gestión de Red (ABM y Tráfico) [2] de este tipo servicio en redes privadas/públicas y en la Economía de escala (Reducción de costos y Escalamiento/Optimización de red) que ello genera, a saber:

1. Bajo costo por llamada, en especial llamadas de larga distancia e internacionales
2. Bajo costo de infraestructura: una vez que la infraestructura IP está instalada, ninguna o casi ninguna infraestructura adicional de telefonía se necesaria;
3. Interoperabilidad con la red pública de telefonía (vía PBX);
4. Integración con servicios y aplicaciones ya existentes dentro de la institución.

Además permite la flexibilidad de ABM/Rutas, cuando se desea hacer un ABM (Alta/Baja/Modificación) de usuarios de softpone/hardphone y rutas estos sistemas son mas fáciles de administrar, que los sistemas de telefonía tradicional los cuales usan comandos Hombre/Máquina propietarios de cada fabricante, es decir, son sistemas fácilmente integrables a internet por lo tanto, cualquier servidor web con un sistema de bases de datos opensource y con un nivel de scripting básico permitiría a cualquier

persona gestionar todo el sistema de telefonía sin tener que recurrir al fabricante ni a un especialista.

Mediante la Gestión de Tráfico Telefónico se puede realizar un control de la calidad de servicio telefónico provisto por el ITSP (Internet Telephony Service Provider) analizando el ASR (Average Success Rate: Porcentaje Promedio de llamadas Exitosas). Al mismo tiempo, si se contratan 2 o mas ITSP se puede cambiar rápidamente de proveedor, con un “click de mouse”, (si el ASR provisto es bajo o si hay aumento de tarifas) o simplemente para enrutar ciertas llamadas por cada proveedor en función del precio/calidad ofrecido al destino de interés. Esta facilidad sería impensable en los sistemas telefónicos tradicionales ya que éstos últimos tienen una conectividad física con cada proveedor y no lógica como sucede con los ITSP’s.

Por lo tanto se produce una reducción de costos, desde el punto de vista económico, ya que se elimina el costo de llamadas entre sucursales/delegaciones distribuidas geográficamente (tráfico on-net). Además todo el tráfico off-net (tráfico saliente de la red académica/corporativa) local y larga distancia se puede sumar para generar una masa crítica de minutos y negociar el precio de toda esa “bolsa” de minutos con aquel ITSP que ofrezca mejor Costo/Beneficio.

Por otro lado se produce una mejora en el escalamiento/optimización de la red, a medida que la red crece, debido a que los sistemas de ToIP OpenSource son mucho mas económicos que los sistemas tradicionales ya que éstos últimos se compran con una cantidad máxima de licencias (usuarios, casillas de voice-mail) por encima de la cual hay que pagar dinero adicional, si la red sigue creciendo. Al mismo tiempo, cada vez que se activa un usuario nuevo se usa la misma infraestructura de red existente (Red LAN) y no se necesita “cablear”, ni “puentear” pares telefónicos para dar servicio de telefonía optimizando de esta forma tiempo y dinero

2. LINEAS DE INVESTIGACION y DESARROLLO

El UTN VoIP Test bed permite la simulación, el estudio y la puesta a punto de la red UTN VoIP, la cual es la red de Telefonía IP de Universidad Tecnológica Nacional.

Para la red UTN VoIP se plantea el siguiente objetivo general [3]:

1. Implementar una red VOIP experimental que cubra las comunicaciones telefónicas entre las Facultades Regionales de la UTN, empleando la infraestructura de red institucional

Y los siguientes objetivos específicos:

1. Realizar un sistema que optimice el uso de las redes de datos, realizando convergencia entre voz, datos y video.
2. Disminuir los costos de comunicaciones telefónicas entre Regionales.
3. Obtener experiencia en el diseño, implementación y manejo de redes convergentes complejas.
4. Crear sistemas tecnológicos de VoIP transferibles a la academia y a la industria.
5. Optimizar el uso de la red Internet 2 y de la red RUT2

Existen varias iniciativas de Telefonía IP en redes académicas, siendo las mas conocidas en Latinoamérica la red RNP [4] de Brasil y la REUNA de Chile [5].

2.1 Metodología

2.1.1 Protocolo SIP

Session Initiation Protocol (SIP) [5] fue desarrollado por el IETF como un estándar de los protocolos de Internet. SIP es un protocolo muy rico y extensible y que tiene muchas capacidades adicionales, más allá de la tarea de señalización usada en VoIP. SIP es un verdadero protocolo maestro para establecer sesiones sobre Internet.

Un sistema VoIP basado en SIP tiene los siguientes componentes:

- UAC (Agente Usuario Cliente): Inicia las peticiones SIP y actúa como agente usuario llamante
- UAS (Agente Usuario Servidor): Recibe las peticiones y devuelve las respuestas en nombre del usuario, actúa como agente usuario llamado
- Gateway: Realiza la interfase entre paquetes de voz ip y la red telefónica publica/privada.

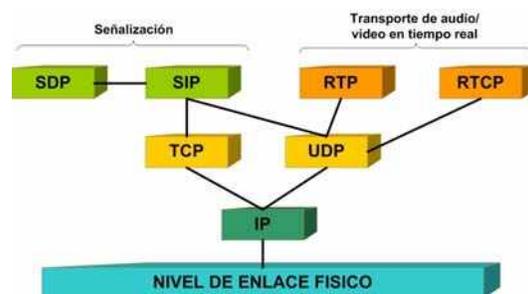


Figura 1: Pila de Protocolos SIP

En la figura 1 se puede ver la pila de protocolos SIP aplicados a VoIP, donde se distinguen los siguientes componentes:

- Protocolo de capa de aplicación para Establecer, Mantener y Terminar sesiones multimedia.
 - o Llamadas en Internet sobre cualquier red IP.
 - o Localización.
 - o Negociación de características entre end-points.

- Basado en RFC 2543 [6], RFC 3216 [7]
- SDP (Session Description Protocol), descripción de sesiones de comunicación multimedia [8].
- RTP (Real Time Protocol), para asegurar el transporte de voz digitalizada [9].
- RTCP (Real Time Control Protocol), que controla la transmisión de voz [10].

2.1.2 UTN VoIP Test Bed

Se investiga y desarrolla un prototipo a escala de VoIP con los distintos componentes de la red, que permite la simulación, estudio y puesta a punto de los procedimientos y métodos de implementación de la red UTN-VoIP. El plan de trabajo implica realizar las siguientes etapas de desarrollo:

1. Se diseña la ingeniería conceptual de la maqueta y se especifican las características principales.
2. Se desarrolla y se montan los componentes de hardware del sistema
3. Se instalan y configuran los componentes de software del sistema
4. Se realizan pruebas de aceptación y performance
5. Se documenta y se relevan los resultados asociados a la red UTN-VoIP

La Maqueta UTN-VoIP se construye en el ambiente XNGN (Experimental Next Generation Network) del Grid TICs y se basa en protocolo SIP para la transmisión, manejo y señalización de paquetes de voz.

El prototipo se realiza en dos etapas. La primera tendiente a resolver las comunicaciones SIP dentro de las redes de cada regional perteneciente a la Red UTN-Voip. La segunda se basa en la administración de las comunicaciones entre regionales de la UTN, utilizando la red RUT2.

2.1.2.1. Primera Etapa

Se montó un servidor B2BUA [11], con software de código abierto basado en estándares, denominado Servidor PBX Mendoza. En la maqueta (Figura 2), en el que se configuraron distintos clientes. Una vez probado el funcionamiento básico de la central PBX, llamadas entre los distintos clientes, tanto en softphones como en teléfonos IP, se comenzaron a configurar las distintas funcionalidades que presta el servidor:

- IVR (Interactive Voice Response) que es un menú con el que el usuario puede interactuar mediante pulsaciones DTMF.

- Transferencia de llamadas, ciega y atendida.
- Buzón de voz, que permite grabar un mensaje en caso de no estar disponible el usuario y además permite enviar el mensaje de voz al usuario por e-mail.
- Call Parking, permite transferir la llamada a una extensión (ciega) del servidor la cual puede ser recuperada desde otro teléfono.

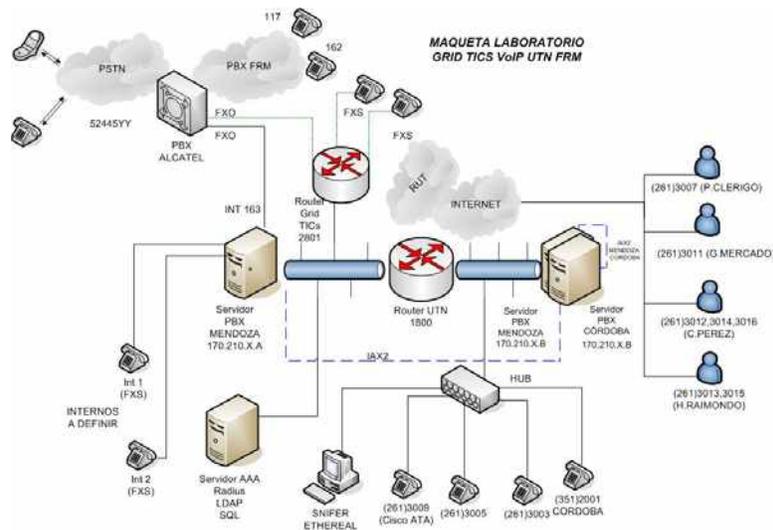


Figura 2: UTN VoIP Test Bed

- Música en espera.
- Follow Me, el usuario puede programar desde su teléfono, en qué extensión se encontrará y la llamada se transferirá de forma automática.
- Meetme, configuración para crear una sala de conferencia, donde los participantes ingresan a una sala virtual de conferencias con solo discar un número.
- Generación de CDR, que permita monitorear las comunicaciones locales

A continuación, se instaló otro servidor con las mismas configuraciones denominado Servidor PBX MendozaB. Entre ambos servidores se probaron comunicaciones entre clientes registrados en cada uno de ellos. Se analizó el comportamiento del protocolo IAX (Inter Asterisk eXchange v2.) Protocolo exclusivo para enlaces entre Asterisk, aunque hay algunos hardphones IAX2 que también lo soportan. En este protocolo los streams de audio y la señalización viajan por el mismo puerto: 4569 UDP. Es un protocolo binario, NO basado en texto. Es posible enviar varias conversaciones por el mismo flujo, lo cual supone un importante ahorro de ancho de banda (overhead de la capa IP y transporte UDP). Usa autenticación – MD5 y necesita una fuente de tiempo: ZapTel (ztdummy).

En este punto se llevaron a cabo distintas capturas de paquetes con Ethereal [12] con el propósito de evaluar el cumplimiento de los estándares por parte de software.

Se realizaron pruebas exhaustivas de monitoreo, para evaluar los distintos codecs utilizados en las comunicaciones.

Se anexó una placa analógica al Servidor PBX MendozaA, con el objetivo de enlazar el servidor con la central telefónica de la universidad. Dicha placa [13] posee 2 puertos FXO y dos puertos FXS. Dicha conexión permite entonces que este servidor actúe como Gateway entre la telefonía SIP y la telefonía tradicional.

Luego de realizadas todas las pruebas de comunicación entre ambos servidores, se configuró un tercer servidor para simular una regional situada en otra provincia (Córdoba) que figura en la maqueta como Servidor PBX Córdoba. Entonces para poder comunicarse con los usuarios registrados en este servidor, realizando llamadas desde los otros servidores (Mendoza), se debía anteponer la característica de la provincia (351). Y para realizar llamadas desde el servidor Córdoba hacia los de Mendoza se debía anteponer (261).

De esta manera se empezó a esbozar un Plan de Numeración [1] y [14] que contemplara la numeración para todas las Regionales :

Los identificadores serán de la forma:

<CA>I<ORG><TERM>

donde:

CA: Código de Área (3 dígitos) – El mismo código DDN de la ciudad usado en la telefonía convencional donde se localiza la regional de la universidad.

I: Identifica explícitamente la naturaleza del número. la presencia del “1” identifica que es una llamada IP, para evitar posibles confusiones con teléfonos válidos de la red pública.

ORG: Prefijo VoIP de la institución (3 dígitos) – este número se agregó con el objeto de proveer en un futuro la apertura del servicio a otras instituciones (centros de investigación, universidades, etc.) que será independiente de la localización geográfica de la institución.

TERM: identificador del cliente VoIP, conformado por cuatro dígitos numéricos.

2.1.1.2.2. Segunda Etapa

La segunda etapa consiste en proveer a la red de una interconexión entre regionales que resulte transparente para los administradores de los servidores implementados en la primera etapa y que están ubicados en cada regional.

Para ello se recurrió a la instalación y configuración de un Proxy, también de código abierto y que cumpliera con los estándares SIP.

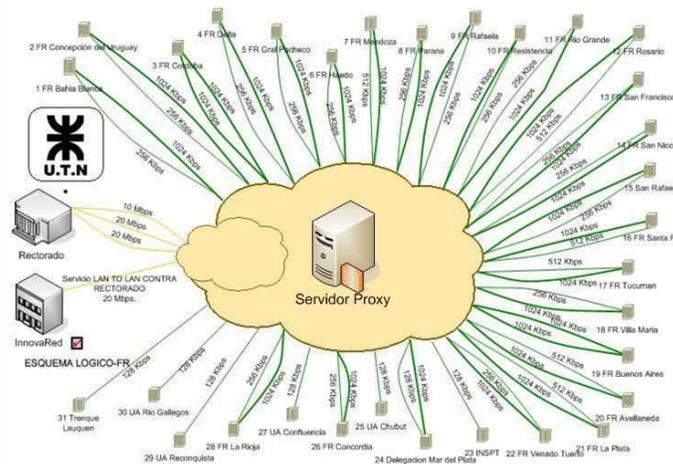


Figura 3: Esquema de UTN VoIP

Todos los servidores locales/regionales se conectarán a este servidor Proxy, quién será el encargado de establecer las comunicaciones entre las regionales y que cuenta con las siguientes características:

- Autenticación de los servidores de las regionales con TLS.
- Registración de los diferentes dominios.
- Generación de CDR: Call Detail Record, registrando el origen de la llamada, el destino y la duración de la misma. Estos datos son utilizados para evaluar el tráfico generado por las llamadas IP y poder establecer las necesidades de ancho de banda necesario para mejorar la performance del servicio.

La segunda etapa aún no está concluida, aunque se está montando y probando el Proxy SIP.

3. RESULTADOS OBTENIDOS/ESPERADOS

Las pruebas demostraron que es posible conseguir la misma calidad que la telefonía tradicional. No hay que descuidar temas fundamentales como los codec a utilizar, el establecimiento de prioridad de tráfico VoIP dentro de la red y la reserva adecuada de ancho de banda según la cantidad de canales de voz que se utilicen.

En cuanto a la fiabilidad, en redes LAN se puede lograr un a gran fiabilidad, como así también en la red RUT2 si se administran la cantidad de canales que se adjudicarán a cada regional,

Los usuarios que se registren a los diferentes servidores, fuera de las regionales, desde Internet contarán con diferentes factores que atentarán contra la calidad y fiabilidad de las comunicaciones,

ya que nos encontramos con diferentes dificultades como QoS, los diferentes tipos de NAT, el DELAY, el JITTER, que ya no podemos administrar por ser redes públicas.

Aunque se logra implementar una buena flexibilidad técnica para ampliar la red, para Operar y Mantener la misma y económica para reducir costos del minuto telefónico.

3.1 Propuestas

Después de hacer una evaluación parcial del proyecto UTN VoIP, surgen un par de propuestas que, de ser implementadas, podrían dotar a la red de características únicas en el país. Tal es la naturaleza de estas propuestas que el nombre de UTN-VoIP debería cambiarse por otro más significativo tal como UTN-Phone. A continuación las propuestas:

Se propone tener un "Proveedor PSTN" único para UTN-Phone. Las redes académicas, distribuidas geográficamente, deberían cursar todo su tráfico saliente, que es el que tienen costo, a través de este ITSP's. La ventaja de esta propuesta es eminentemente económica, ya que se negociaría con un solo proveedor el tráfico de voz de todas las Facultades Regionales de la UTN. En esta propuesta aún las Regionales deben mantener la conexión con el proveedor local para el tráfico de las llamadas entrantes.

La otra propuesta consiste en obtener para UTN-Phone la licencia de Operador de Telecomunicaciones. La gestión de la licencia de servicios de telecomunicaciones, le permitiría a las redes académicas, interconectarse con la RTCP (Red Telefónica Pública Conmutada), mediante protocolo SS#7, obteniendo un sistema de numeración propio. Esto permitiría reducir aún mas los costos y realizar I+D, tal como la tecnología ENUM.

4. FORMACION DE RECURSOS HUMANOS

En la UTN-VoIP, al ser una red nacional, están involucrados diferentes integrantes de la comunidad académica de la UTN en la distintas Facultades Regionales.

En el caso de la Facultad Regional Mendoza y el GridTICs se han formado tres maestrandos en Redes de Datos (UNLP), dos Becarios Graduados UTN, dos Becarios Alumnos Telefónica y dos Becarios Alumnos UTN. Además el Laboratorio de VoIP ofrece prácticas para las cátedras de Redes de Datos y Comunicaciones, tanto de grado como de postgrado.

5. BIBLIOGRAFIA

[1] Bill Douskalis, "IP Telephony, the Integration of Robust VoIP Services", Prentice Hall PTR, ISBN-10: 0130141186, 1999

[2] "The Real World of VoIP", NetworkWorld Executive Guide

[3] G. Mercado, A. Lima, P. Clérigo, G. Gosseto, "UTN VoIP: Especificaciones y Plan de Proyecto", Documento Interno UTN FRM, Agosto 2007.

[3] Rede Nacional de Ensino e Pesquisa, <http://www.rnp.br/>

[4] Red Universitaria Nacional de Chile, <http://www.reuna.cl/>

[5] H. Sinnreich, A. Johnston y R. Sparks, "SIP Beyond VoIP", VON Publishing LLC, ISBN 0-9748130-0-1, 2005.

[6] M. Handley H. Schulzrinne - E. Schooler - J. Rosenberg "SIP: Session Initiation Protocol", RFC 2543, IETF, March 1999

[7] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Saprks, M. Handley y E. Schooler, "SIP; Session Initiation Protocol", RFC: 3261, IETF, June 2002

[8] M. Handley V. Jacobson "SDP: Session Description Protocol", RFC 2327 April 1998

[9] H. Schulzrinne S. Casner R. Frederick V. Jacobson, "RTP: Real-time transport protocol", RFC 1889 January 1996

[10] C. Huitema "RTCP: Real-time transport control protocol", RFC 3605 October 2003.

[11] Asterisk: The Open Source PBX & Telephony Platform <http://www.asterisk.org/>

[12] Ethereal: A Network Protocol Analyzer <http://www.ethereal.com/>

[13] VoIP Telephony Solutions - *Digium*® The Asterisk Company <http://www.digium.com/>

[14] Jim Van Meggelen, Leif Madsen, Jared Smith, "The Future of Telephony", O'Reilly Media Inc, 2005

◇

Análisis experimental del comportamiento del protocolo 802.11e en enlaces troncales.

Roa, Pablo F.
Departamento de Informática.
Facultad de Ingeniería y Ciencias Hídricas
Universidad Nacional del Litoral
Santa Fe, tel +54 342 4575233 Int 143
proa@fich.unl.edu.ar

Contexto.

En la FICH existe el proyecto de investigación CAI+D *“Desarrollo de sistemas de adquisición y transmisión de datos hidrodinámicos empleando tecnología de servidores microweb y su disposición en Internet mediante modelo cliente – servidor”* donde los temas de transmisión de datos son relevantes si empleamos tecnologías wireless. Si bien este trabajo no se encuentra relacionado directamente con el núcleo del proyecto de investigación; realiza aportes que pueden ser tomados en cuenta a la hora de realizar un proyecto complementario. Por otro lado desde la cátedra de Redes y Comunicación de datos II, se promueve mediante estas líneas de investigación, presentar la iniciativa para desarrollos posteriores que terminen en proyectos finales de carrera de grado y / o formadores de proyectos de investigación.

Resumen.

La comunicación inalámbrica es uno de los principales medios de interconexión en teleinformática, mas especialmente debido al auge que promovió la norma 802.11 al ser de uso libre. La norma generó un gran campo de investigación, en el área de señales y modulación, como así también en la generación de nuevos protocolos y el análisis del comportamiento de los tráficos. Si bien el ambientes cableados ya existía una norma que promovía la calidad de servicio, su utilización es reciente en redes wireless.

En este trabajo se realiza un estudio del comportamiento del tráfico de enlaces troncales bajo diferentes situaciones, utilizando la norma 802.11e o también denominada WMM. Si bien existen numerosas simulaciones, los análisis de tráfico real no son frecuentes. Se encuadra en las líneas de investigación de comportamiento de protocolos y análisis de tráfico reales no simulados; en la transmisión de datos mediante radio frecuencia.

Palabras clave: redes de datos, análisis de tráfico, wireless, 802.11e.

1. Introducción.

Resulta notable la expansión en los últimos años, de todo tipo de interconexión, sean cableadas como no cableadas, fundamentalmente impulsados por la necesidad de acceso a Internet. Una de estas tecnologías (802.11) ya se encuentra en una etapa madura de evolución, donde se presentan nuevas características y líneas de desarrollo donde podemos destacar la más reciente 802.11n. Tabla 1.

También podemos destacar que existen implementaciones no estándar que posibilitan mejorar la transferencia bajo condiciones de mucho ruido o bien incrementar el throughput (volumen de datos útiles) del cual este trabajo no aborda.

Los equipos recientes están incorporando opciones que los usuarios desconocen su utilización su comportamiento y menos su fin. Al habilitarlos pueden incrementar su rendimiento (bajo determinadas circunstancias) con determinados equipos clientes (remotos)

pero en detrimento de los demás conectados inalámbricamente.

Tabla 1.

Group Task		Año
802.11	Estándar original	1997
802.11a	Estándar en 5.3-5.8 Ghz	1999
802.11b	Estándar en 2.3-2.4 Ghz	1999
802.11c	Modo Bridge	
802.11d	Soporte de Roaming	2001
802.11e	Soporte de QoS	2005
802.11g	Transferencia de 54 Mbps en 2.4 Ghz	2003
802.11h	Mejoras en la transmisión de 5.8 Ghz	2003
802.11i	Extensiones en la Capa MAC	2004
802.11n	Transmisiones hasta 300 Mbps.	Activo
802.11p	Acceso inalámbrico para vehículos.	Activo
802.11s	Redes MESH	Activo

La forma de acceso al medio del protocolo 802.11 juega un papel importante en los retardos debido a los tiempos interframe que se establecen y a la cantidad de estaciones. En nuestro análisis esta problemática se minimiza debido a la implementación en modo bridge donde solo se establecen dos estaciones.

Los enlaces troncales sirven para intercomunicar distintos Access Point (AP), de manera de conducir el tráfico desde el NOC central hasta cada uno de los AP; este trabajo usualmente se realiza mediante 802.11a. Un inconveniente que se presenta es la acumulación desordenada del tráfico en los troncales, donde según su procedencia no existe ningún clasificador a nivel de capa 2. La norma 802.11 no establece ningún tipo de priorización a no ser que se realizara mediante (Type of Service) o *DSCP* (Diffserv Codepoint).

A medida que el protocolo IP posibilita que nuevas aplicaciones con características especiales puedan ser utilizadas

masivamente (como la telefonía IP, video streaming); se hace imprescindible la utilización de priorización no solo en capa 3 sino también en capa 2.

Si bien 802.11e no posee la habilidad de reservar ancho de banda, el protocolo provee la utilización de 4 colas de prioridad que analizaremos en el item siguiente. Los escenarios actuales son mucho más complejos que el análisis que aborda este trabajo, debido a los múltiples enlaces troncales donde una antena debe soportar varias conexiones aunque en un número menor al de la implementación directa en modo Access Point. Figura 1.

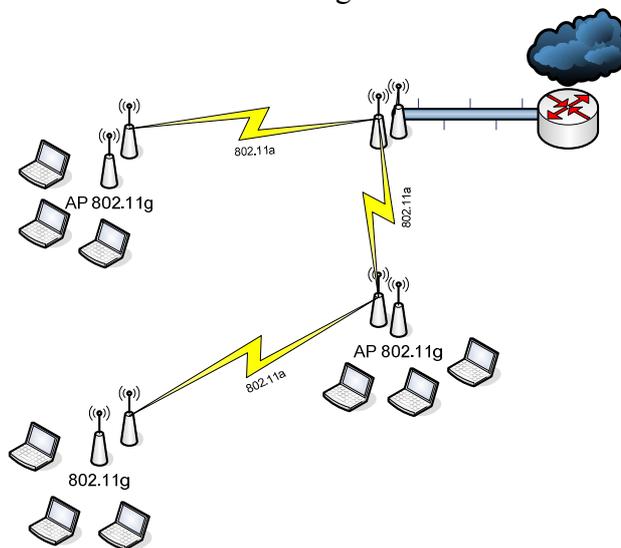


Figura 1

2. El protocolo 802.11e.

Este protocolo se basa en EDCA (Enhanced Distributed Channel Access Function) en el que establece la posibilidad de priorizar distintos tráficos a la hora de ganar el acceso al canal durante el período de contención, y mantiene la compatibilidad con el modo DCF.

La figura 2 compara el estándar 802.11 con el protocolo 802.11e. Podemos observar la implementación de nuevos períodos (AIFS) entre los parámetros más destacados, destinados a establecer períodos de contención en conjunción con la variación del tamaño de la ventana de contención CW_{min} y CW_{max} .

De esta manera, el protocolo 802.11e permite "dividir" el tráfico en 4 categorías: Background, Best Effort, Video y Voice,

permitiendo definir una prioridad a cada uno para definir las políticas de QoS.

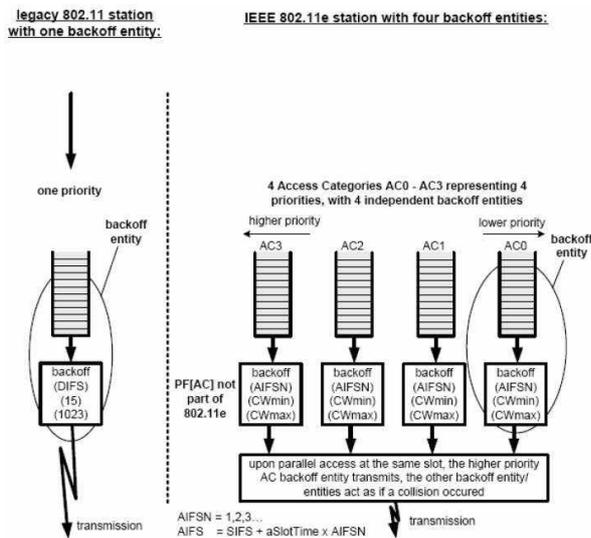


Figura 2.

3. Implementación.

Nuestra implementación ha utilizado dos equipos Routerboard corriendo el Sistema Operativo Mikrotik versión 3.20, por intermedio de un enlace en 5.8 Ghz, conectando dos redes LAN de las cuales se enviaron los tráficos de voz y datos. Para minimizar los efectos de ruidos e interferencias como así también variaciones en las señales; las placas se dispusieron en un entorno físico de no más de 2 mts. La figura 3 muestra la implementación de la topología física bajo prueba.

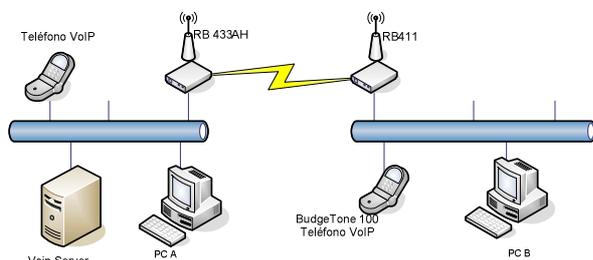


Figura 3.

Para tener un marco referencial del comportamiento, se plantea medir el throughput de datos TCP y UDP sin WMM, a distintas velocidades de conexión (48 Mbps, 36 Mbps, 24 Mbps, 18 Mbps, 11 Mbps, 5.5 Mbps. En ambas redes LAN se disponen de

switchs de 100 Mbps habilitados con 802.1p, 2 colas por puerto. En una segunda etapa se pretende medir el flujo de datos en condiciones donde el tráfico consiste solamente en VoIP, implementando un servidor bajo protocolo SIP, teléfonos de VoIP y computadoras (PC A y PC B) que posean una aplicación de VoIP.

En una tercer etapa se pretende generar tráficos de VoIP y datos en forma concurrente para verificar el funcionamiento sin la implementación de WMM.

En una cuarta etapa se pretende habilitar 802.11e y enviar sólo tráfico VoIP; con posterioridad y de forma concurrente el tráfico de datos y de voz.

Las herramientas de medición son:

- ✓ VQmanager
- ✓ Traffic Flow.
- ✓ Ethereal traffic sniffer.
- ✓ Observer versión 11.

El servidor empleado como PABX es el 3CX de 3Com en su versión 6.0 donde se utilizó SIP como protocolo de VoIP con codecs PCMU, PCMA. En la figura 4 se pueden observar los dos Routerboard y el teléfono IP.



Figura 4.

4. Resultados.

El proyecto se encuentra en la fase 1, donde se establecen las condiciones basales de tráfico. Se espera obtener para el mes de mayo, las mediciones de las 4 fases.

5. Conclusiones.

Aun sin conclusiones.

6. Bibliografía

- [1] J. W. Robinson y T. S. Randhawa. Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function. *IEEE Journal on Selected Areas in Communications*, 22(5):917–928, Junio 2004.
- [2] A. Banchs y L. Vollero. A Delay Model for IEEE 802.11e EDCA. *IEEE Communications Letters*, 9(6):508–510, Junio 2005.
- [3] Z. Kong, D. H. K. Tsang, B. Bensaou, y D. Gao. Performance Analysis of IEEE 802.11e contention-based channel access. *IEEE Journal on Selected Areas in Communications*, 22(10):0–0, Diciembre 2004.
- [4] J. Hui y M. Devetsikiotis. A Unified Model for the Performance Analysis of IEEE 802.11 EDCA. *IEEE Transactions on Communications*, 53(9):1498–1510, Septiembre 2005.
- [5] A. Banchs y P. Serrano. Revisiting 802.11e EDCA Performance Analysis.
- [6] B.-J. Kwak, N.-O. Song, y L. E. Miller. Analysis of the Stability and Performance of Exponential Backoff. En *Proceedings of WCNC'03*, Marzo 2003.
- [7] P. Serrano, A. Banchs, T. Melia, y L. Vollero. Performance Anomalies of Nonoptimally Configured Wireless LANs. En *Proceedings of IEEE WCNC '06*, Las Vegas, Nevada, EEUU, Abril 2006.

Cluster Modular Autocontenido

Gabriel E. Arellano, Leonardo J. Hoet
Departamento Ingeniería en Sistemas de Información
Facultad Regional Concepción del Uruguay
Universidad Tecnológica Nacional
[arellanog,hoetl]@frcu.utn.edu.ar

RESUMEN

Desde hace años es común en organizaciones educativas y de investigación el desarrollo de actividades que requieren grandes capacidades de cálculo. La solución más común a estas necesidades es la compra o construcción de un cluster HPC (High Performance Computing), pero construir un cluster acarrea problemas y costos inesperados problemas al momento de utilizarlo y mantenerlo en operaciones. Nuestro proyecto busca desarrollar y documentar técnicas para construir un cluster HPC que sea fácilmente ampliable y que minimice los problemas que normalmente acarrea la operación del mismo. Los resultados preliminares son alentadores, en este punto podemos asegurar un ahorro de energía de hasta un 27%, un ahorro de espacio de hasta el 30%, y un ahorro en los costos de hasta el 11% respecto de un cluster HPC “tradicional”.

Palabras Clave: cluster, hpc, modular, autocontenido, técnicas de construcción, cómputo paralelo.

1. INTRODUCCIÓN

En la actualidad es casi inconcebible pensar en computo intensivo sin asociarlo directamente con un cluster de computadoras personales. Un cluster de computadoras es un conjunto de computadores independientes (llamadas nodos) funcionando como si fueran un solo equipo (con la suma de los recursos de sus elementos).

Es por esto que cualquier organización que necesite capacidades de cálculo en algún momento se ve enfrentada a la posibilidad de adquirir o construir su propio cluster de computadoras. Al menos al principio, la organización optará por la construcción de un cluster de computadores personales, esto se debe principalmente a cuestiones de oportunidad, costos y escala [sterling02].

La tarea de construir un cluster (independientemente de su tamaño), involucra mucha planificación y gran cantidad de decisiones: composición de cada nodo, cantidad de nodos, comunicación de los nodos, infraestructura [steiner01] (espacio físico, soportes, alimentación eléctrica, refrigeración, protección y seguridad), sistema operativo a emplear, aplicaciones a instalar, etc.

Uno de los aspectos antes mencionados es el que suele causar mayores dificultades ya que se lo suele subestimar en la planificación, nos referimos a la infraestructura necesaria para dar soporte al cluster [feng04].

En organizaciones pequeñas suele ser complejo conseguir un espacio físico que cuente con los requerimientos de un cluster (alimentación eléctrica estabilizada y/o ininterrumpida para cada uno de los nodos, control de temperatura, control de acceso), además es necesario alojar los nodos y los dispositivos de comunicaciones de manera ordenada, lo que hace necesario estanterías y mecanismos para ordenar los cables de cada nodo [kok01].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Nuestro objetivo es desarrollar, documentar y verificar un conjunto de técnicas para la construcción, instalación y operación de un cluster para cálculo intensivo tratando de minimizar los inconvenientes comúnmente asociados a este tipo de actividades.

En particular nos interesa atacar los aspectos que suelen obstaculizar más gravemente un proyecto de implementación de un cluster de computadoras, y que en muchos casos, termina disuadiendo a la organización de la idea de tener su propio cluster.

Al final de nuestro proyecto habremos generado un conjunto de técnicas fácilmente transferibles y de simple implementación que permitirán a cualquier

organización de pequeño tamaño construir un cluster de pequeñas dimensiones físicas, económico y escalable a sus necesidades futuras.

Los problemas que nos interesa explorar son:

- Estrategias de construcción que minimicen el espacio requerido y los requisitos técnicos del citado espacio.
- Conjuntos de componentes que reduzcan los costos de construcción y operación, y que sean fácilmente asequibles en nuestro país.
- Estrategias que faciliten la instalación, operación y mantenimiento del cluster.

Como se mencionó anteriormente, uno de los escollos más grandes en la construcción de un cluster es la infraestructura necesaria: estanterías, conectores de alimentación, refrigeración, control de acceso (estos dos últimos que suelen hacer necesario reservar un espacio físico exclusivamente para el cluster). Para solucionar este problema proponemos como solución la utilización de racks¹ normalizados para alojar en su interior los elementos del cluster. De esta manera el cluster se encontrará “autocontenido”, es decir todos sus elementos estarán dentro del rack, para ponerlo en funcionamiento sólo será necesario conectar un único tomacorriente de alimentación y un sólo patchcord a la red de computadoras de la organización.

Aclaremos que la utilización de racks en la construcción de clusters de computadoras no es para nada nuevo, todos los clusters en la lista de las computadoras más poderosas del mundo emplean racks para alojar sus nodos[top500], pero para ello utilizan hardware especializado (y bastante costoso), como ser gabinetes rackeables² junto a microprocesadores, disipadores y periféricos diseñados específicamente para caber dentro de estos gabinetes.

Nuestro enfoque respecto al rack difiere principalmente en que nuestro objetivo primordial en todos los casos es emplear hardware que se encuentre comúnmente en el mercado local, de esta manera nuestros resultados podrán ser fácilmente recreados por cualquier organización que lo desee. Es por ello, que en lugar de emplear gabinetes rackeables montaremos los componentes de cada par de nodos (motherboard, dispositivos de almacenamiento y fuente de alimentación) sobre bandejas normalizada³ de 19”.

1 Un rack es un gabinete con medidas normalizadas para que sea compatible con equipamiento de diversos fabricante.

2 Gabinetes especiales diseñados para ser colocados dentro de los racks normalizados.

3 Bandejas metálicas que están diseñadas para ser colocadas dentro del racks normalizados.

Otro inconveniente importante es el consumo de energía eléctrica (con su costo y el correspondiente calor disipado), en este sentido seguiremos dos estrategias, la primera será la utilización de memorias CompactFlash (en conjunto con adaptadores CompactFlash - IDE) en lugar de emplear discos rígidos, lectoras de CD-ROM o unidades de disquetes, y la segunda, es el uso de una sola fuente de alimentación cada dos nodos.

Desde el punto de vista de la organización de las actividades, para cumplir con los objetivos antes mencionados, nuestro proyecto se encuentra dividido en tres etapas, encontrándonos actualmente al final de la primera de ellas.

La primera etapa consiste en actividades de relevamiento de las aplicaciones habituales de un cluster, los distintos tipos de clusters, las distintas técnicas de comunicación que emplean, las herramientas de uso y programación que suelen incluir y las distintas distribuciones y herramientas de instalación disponibles para instalar/configurar clusters. En simultáneo con estas actividades se realiza un relevamiento del hardware disponible en el mercado y las distintas alternativas de montaje y construcción de clusters. La finalización de esta primera etapa tendrá como resultado la construcción de un cluster con un número parcial de nodos (3 en primera instancia y 8 al final de la etapa), el que permitirá realizar verificaciones empíricas de nuestros supuestos.

La segunda etapa dará inicio con un cluster modular autocontenido con 8 nodos y utilizando éste, podremos probar y perfeccionar las distintas técnicas y estrategias de instalación y uso del cluster relevadas en la etapa anterior. En paralelo a estas actividades, se llevará a cabo un análisis de los distintos mecanismos disponibles para evaluar la performance de clusters de computadoras, los cuales servirán en la etapa final para evaluar de manera objetiva las técnicas perfeccionadas en la segunda etapa. Además, en el transcurso de la segunda etapa, se agregarán 8 nodos adicionales con lo cual al final de ésta contaremos con un cluster de 16 nodos.

En la tercera y última etapa procederemos a evaluar la efectividad y conveniencia de las técnicas perfeccionadas en la segunda etapa y generar un documento donde se encuentre plasmado el know-how adquirido a lo largo del proyecto. Al finalizar esta etapa se encontrará operando nuestro cluster modular autocontenido con un mínimo de 16 nodos. Igualmente se espera para ese momento, haber agregado otros 8 nodos, llegando al máximo de las capacidades del rack (1 coordinador y 24 nodos esclavos).

3. RESULTADOS Y CONCLUSIONES PRELIMINARES

En este punto los resultados son muy prometedores. Podemos garantizar una reducción del espacio requerido de entre 23 y 30% respecto de un cluster que emplee gabinetes tradicionales en sus nodos⁴, e inclusive una reducción de entre el 25 y el 75% del espacio de rack requerido respecto de un cluster armado utilizando gabinetes rackeables (que además son entre un 50 y un 150% más costosos que los tradicionales), ya que cada gabinete normalizado tiene un alto de entre 2U. (88,9mm.) y 4U. (177,8mm.)⁵, cabe recordar que nuestro cluster utiliza 3U. (133,4 mm.) para alojar cada par de nodos.

Respecto al consumo de energía, nuestros cálculos y observaciones permiten estimar una reducción de entre el 15 y el 25% respecto de un cluster tradicional (cada nodo con su gabinete, fuente de alimentación y una disquetera, disco rígido o lectora de CD-ROM). Esta reducción se debe principalmente al empleo de una fuente de alimentación cada dos nodos y a la utilización de memorias CompactFlash en lugar de disqueteras, discos rígidos o lectoras de CD-ROM.

Recordemos además que una fuente de alimentación no entrega toda la potencia que consume, parte de ella (en fuentes de buena calidad no suele ser inferior al 30%) se pierde [aebischer02], por lo que al usar 12 fuentes de alimentación en lugar de 24 logramos una reducción de las citadas pérdidas. Mediciones de consumo energético realizadas en nuestro cluster experimental dieron como resultado que cada uno de nuestros nodos, tiene un consumo máximo de 71 Watts, por lo que si utilizáramos una sola fuente por cada nodo, ésta se encontraría tan subutilizada

Una memoria CompactFlash (con un adaptador CompactFlash - IDE) consume un 90% menos que una disquetera, un 95% menos que un disco rígido y 97% menos que una lectora de CD-ROM. Por lo que al utilizar éstas últimas en nuestros nodos logramos un ahorro en consumo de energía en dispositivos de almacenamiento del 91% respecto de un cluster donde todos los nodos tengan su propio disco rígido, del 94% respecto de un cluster donde los 24 nodos esclavos tengan lectoras de CD-ROM, y del 84% respecto a un cluster donde los 24 nodos esclavos usen disqueteras.

4 El cálculo se realizó teniendo en cuenta gabinetes mid-tower y sin tener en cuenta el espacio entre gabinetes, ni las estanterías necesarias.

5 Existen gabinetes con una altura inferior a 2U. pero sólo pueden alojar hardware especializado, por lo cual no fueron tenidos en cuenta.

Respecto a costos de adquisición, nuestro Cluster Modular Autocontenido sería un 11% menos costoso que un cluster de 24 gabinetes independientes cada una con su disco rígido, un 2,9% más caro que un cluster con 23 gabinetes independientes con lectoras de CD-ROM además de un coordinador con disco rígido, y un 4,4% más costoso que un cluster compuesto de 23 gabinetes independientes con disqueteras de 3½" y un coordinador con disco rígido.

Cabe resaltar las ventajas de emplear memorias CompactFlash en los nodos esclavos en lugar de discos rígidos (alto costo y mayor consumo de energía), lectoras de CD-ROM (sólo lectura y alto consumo eléctrico) y disqueteras (poca capacidad, baja velocidad y pobre confiabilidad).

Claro está, la mayor ventaja de nuestro enfoque es el hecho de que nuestro cluster es modular (se le podrían anexar otros racks (cada uno con 24/26 nodos adicionales) y además está autocontenido, es decir todos componentes, cables, dispositivos de conectividad, etc. están dentro del rack, por lo cual para hacerlo funcionar sólo es necesario conectar el tomacorriente a la red eléctrica y el nodo coordinador a la red de la organización.

Otra ventaja de emplear un rack es la facilidad con la cual se podría cambiar de ubicación el cluster, e incluso, en nuestro caso, ser llevado al aula (el rack del prototipo cuenta con ruedas).

4. REFERENCIAS

- [aebischer02] B. Aebischer and A. Huser. "Energy efficiency of computer power supply units - Final report" - Swiss Federal Office of Energy, Technical Report 2002.
<http://www.bfe.admin.ch/dokumentation/energieforschung/index.html?lang=en&publication=9436>
- [feng04] W. Feng, Green Destiny: "A 240 Node Energy-Efficient Supercomputer in Five Square Feet", <http://sss.lanl.gov/presentations/041015-IEEE-DVP.pdf>
- [kok01] Kok, J., Elzinga, E., and Wolffe, G. "Notes on constructing a parallel computing platform". J. Comput. Small Coll. 17, 1 (Oct. 2001), 71-80.
- [steiner01] Steiner, S. 2001. "Building and installing a Beowulf cluster". J. Comput. Small Coll. 17, 2 (2001), 78-87.
- [sterling02] Sterling, T. (ed), "Beowulf Cluster Computing with Linux", MIT Press, Cambridge, Massachusetts, 2002.
- [top500] TOP500: <http://www.top500.org/>

DISPOSITIVO DE ALMACENAMIENTO ESTANDAR PARA SOLUCION EMBEBIDA

Di Giulio, Pablo Andrés / Grupo T.D.A. / Departamento de Ingeniería Electrónica / U.T.N. Facultad Regional San Francisco

CONTEXTO

El grupo T.D.A. (Técnicas Digitales Aplicadas) vienen llevando a cabo diversas actividades en lo referente a implementaciones digitales. Actualmente se están llevando a cabo muchos trabajos de I/D en lo que tiene que ver con la interconectividad con la PC, protocolos de comunicación, etc.

RESUMEN

Los dispositivos de almacenamiento masivo constituyen una parte muy importante de cualquier sistema o instalación informática. Hoy en día son la columna vertebral que permite que los diversos sistemas y dispositivo interactúen.

La estandarización en la forma en que se almacena y transporta esta información es muy importante, fundamentalmente a la hora de hacer un desarrollo electrónico propio que utilice un dispositivo de almacenamiento de uso estándar (CD, DVD, Pendrive, MMC, SD, etc.).

El objetivo del presente trabajo de I/D es evaluar los distintos dispositivos de almacenamiento estándar que existen en el mercado y verificar que factibilidad existe de utilizarlo en un desarrollo electrónico propio que utilice un microcontrolador como CPU.

Analizadas las diversas alternativas, observamos que las tarjetas de memorias SD son la que menos recursos del microcontrolador requieren para su manejo. Se desarrollo un pequeño hardware para el manejo físico de dicho dispositivo. Por otra parte se desarrollo una pila (Stack) de software para el acceso al dispositivo además del sistema de archivos que permita comprender la información que está montada sobre el mismo.

Como conclusión, se exponen los recursos mínimos necesarios para poder utilizar este medio de almacenamiento masivo en una aplicación propietaria.

Palabras clave: *Arquitectura de PC, Dispositivo de almacenamiento masivo, Sistema de archivos, Sistema Operativo, Microcontroladores.*

1. INTRODUCCION

El presente trabajo surgió debido a la necesidad del grupo, de utilizar un medio de almacenamiento de información estándar en el mercado para mover información desde un desarrollo electrónico propio a una PC.

Normalmente, la mayoría de los desarrollos electrónicos propietarios, utilizan como controlador un microcontrolador, por ejemplo, control de CNC, datalogger, estación meteorológica, control de potencia, instrumentos de medición, etc.

Un microcontrolador (MCU - Micro Controller Unit) es un circuito integrado que incluye en su interior las tres unidades funcionales de una computadora: CPU, memoria y unidad de entrada/salida, pero con capacidades limitadas. Los MCU se clasifican según su estructura interna en 8-bit, 16-bit y 32-bit, la cual determina el poder del mismo, siendo el MCU de 8-bit el de menor performance y el de 32-bit el de mayor.

A la hora de diseñar un sistema electrónico con MCU es muy importante tener en cuenta la cantidad de recursos que se va a necesitar de este para correr la aplicación, ya que normalmente y en los MCU de línea baja, los recursos son limitados. Los puntos a tener en cuenta, en la elección de un MCU para una determina aplicación son los siguientes:

- Cantidad Memoria Flash
- Cantidad Memoria RAM
- Velocidad de reloj interna
- Tensión de alimentación
- Consumo
- GPIO (General Propouse Input Output) a utilizar.
- Protocolos de comunicaciones (USB, Serial, Ethernet, etc.)

Uno de los puntos a tener en cuenta, en la elección del MCU a utilizar en una determinada aplicación, es si requiere del uso de la PC para algún servicio, por ejemplo visualizar graficas, realizar cálculos de mayor complejidad, realizar backup de la información, etc. Si la aplicación requiere de la PC, debemos evaluar si nuestra aplicación va a estar conectada de forma directa a la PC, a través de un protocolo de comunicación (USB, Serial, Ethernet) o si la aplicación no va a estar conectada todo el tiempo y solamente va a almacenar la información en una memoria y luego descargarla.

Habitualmente, en las aplicaciones donde la PC esta distante, se utiliza un medio de almacenamiento/transporte de información que sea de uso estándar, pudiéndose encontrar repuestos de forma masiva a la hora del recambio y sobre todo ahorrarle trabajo al diseñador del software, ya que el uso de un dispositivo de almacenamiento estándar es soportado por cualquier S.O. (Sistema Operativo).

Por lo tanto la mayor complejidad queda a cargo del desarrollo electrónico, ya que su aplicación, debe acceder a este dispositivo de almacenamiento, poder leer y escribir información de forma que el S.O. de la PC comprenda dicha información. Para poder utilizar el medio de almacenamiento primero debemos acceder al dispositivo y luego montar el sistema de archivos (Filesystem) para poder acceder a la información de forma estructurada y comprenderla.

2. DISPOSITIVOS DE ALMACENAMIENTO MASIVO

Para seleccionar el dispositivo que utilice la menor cantidad de recursos de nuestro MCU realizamos una evaluación de los distintos medios de almacenamiento que existen en el mercado. Evaluamos la capacidad de los mismo, velocidad, hardware adicional requerido, complejidad del acceso, etc.

Como conclusión de la evaluación de cada uno de los dispositivos de almacenamiento, confeccionamos la Tabla 1., que resume las ventajas y desventajas de cada unos de los dispositivos, al ser utilizados en un desarrollo electrónico propio como medio de transporte de información.

	Ventajas	Desventajas
CD-ROM, DVD	Velocidad	Dimensiones físicas grande
	Capacidad de almacenamiento moderada	Complejidad para su manejo Requiere de partes móviles
Disco Rígido	Velocidad	Dimensiones físicas grande
	Capacidad de almacenamiento	Complejidad para su manejo Alto costo
Pendrive	Velocidad	Requiere del manejo de USB Host
	Capacidad de almacenamiento moderada Bajo costo Dimensiones físicas pequeñas	
Tarjetas SD/MMC	Velocidad moderada	
	Bajo costo Capacidad de almacenamiento moderada Dimensiones físicas pequeñas	

Tabla 1: Comparativa de los medios de almacenamiento del mercado hoy en día.

Los pendrive son dispositivo que podrían ser utilizados en este estudio. El principal inconveniente, para su manejo, es que el MCU que

comande a este debe poseer la capacidad de manipular el protocolo USB Host, lo cual no se consigue en todos los MCU. Como podemos ver en la Tabla 1., las tarjetas de memoria tienen las mejores prestaciones, ya que son de bajo costo, fácil acceso y velocidad moderada.

2.1.1 SECURE DIGITAL

Las SD es un formato de tarjeta de memoria flash que se utiliza en dispositivos portátiles tales como cámaras fotográficas digitales, ordenadores PDA y Palm, entre otros.

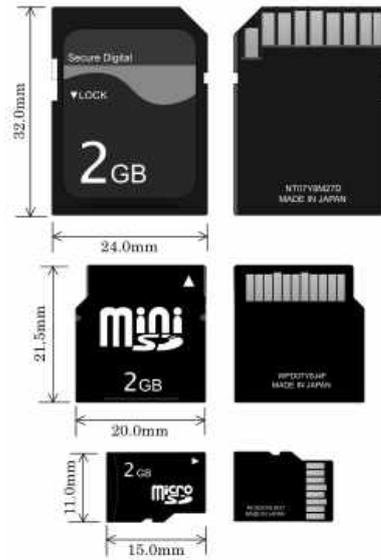


Figura 1: Dimensiones físicas de SD.

Las dimensiones físicas de las tarjetas SD son reducidas, Figura 1, las cuales la hacen muy útiles en aplicaciones donde las dimensiones físicas del producto final son un punto a tener en cuenta. Su costo en el mercado es bajo y existen distintos tipos de zócalos los cuales permiten su contención. Otro punto a favor de dichas memorias es que no requieren de partes móviles a la hora de leerlas o escribirlas, esto es importantísimo cuando la aplicación debe ser montada sobre partes móviles, lo que permite tener menor cantidad de fallas y reajustes del dispositivo. Un punto a favor muy grande además, es que tiene un tipo de interface (protocolo) que es estándar en la mayoría de los MCU grandes y pequeños.

2.1.2 INTERFAZ

Las memorias SD disponen de nueve pines de interface (clock, command, 4 x data, 3 power lines) diseñadas para aplicaciones de niveles de tensión bajos.

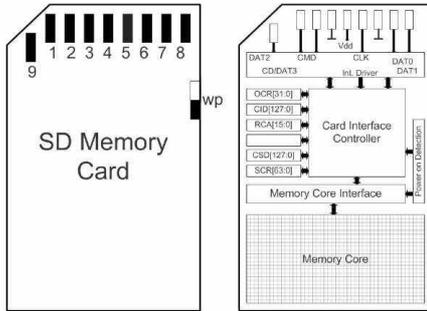


Figura 1: Interface de la tarjeta SD.

2.1.3 TIPOS DE TRANSFERENCIA

Existen 3 modos de transferencia soportados por SD:

- Modo SPI: entrada y salida serial separadas. Requiere de 4 líneas (MISO, MOSI, CLK, CS)
- Modo un-bit SD: comandos separados del canal de datos. Es un tipo de transferencia propietaria.
- Modo cuatro-bit SD: utilizado para transferencia paralelas. Utiliza más pines.

Las tarjetas de baja velocidad soportan tasas de transferencias de 0 a 400 Kbps y modo de transferencia un-bit SD, mientras que las tarjetas de alta velocidad soportan tasas de transferencia de 0 a 100 Mbps en el modo de cuatro-bit, y de 0 a 25 Mbps en el modo un-bit SD.

Dentro de los tipos de transferencia, el modo SPI (Serial Peripheral Interfase) es el más interesante de todos ya que este tipo de bus se encuentra en la mayoría de los MCU de línea baja y no requiere de mucho conocimiento para su uso ya que es muy simple y básico.

2.1.4 MODO SPI

Básicamente el protocolo SPI consiste en el intercambio de información entre el MCU (master) y la tarjeta (slave). Este intercambio se lleva a cabo mediante el envío de comandos por parte del controlador y de respuestas por parte de las tarjetas. Así, en la lectura, el MCU envía el comando de petición de lectura a la tarjeta y esta le envía la respuesta de confirmación seguida del bloque de datos con la información contenida a partir de la dirección solicitada, mientras que el controlador envía mensajes vacíos (tokens) para permitir la salida de datos desde la tarjeta. En la escritura el proceso es parecido, el MCU indica a la tarjeta mediante el comando de escritura que quiere escribir información en una determinada dirección, esta le responde indicando que esta lista y a continuación el MCU envía el bloque de datos a escribir.

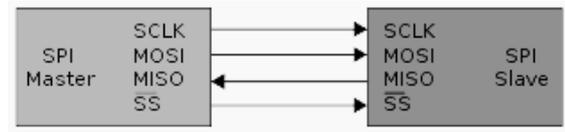


Figura 2: Esquema de comunicación

Las operaciones que no requieren intercambio de datos funcionan de igual forma pero sin usar bloques de datos. En la Figura 2, se esquematiza el modelo de comunicación

2.1.5 COMANDOS

Para realizar transacciones entre el MCU y la tarjeta se deben enviar comandos, los cuales tienen la siguiente estructura:

Byte 1		Byte 2-5	Byte 6
7	6	5-0	7
0	1	Command	0
		Command Argument	1
			CRC

Tabla 2: Frame de comando.

Como podemos observar en la Tabla 2, los primeros dos bits del primer byte son 01 lo que indica que se está enviando un comando. Los comandos están codificados en binario natural como se ilustra en [2]. El último byte (byte 6) es el byte de CRC para la verificación de errores y en realidad en el protocolo SPI no se utiliza, a no ser que en el registro de configuración se especifique que se desea utilizar CRC. Estos son algunos de los principales comandos:

CMD	Argum.	Respu.	Descripción
CMD0	No	R1	Resetea la tarjeta
CMD1	No	R1	Inicializa la tarjeta
CMD9	No	R1	Pide a la tarjeta su información CSD
CMD10	No	R1	Pide a la tarjeta su identificación CID
CMD13	No	R2	Consulta el estado de la tarjeta
CMD16	[31..0] Long. del bloque.	R1	Establece la longitud (en bytes) del bloque para los datos en las operaciones de lectura y escritura.
CMD17	[31..0] Long. del bloque.	R1	Lee un bloque del tamaño indicado por el comando 16.
CMD24	[31..0] Long. del bloque.	R1 R1 R1	Escribe un bloque del tamaño indicado por el comando 16.

Tabla 3: Comandos.

Toda la comunicación entre la tarjeta y el controlador se realiza según el orden del esquema,

de izquierda a derecha, es decir que primero se transmite el bit de más peso (bit 7) del byte 1, y por último el bit de menos peso (bit 0) del byte 6, es decir es una transferencia More Significant Bit First (MSB First).

Las respuestas de la tarjeta son bloques formados por 1 o 2 bytes, dependiendo del tipo de respuesta que se trate. El tipo de respuesta es función del comando, es decir que cada comando tiene asociado un tipo de respuesta.

2.2 SISTEMA DE ARCHIVOS (FILESYSTEM)

Una vez investigado el funcionamiento de la tarjeta de memoria SD, realizamos las pruebas de escritura, lectura y borrado. Con esto ya logramos desarrollar la capa de acceso a la memoria. El siguiente paso es desarrollar el sistema de archivos para poder acceder a la información que está en la memoria de forma coherente.

Desde el punto de vista del usuario de la PC todo se ve tan fácil: se copian archivos de un directorio a otro, se formatea un disquete o se borra un archivo, y todo esto con un solo comando. Pero detrás del telón existe un largo camino hasta que el controlador ha localizado la posición de un archivo o un espacio libre en la memoria que se pueda emplear para guardar un archivo. Ya que las memorias no quieren saber nada de archivos y sub-directorios; sólo conocen de posiciones de memoria; el S.O. (Sistema Operativo) ha de equipar a cada almacenamiento masivo con un sistema de archivos, que haga posible el salto del nivel físico al nivel lógico. Este sistema de archivos se compone de una serie de estructuras de datos, que describe el tamaño de la memoria su contenido.

2.2.1 ESTRUCTURA DEL VOLUMEN

Existen distintos tipos de sistemas de archivo, en nuestro caso, el sistema de archivos implementado es FAT. Cuando se habla de sistemas de archivos, se tiene que tener bien en claro el termino de volumen. Como se detalla en [3] es importante anotar que cada volumen tiene una estructura homogénea, independientemente de si está en un disquete o en un disco duro. El tamaño no tiene ninguna importancia, ya que simplemente afecta al tamaño de las diferentes estructuras de datos, que se necesitan para la gestión del volumen.

Un volumen está estructurado de la siguiente forma:

Sector de arranque
Primera tabla de localización de archivos (FAT)
Una o más copias de la FAT
Directorio raíz (eventualmente con etiqueta de volumen)
Zona de datos para archivos y sub-directorios

Tabla 4: Estructura de un volumen

Además de conocer el significado de volumen, se nombraran otros términos como sectores que son

bloques de memoria que contienen una cierta cantidad de bytes. Dependiendo del sistema de archivos implementado este tamaño puede variar

2.2.2 SECTOR DE ARRANQUE

La estructura del sector de arranque es la primera que se crea luego de formateado el dispositivo de almacenamiento. Contiene todo lo referente al volumen como número de sectores por volumen, cantidad de bytes por volumen. También contiene el llamado Bootstrap-Loader, mediante el cual se puede arrancar el SO. A esta circunstancia es a la que se la llama sector de arranque.

2.2.3 LA TABLA FAT

La zona correspondiente a la FAT es donde se encuentra la administración de espacios. En ella se detallan los archivos existentes y los espacios libres. Cada entrada a la tabla corresponde a un corresponde con un numero de sectores. Estos grupos de sectores se llaman cluster y se define en [3]. Dependiendo del tipo de dispositivo el valor de la cantidad de sectores por cluster puede cambiar Como se ejemplifica en [4] el tamaño de la entrada FAT depende de una cierta cantidad de bit. En nuestro caso la FAT implementada es FAT16 por lo que el tamaño de la entrada a la tabla es de 16-bit, pudiéndose direccionar hasta 65535 clusters.

La tabla FAT además de almacenar si un cluster está ocupado, indica la posición a donde se guardan los archivos. Cada entrada a la FAT tiene la siguiente codificación:

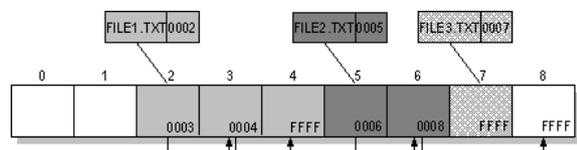


Figura 3: Estructura de la FAT.

Como podemos observar en la Figura 3, las dos primeras posiciones de memoria de la zona de la tabla FAT no se deben escribir, la ubicación de los archivos recién comienza recién en la tercera posición de memoria. El ejemplo se puede observar en [4].

2.2.4 DIRECTORIO RAIZ

Luego de la tabla FAT y su copia viene zona llamada "Directorio Raíz". En él se ubican todas las informaciones relevantes sobre los archivos y subdirectorios allí contenidos.

Cada archivo o sub-directorio posee una estructura o entrada en el directorio raíz, el detalle de esta zona se especifica en [3].

2.2.5 ZONA DE DATOS

Al directorio raíz le sigue la zona de datos cuyo cluster se emplea para guardar archivos y subdirectorios. Para cada cluster en esta zona

naturalmente existe una entrada en la FAT, que se puede modificar mediante la manipulación de archivos en esta zona.

2.3 ENSAMBLE FINAL - LECTURA DE UN ARCHIVO

Una vez estudiado el sistema de archivos FAT, se desarrollaron las estructuras correspondientes a las distintas zonas del volumen. Con dichas estructuras y las librerías desarrolladas para la lectura y escritura de la SD se generaron las funciones para el acceso a la información. En nuestro caso, se realizaron las funciones de lectura de archivos únicamente, quedando para una segunda etapa las rutinas de escritura.

2.3.1 MONTANDO EL SISTEMA DE ARCHIVOS

Antes de realizar la lectura de un archivo se debe montar el sistema de archivo sobre el MCU. En este proceso se extrae la información almacenada en la primera zona del volumen (Sector de arranque). De dicha zona se utilizan los siguientes datos:

- Bytes por sector (offset 0x0B)
- Sectores por cluster (offset 0x0D)
- Entradas al directorio raíz (offset 0x11)
- Numero de sectores por FAT (offset 0x16)

Con dicha información se calculan las posiciones de las demás zonas del volumen.

2.3.2 LECTURA

El proceso de lectura a grandes rasgos es simple. Inicialmente se debe ingresar a la zona directorio raíz, donde se busca a través de las estructuras el nombre del archivo deseado. Una vez ubicada la estructura se extrae la posición que ocupa este en la tabla FAT. Con el valor extraído, se indexa en la tabla FAT y se extrae el offset que hay que aplicarle a la zona de datos para ubicar la información del archivo buscado.

Si el archivo deseado se encuentra en una carpeta, las mismas se encuentran en la zona de datos, pero a diferencia de los archivos, en vez de contener información contienen la estructura de los archivos como las que se encuentran en la zona directorio raíz.

3. RESULTADOS OBTENIDOS

El objetivo central de este trabajo fue investigar y desarrollar una capa de software para ser embebida en un MCU, que permita acceder a un dispositivo de almacenamiento masivo estándar para transportar y almacenar información. Este dispositivo fue seleccionado luego de realizar un análisis del consumo de recursos, que requiere la gran mayoría de los dispositivos de almacenamiento que existe hoy en el mercado.

Las tarjetas de memoria SD son las que fueron seleccionadas, ya que ofrecen las mejores prestaciones,

son pequeñas físicamente, su modo de acceso es simple (SPI – A través de 4 líneas), no poseen partes móviles y su velocidad de acceso es moderada para las aplicaciones que se pueden llegar a desarrollar.

Se desarrolló el código fuente sobre un MCU MC9S08AW60 de la compañía Freescale Inc., [5].

Como resultado se exponen los mínimos recursos que debe disponer el MCU que comanda el desarrollo, para poder comandar un dispositivo estándar como son las tarjetas de memoria SD:

- Tensión de alimentación 3.3V
- Memoria Flash: 2 KBytes.
- Memoria RAM: 600 Bytes.
- Modulo SPI
- Velocidad máxima de SPI CLK 2.5 MHz.
- Compilador ANSI C.

Los resultados anteriores tienen las siguientes consideraciones:

- El código fuente realizado sobre un MCU, cuyo set de instrucciones es del tipo CISC, por lo que el tamaño de la memoria flash especificada como mínima puede variar si el set de instrucciones es del tipo RISC.
- La cantidad de memoria RAM especificada contempla un buffer de datos de 512 bytes. Este se utiliza tanto para lectura como escritura. La transferencia de escritura de la tarjeta requiere del envío de 512 bytes pero en el proceso de lectura de la memoria se puede especificar, a través del CMD16, un tamaño de lectura que puede variar de 1 a 512 bytes. Si el MCU dispone de escasa memoria RAM se puede reducir el tamaño de este buffer.

4. FORMACION DE RECURSOS HUMANOS

En la actualidad existen trabajos similares, muchos de ellos tienen el inconveniente de que no indexan a través de sub-directorios en busca de archivos, por lo que solamente podemos manejar archivos que están en la raíz. Esta opción que se propone, de buscar a través de sub-directorios, es muy interesante ya que el costo en la memoria flash es mínimo al igual que la latencia de lectura. Actualmente estamos trabajando en conjunto con el Laboratorio de DSP de la U.N.C. el cual tiene muchos años ya y nos tendió su mano este proyecto de I/D.

5. BIBLIOGRAFIA

- [1] SD Association. “www.sdorg.org”
- [2] SD Media Format Expand the MAXQ2000's Space for Nonvolatile Data Storage. Application note 3969. Maxim IC.
- [3] Tischer M., Jennrich B., (1996). PC Interno 5, Programación de sistemas. Marcombo.
- [4] NTFS.com. “www.ntfs.com”
- [5] Freescale Inc. “www.freescale.com”

WICC 2009

Análisis de compresión de un sistema de archivos sobre memoria Flash usando OLPC¹

Lic. Matías Zabaljáuregui

Mg. Lia Molinari

Lic. Javier Díaz (Director del LINTI)

{matiasz, Imolinari, javierd}@info.unlp.edu.ar

Calle 50 y 120 – 2do Piso TE: 0221-4223528. FAX: 0221-427-3235

L.I.N.T.I.. - Facultad de Informática – U.N.L.P.

Resumen

Las ventajas de las técnicas de compresión al vuelo (compresión on the fly) se han estudiado desde hace ya una década para su posible utilización en los diferentes subsistemas del kernel de un sistema operativo. Se han planteado soluciones de compresión transparente para el usuario de datos almacenados en memoria o disco y datos enviados por la red, con resultados notables.

Una alternativa para implementar esta idea de manera transparente para el usuario y sin tener que realizar modificaciones importantes al diseño del sistema operativo es incluir funcionalidad extra a un sistema de archivos, de manera tal que cuando el usuario escriba datos al sistema de archivos, éste los comprima justo antes de volcarlos en el disco y cuando el usuario intente recuperar sus datos, el sistema de archivos los descomprima.

Este trabajo estudia las ventajas de aplicar esta última idea a un sistema de archivos especialmente diseñado para memorias Flash, para el proyecto OLPC.

1 Sobre la compresión transparente en sistemas de archivos

Las técnicas de compresión de datos han ayudado a hacer un uso más eficiente del espacio en disco, el ancho de banda de red y otros recursos similares.

La mayoría de las aplicaciones de compresión requieren una acción explícita por parte del usuario para comprimir y descomprimir archivos de datos. Sin embargo, hay algunos sistemas en los cuales la compresión y descompresión de archivos es hecha de manera transparente por el sistema operativo. Un archivo comprimido requiere menos sectores en el disco y además, el tiempo adicional requerido para la compresión y descompresión de archivos de datos puede compensarse con el tiempo ganado al reducir el acceso al disco (uno de los dispositivos más lentos en una computadora).

Existen un conjunto de proyectos que han intentado incluir alguna forma de compresión de datos como funcionalidad transparente del sistema de archivos de un sistema operativo. Estos proyectos, en general, coinciden con las características buscadas en sus diseños: mantenimiento de performance, transparencia al usuario (por ejemplo, el tamaño y otros atributos del archivo visibles para el usuario, deberían permanecer igual que los del archivo sin comprimir), compatibilidad binaria con los programas existentes. Otro importante punto es a qué nivel se realiza la compresión: a nivel de archivo o a nivel de bloque lógico.

¹ OLPC: One Laptop Per Child

La compresión a nivel de archivo lleva a una mayor degradación de la performance cuando se leen y se escriben pequeñas partes del archivo, ya que el archivo entero necesita ser comprimido para cada escritura y necesita ser descomprimido para cada lectura. La ventaja de performance de comprimir datos al nivel de bloque lógico supera ampliamente la dificultad asociada con su implementación, por lo tanto, se suele optar por esta alternativa. Otra de las ventajas de usar bloques lógicos como la unidad de compresión es que si un bloque de disco se corrompe, sólo los datos correspondientes a ese bloque lógico son perdidos. Si el archivo fuera la unidad de compresión, todos los datos del archivo se perderían si se corrompe sólo un bloque. Una desventaja importante de este diseño es la necesidad de modificar una parte del Virtual File System [1] del kernel ya que aunque el tamaño de bloque lógico siempre es un múltiplo entero del bloque físico, el número de bloques físicos requeridos para almacenar un bloque lógico depende de cuanto sea comprimido ese bloque de datos.

Un objetivo deseable de un sistema de archivos comprimido es que soporte múltiples técnicas de compresión dinámicamente. La elección de la técnica particular de compresión podría hacerla el administrador del sistema y el usuario. El administrador del sistema puede elegir la técnica de compresión a ser usada en el momento de montar el sistema de archivos. Por otro lado, los beneficios de la compresión no son significativos en el caso de archivos muy pequeños, por lo que se podría comprimir sólo archivos que tengan un tamaño mayor que un determinado umbral. Valga como ejemplo el caso de los directorios, que, como suelen ser archivos de tamaño pequeño, y más frecuentemente accedidos, podría ser conveniente no comprimirlos.

1.1 Sistemas de archivos estructurados como Log

La idea fundamental de un sistema de archivos estructurado como log es mejorar la performance de escritura almacenando temporalmente todas las modificaciones en el file cache y luego escribiendo todos los cambios secuencialmente en una única operación de transferencia al disco. La información escrita al disco incluye bloques de datos de archivos, atributos de archivos, bloques índices, directorios y casi toda la información necesaria para administrar un sistema de archivos.

Para entender como funciona un sistema de archivos estructurado como log, es necesario explicar dos ideas básicas. La primera relacionada con almacenar el sistema de archivos completo como un árbol, proveyendo de esta forma un único punto de inicio para todo direccionamiento. La segunda involucra el dividir el dispositivo de almacenamiento en segmentos relativamente grandes en los cuales se escriben los nuevos datos.

Generalmente los sistemas de archivos Unix estructuran sus datos como un árbol, con el directorio root como raíz. Sin embargo, ciertos metadatos del sistema de archivos, como la tabla de inodos y el mapa de bloques de disco libres, no son direccionados como parte del árbol. En su lugar, estos elementos tienen ubicaciones fijas en el dispositivo de almacenamiento.

Es posible almacenar los metadatos en archivos especiales para lograr un sistema de archivos completamente direccionado como un árbol y de esta forma el único dato que tiene una ubicación fija en el dispositivo es el puntero a la raíz del árbol. La raíz entonces puede apuntar, posiblemente vía bloques indirectos, al archivo especial que contiene todos los inodos. Uno de estos inodos puede ser el que identifica a un archivo que contiene el mapa de bloques libres, otro inodo puede contener el directorio root, y siguiendo este patrón puede encontrarse todos los datos y metadatos del sistema de archivos.

Teniendo el sistema de archivos estructurado como un árbol, se dispone de gran flexibilidad para almacenar los datos y metadatos en cualquier parte del dispositivo. Como no hay datos con ubicaciones fijas, cualquier cosa puede almacenarse en cualquier lugar del disco. Esto nos lleva al segundo aspecto fundamental de un sistema de archivos estructurado como log: los segmentos.

El dispositivo se divide lógicamente en un número de segmentos relativamente grandes. Cuando se necesita escribir datos al disco, se busca un segmento no usado y se escriben todos los datos junto

con los metadatos apropiados. Todos estos datos son escritos secuencialmente en uno o más segmentos y finalmente la dirección del inodo raíz se almacena en una ubicación fija del dispositivo. Desde ese inodo puede recorrerse completamente el estado actualizado del sistema de archivos.

1.2 Compresión en un sistema de archivos estructurado como log

La compresión transparente de datos no se ha generalizado en los sistemas de archivos convencionales, principalmente por dos razones.

Por un lado, los algoritmos de compresión no proveen una compresión uniforme de todos los datos. Cuando un bloque de un archivo es sobrescrito, los nuevos datos pueden ser comprimidos con otra relación con respecto a los datos anteriores, por lo tanto, el sistema no podría simplemente sobrescribir los bloques originales. Si los nuevos datos comprimidos son más grandes que los anteriores, deben ser escritos en otra ubicación; si son más pequeños, el sistema de archivos puede o bien reusar el espacio original de los datos (desaprovechando parte del mismo) o liberar el espacio y alocar un espacio nuevo más pequeño. En cualquier caso, el espacio de disco tiende a volverse fragmentado, lo cual reduce la efectividad de la compresión.

Por otro lado, los mejores algoritmos de compresión son adaptativos, es decir que usan patrones descubiertos en una parte de un bloque para hacer un trabajo más efectivo en comprimir la información de otras partes.

Estos algoritmos trabajan mejor en bloques de datos de gran tamaño, en lugar de bloques pequeños. La tendencia en la variación en la relación de compresión, aunque para diferentes algoritmos y diferentes datos, suele ser la misma: los bloques más grandes logran mejor relación de compresión [10].

Sin embargo, es difícil lograr que bloques de datos lo suficientemente grandes sean comprimidos a la vez. La mayoría de los sistemas de archivos usan tamaños de bloques que son demasiado pequeños para una compresión de datos efectiva, e incrementar el tamaño del bloque significaría una gran desperdicio de espacio por fragmentación. Además, comprimir múltiples bloques simultáneamente y de manera eficiente resulta difícil ya que los bloques de disco adyacentes no siempre se escriben al mismo tiempo. Comprimir archivos completos tampoco presenta una solución en la práctica ya que en gran parte de los sistemas de archivos la mayoría de los archivos son de sólo algunos kilobytes, además de los problemas de eficiencia mencionados anteriormente.

Afortunadamente, en un sistema de archivos estructurado como log, la estructura de datos principal en el disco es un log escrito secuencialmente.

Todos los datos nuevos, incluyendo las modificaciones a los archivos existentes, son escritos al final del log. Esta técnica fue introducida por Rosenblum y Ousterhout en un sistema de archivos llamado Sprite LFS[9]. El objetivo principal de LFS es proveer performance eliminando las búsquedas de disco en las escrituras.

Este tipo de sistema de archivos es ideal para agregar compresión transparente de los datos. Simplemente se comprime el log a medida que se va escribiendo. Como no se sobrescriben bloques, no es necesario realizar acciones especiales cuando los nuevos datos se comprimen con otra relación y, como los bloques son escritos secuencialmente en el disco, se elimina la fragmentación.

2 DEFLATE/zlib

Entre otros [11], DEFLATE es un algoritmo de compresión de datos sin pérdida que usa una combinación del método LZ77 y la codificación de Huffman. Fue definido originalmente por Phil Katz para la versión 2 de su herramienta PKZIP y luego fue especificado formalmente en la RFC 1951.

zlib[13] es una librería de compresión open-source que surgió como una abstracción del algoritmo DEFLATE utilizado en el programa de compresión de archivos gzip. La primer versión fue publicada en 1995 para ser utilizada en la librería de imágenes libpng.

Existen cientos de aplicaciones para sistemas operativos tipo UNIX que se basan en este algoritmo para comprimir sus datos y su uso en otras plataformas, como Palm OS y otros sistemas embebidos, está creciendo notablemente, ya que su código es portable y tiene un footprint de memoria principal relativamente pequeño.

La compresión se logra a través de dos pasos:

- La búsqueda de strings duplicados y su reemplazo por punteros (pares distancia, tamaño), utilizando el algoritmo basado en diccionario LZ77.
- El reemplazo de los símbolos de salida del proceso anterior con nuevas codificaciones basadas en la frecuencia de uso, utilizando la codificación Huffman.

En DEFLATE se presenta una sutil modificación del método Huffman para evitar tener que almacenar y/o transmitir el árbol generado por el algoritmo. En la variación utilizada por el estándar DEFLATE, hay dos reglas adicionales:

- Los elementos que tengan códigos más cortos se colocan a la izquierda de aquellos con códigos más extensos
- Entre los elementos con códigos de igual longitud, aquellos que están primeros en el orden del conjunto de elementos se ubican a la izquierda.

Cuando se imponen estas dos restricciones en la construcción del árbol de Huffman, existe un único árbol para cada conjunto de elementos y sus respectivas longitudes de código. Estas longitudes son los únicos datos requeridos para reconstruir el árbol y, por lo tanto, lo único que se almacenará y/o transmitirá.

Por otro lado, el compresor de DEFLATE ofrece gran flexibilidad a la hora de comprimir los datos, pues hay tres modos de compresión disponibles.

3 La tecnología Flash y JFFS2

Flash es una tecnología de almacenamiento cuya utilización está creciendo notablemente en sistemas embebidos porque provee almacenamiento basado en estado sólido con alta densidad y confiabilidad, a un costo relativamente bajo.

Las memorias Flash son de tipo no volátil, una característica muy valorada para la gran cantidad de aplicaciones en las que se emplea este tipo de memoria, principalmente pequeños dispositivos basados en el uso de baterías como teléfonos móviles, PDA, pequeños electrodomésticos, cámaras de fotos digitales, reproductores portátiles de audio, etc. [2] [3]

Los objetivos del diseño de Journaling Flash File System (JFFS) [12] [8] están determinados en gran medida por las características de la tecnología Flash y de los dispositivos en los que se supone que será utilizado. Como los dispositivos embebidos y alimentados por baterías usualmente son tratados por los usuarios como simples electrodomésticos o aparatos electrónicos, es necesario asegurar una operación confiable frente a interrupciones en la alimentación eléctrica.

El JFFS original es un sistema de archivos basado en logs puro.

4 Las pruebas realizadas con OLPC

Las pruebas de performance fueron realizadas sobre una plataforma embebida que utiliza un chip Flash NAND de 512 MB como dispositivo de almacenamiento secundario. El hardware de prueba consiste en la placa base de la primer generación de prototipos del proyecto One Laptop Per Child gestado en Media Labs del MIT[7].

Las pruebas consistieron en medir el tiempo que consumen las operaciones de lectura y escritura de bloques de datos en la memoria Flash de la OLPC utilizando JFFS2 con y sin compresión, y en calcular las latencias introducidas por las funciones de compresión y descompresión implementadas en el kernel. De esta forma, se logró demostrar que aunque la compresión introduce una latencia no

despreciable, el tiempo final de la operación de escritura resulta menor que en el caso de no utilizar compresión.

Luego de analizar distintas herramientas de benchmarks para sistemas de archivos, se realizó un primer juego de pruebas con aquella que ofrecía la mayor flexibilidad para medir latencias en las operaciones de lectura y escritura de bloques de datos.

Iozone[6] es un instrumento open-source que permite medir el rendimiento y latencia en varias de las operaciones posibles sobre un sistema de archivos. Sin embargo, al realizar las pruebas se alcanzaban relaciones de compresión que excedían los promedios normales de los algoritmos de compresión utilizados. Al estudiar el código fuente de Iozone, se pudo comprobar que éste genera los datos a ser escritos en el sistema de archivos utilizando un único patrón de caracteres que se repite hasta completar un archivo del tamaño deseado por el usuario.

Para las pruebas realizadas para este trabajo, se escribieron dos programas específicos que miden el tiempo consumido por cada operación de bloque escrito o leído. [4] [5]

Por otro lado, para estudiar la latencia introducida por las funciones de compresión y descompresión, en primer instancia se buscaron mecanismos ofrecidos por el kernel linux para medir tiempos. Fue necesario realizar la medición en espacio de kernel para poder aislar únicamente la latencia de las funciones que implementan la compresión y descompresión de los datos, sin incluir el tiempo consumido en otras tareas realizadas por las system calls read() y write().

5 Los resultados obtenidos

Aunque se disponen de los datos detallados para la transferencia de cada bloque, a continuación se realiza el análisis estudiando las latencias promedios para minimizar el efecto del read-ahead en el caso de las lecturas y la activación del garbage collector de jffs2 en el caso de las escrituras.

Como se puede observar en el Cuadro 1, en el caso de la configuración de jffs2 sin comprimir, el tiempo promedio que insume escribir un bloque de 4096 bytes es de 8240 microsegundos, mientras que la latencia producida al invocar a la función jffs2 compress(), es de 1,5 microsegundos. Como en este caso no se están comprimiendo los datos, la invocación a la función virtual de compresión simplemente retorna sin realizar ninguna acción. Por lo tanto, los 8240 microsegundos incluyen el tiempo de ejecución de la system call más el tiempo de acceso y transferencia de datos al dispositivo Flash.

Si se analiza el caso de la escritura en el sistema de archivos jffs2 con compresión zlib, se verifica que aunque se incrementa la latencia promedio de la función de compresión a 2195 microsegundos, el tiempo promedio total de la system call es de 5690. Es decir se logra una reducción significativa en el tiempo de escritura del dispositivo Flash.

Cuadro 1: Resultados de las pruebas

	write()	compress()	read()	decompress()
sin compresión	8240,17	1,44	1735,98	–
compresión zlib	5689,69	2194,88	2390,05	633,51

Sin embargo, deben destacarse dos hechos que se manifiestan en las pruebas y comprueban algunas afirmaciones que suelen aparecer en la literatura relacionada. Por un lado, se puede verificar que el algoritmo zlib es varias veces más rápido al descomprimir que al comprimir. compress() introdujo una latencia promedio de 2194,88 microsegundos por bloque mientras que decompress() sólo agregó 633,51 de tiempo promedio a la lectura de un bloque. Esta característica puede ser importante a la hora de elegir un sistema de archivos para un sistema embebido. Por otro lado, muchos benchmarks para sistemas de archivos suponen que todo lo que se lee del dispositivo debió ser escrito en algún momento, con lo cual plantean el tiempo combinado de escritura y lectura como

el indicador más importante. Por lo tanto, si se analiza la performance general, teniendo en cuenta el tiempo que consumen las operaciones de lectura y escritura en conjunto, se llega a la conclusión de que el uso de compresión incrementa la eficiencia del sistema de archivos y el dispositivo.

6 Conclusiones

El mundo de los sistemas embebidos ha venido creciendo de manera sostenida en los últimos años y, en este contexto, el futuro de la tecnología Flash es realmente alentador. Ya que se tiende a la ubicuidad de las computadoras y electrodomésticos inteligentes e integrados, se presenta la necesidad de contar con memorias pequeñas, baratas y con grandes capacidades de almacenamiento.

Además de sus otras ventajas, los sistemas de archivos estructurados como logs ofrecen una oportunidad en la cual la compresión de datos puede ser utilizada sin los problemas de asignación y fragmentación que se presentan en las aproximaciones más convencionales.

Utilizando una técnica de compresión suficientemente rápida, el tiempo utilizado en la compresión y descompresión se compensa por el tiempo ganado por la menor cantidad de accesos al dispositivo físico.

Claramente, esta afirmación está condicionada por la relación entre la velocidad de la CPU y la velocidad del controlador del dispositivo de almacenamiento. Sin embargo, ya que las velocidades de CPU están creciendo más rápido, en términos relativos, que las velocidades de los discos, se puede esperar que la compresión al vuelo disminuya la latencia y mejore el rendimiento de los sistemas de almacenamiento.

7 Referencias

- [1] Daniel P. Bonet and Marco Cesati. Understanding Linux Kernel 3rd Edition. O'Reilly, 2005.
- [2] Intel Corporation. Understanding the Flash Translation Layer (FTL) Specification.
<http://developer.intel.com/design/flcomp/applnots/297816.htm>.
- [3] Linux Devices. Linux Devices. <http://www.linuxdevices.com>.
- [4] GNU. The GNU C Library Reference Manual. www.gnu.org.
- [5] Intel. Using the RDTSC Instruction for Performance Monitoring.
<http://developer.intel.com>.
- [6] iozone. iozone. <http://www.iozone.org>.
- [7] MIT. One Laptop Per Child. <http://wiki.laptop.org>.
- [8] Inc Red Hat. eCos – Embedded Configurable Operating System.
<http://sources.redhat.com/ecos/>.
- [9] Mendel Rosenblum and John K. Ousterhout. The Design and Implementation of a Log-Structured File System. ACM Transactions on Computer Systems.
26
- [10] Matthew Simpson. Analysis of Compression Algorithms for Program Data. University of Maryland.
- [11] Paul R. Wilson, Scott F. Kaplan, and Yannis Smaragdakis. The Case for Compressed Caching in Virtual Memory Systems. Dept. of Computer Sciences, University of Texas, Austin.
- [12] David Woodhouse. The Journalling Flash File System. ACM Transactions on Computer Systems.
- [13] Zlib. Informacion Tecnica sobre zlib. <http://www.zlib.net>.

EVALUACION Y SIMULACION DEL RENDIMIENTO DE LOS PROTOCOLOS DE RUTEO PARA MANET BAJO RESTRICCIONES DE QoS

Maria A. Murazzo, Nelson R. Rodríguez, Matías Martínez
maritemurazzo@gmail.com, nelson@iinfo.unsj.edu.ar, matiasmartinezinfo@gmail.com

*Universidad Nacional de San Juan – Facultad de Ciencias. Exactas, Físicas y Naturales -
Departamento de Informática*

CONTEXTO

El presente trabajo se encuadra dentro del área de I/D Arquitectura, Redes y Sistemas Operativos y se enmarca dentro del proyecto de investigación Arquitectura de acceso a servicios Web desde dispositivos móviles heterogéneos (21 E 830) donde se está trabajando en la línea de investigación de QoS en redes móviles como soporte para el desarrollo de aplicaciones ubicuas.

RESUMEN

Las redes inalámbricas han tenido un gran impacto en la industria y en los ámbitos académicos. Las MANETs (Mobile Ad-hoc NETWORK) agregan a las limitaciones de las redes inalámbricas características de movilidad, debido a esto, es muy difícil mantener adecuados niveles de QoS (Quality of Service).

Para poder realizar el análisis de parámetros de QoS, se propuso trabajar con protocolos de ruteo reactivos debido a su menor uso de recursos y su mejor adaptabilidad a la movilidad.

Por lo tanto se decidió trabajar con los protocolos AODV (Ad hoc On-Demand Distance Vector) y DSR (Dynamic Source Routing), y realizar el análisis de parámetros de QoS bajo diferentes escenarios con el objetivo de evaluar su comportamiento.

Palabras clave: QoS, MANET, AODV, DSR, NS-2

1. INTRODUCCION

Una red móvil ad hoc (MANET) es una red de comunicaciones formada espontáneamente por un conjunto de dispositivos móviles inalámbricos que son capaces de comunicarse

entre sí, sin la necesidad de una infraestructura de red fija o gestión administrativa centralizada.

Debido a que el alcance de transmisión de los dispositivos es limitado, pueden llegar a ser necesarios nodos intermedios para transferir datos de un nodo a otro. Por ello, en una red MANET cada nodo puede operar como fuente, destino o router (naturaleza “multihop”).

Los nodos son libres para moverse arbitrariamente, produciendo cambios en la topología de la red. El grado de movilidad y cambio de la topología depende de las características de los nodos [1]. Además, las variaciones en el canal de radio y las limitaciones de energía de los nodos pueden producir cambios en la topología y en la conectividad. Por lo que, las MANETs deben adaptarse dinámicamente para ser capaces de mantener las conexiones activas a pesar de estos cambios.

Estas redes son principalmente útiles en ambientes militares y de rescate, donde hay necesidad de comunicación ubicua. El poder ofrecer niveles de QoS en MANETs sigue siendo un tema abierto para la comunidad investigadora, y supone un reto interesante, dadas las dificultades que conlleva [2].

En tal sentido es necesario dotar a las MANET de la posibilidad de proporcionar niveles adecuados de QoS a las aplicaciones que lo requieran, como la voz (por ejemplo, VoIP) y el video (por ejemplo, video-streaming). Para lograr esto es necesario realizar la evaluación de múltiples parámetros tales como ancho de banda mínimo esperado, porcentaje máximo de pérdidas de datos, retardo máximo y variación del retardo máxima (jitter).

Si bien este problema está prácticamente resuelto en redes fijas, las características de las redes MANETs hacen necesario un nuevo estudio para afrontar este problema.

La topología dinámica, la naturaleza multihop y los escasos recursos de los nodos hacen necesario que los mecanismos de provisión de QoS sean lo más ligeros posibles, en cuanto a carga de procesamiento (CPU), como de recursos de red (ancho de banda), para evitar que el throughput o capacidad disponible por nodo se reduzca drásticamente [3].

Con el objetivo de permitir el soporte de QoS en redes MANETs es necesario que se definan cuales son las métricas que se deben tener en cuenta. La QoS se define como un conjunto de requerimientos que deben ser satisfechos por la red; y la red espera garantizar un conjunto de atributos predefinidos por los usuarios, en términos de performance extremo a extremo.

Estos parámetros pueden ser retardo, ancho de banda, probabilidad de pérdida de paquetes y variaciones de retardo; en las redes MANET también se consideran importantes el consumo de energía y el área de cobertura de los nodos.

Lo más importante a tener en cuenta en las MANET a la hora de proveer QoS, es que la dinamicidad de su topología, modifica constantemente los nodos vecinos, así como el estado de los enlaces, provocando fluctuaciones en los niveles de recursos, esto conduce a que la provisión de QoS es compleja en este tipo de redes.

Con el objeto de encontrar una solución óptima a los problemas planteados es que se decidió trabajar la QoS como un problema exclusivo de la capa de red. En tal sentido se optó por el análisis de los protocolos de ruteo de MANET convencionales [4] y no los que poseen soporte de QoS.

Esto se debe a que los protocolos de ruteo con QoS deben trabajar junto con la señalización de QoS [5] para establecer caminos, a través de la red, que reúnan los requerimientos de

QoS extremo a extremo, lo que provoca desperdicio de ancho de banda.

De manera general, los protocolos de ruteo para redes MANET se clasifican en dos categorías principales: proactivos y reactivos.

Los protocolos proactivos mantienen tablas que almacenan la información de ruteo y periódicamente, disparan un mecanismo de propagación de actualización a través de la red, con el fin de conocer el estado de la red. Esto puede causar una cantidad importante de paquetes de señalización que afecte el ancho de banda, el throughput y el consumo de energía. Dicha utilización de recursos no siempre es deseable porque también se realiza para caminos que nunca son utilizados.

La ventaja es que las rutas a cada destino están siempre disponibles, pero la principal desventaja es la escalabilidad para funcionar adecuadamente cuando en la red presenta una alta tasa de movilidad o cuando hay un gran número de nodos en la red.

Ejemplo de protocolo de esta categoría son: OLSR (Optimized Link-State Routing) [7] y DSDV (Dynamic Destination Sequenced Distance Vector) [8].

Como solución a estos inconvenientes aparecen, los protocolos bajo demanda o reactivos, que se caracterizan por iniciar un mecanismo de descubrimiento de ruta cuando una fuente necesita comunicarse con un destino al cual no sabe cómo llegar. De manera general, el ruteo por demanda requiere menos overhead que los proactivos, pero incurre en un retraso de descubrimiento de ruta cada vez que se requiere un nuevo camino.

Las diferencias entre los protocolos por demanda están en la implementación del mecanismo de descubrimiento de ruta y en las optimizaciones del mismo.

Ejemplos de protocolos reactivos son: DSR (Dynamic Source Routing) [9] y AODV (Ad hoc On-Demand Distance Vector) [10].

Seguridad y Privacidad en la Plataforma Android

M. Capobianco

A. G. Stankevicius

J. Echaiz

Grupo de Investigación en Software Libre
Dpto. de Cs. e Ing. de la Computación – Universidad Nacional del Sur
Tel.: 0291-4595135 – Av. Alem 1253, Bahía Blanca
{mc, ags, je}@cs.uns.edu.ar

Resumen

Conforme crece el uso y la adopción de los dispositivos móviles, se espera que brinden funcionalidades cada vez más complejas. En consecuencia, la demanda de nuevas aplicaciones aumenta considerablemente. Android es una nueva plataforma licenciada como software libre que será clave para cubrir esta demanda, proveyendo un sistema operativo, middleware y las herramientas necesarias para el desarrollo de aplicaciones por medio del lenguaje Java. En este proyecto analizaremos la seguridad en la plataforma Android y determinaremos si esta permite salvaguardar la privacidad de los usuarios. En base a este análisis, propondremos soluciones a los problemas identificados, colaborando con el refinamiento de la plataforma.

Palabras clave: Frameworks de Desarrollo, Seguridad en Sistemas, Tecnologías Móviles.

1. Contexto

La línea de investigación propuesta en este documento fue presentada como un nuevo Proyecto de Grupo de Investigación (PGI) en la secretaría de Ciencia y Tecnología de la Universidad Nacional del Sur durante la convocatoria 2009 y se encuentra actualmente en proceso de evaluación.

2. Introducción

La importancia de los dispositivos móviles ha crecido en forma considerable en los últimos diez años, llegando estos a ocupar un rol preponderante en nuestras vidas. Hoy en día nos resulta difícil imaginarnos como comunicarnos sin nuestro teléfono celular o como organizar nuestro día sin nuestro PDA.

Conforme crece el uso y la adopción de estos dispositivos, los usuarios esperan obtener funcionalidades cada vez más complejas de los mismos [12, 11]. En el último congreso mundial sobre dispositivos móviles realizado este año en la ciudad de Barcelona, se publicitaron equipos que son capaces de brindar televisión móvil, pantallas sensibles al tacto y navegación por Internet a alta velocidad entre otras cosas. Una de las tendencias que se evidenció en el mencionado congreso fue la creciente importancia que le están dando los fabricantes no solo al diseño y a las funciones de los teléfonos, sino a los servicios y los contenidos que estos brindan. Los dispositivos móviles de última generación presentan una variada oferta de contenido multimedia, como juegos, cine y hasta televisión, que ahora se puede transmitir al teléfono de forma fluida gracias a la mayor capacidad de las redes celulares de tercera generación (3G).

En este contexto, las aplicaciones que corran sobre estos dispositivos móviles tendrán un rol cada vez más importante. En este sentido nos encontramos en la actualidad en un punto de inflexión, dado que las aplicaciones de los teléfonos celulares se venden junto con el hardware, tal como sucedía con las primeras computadoras, en donde el software no se vendía por separado sino que era ofrecido por el fabricante del hardware junto con la adquisición de la computadora. Pero esto está cambiando rápidamente y en un futuro cercano, cada usuario de un teléfono celular será capaz de personalizar el software del mismo incluyendo en éste las aplicaciones que le resulten más interesantes y que estén estandarizadas para su región geográfica independientemente del modelo de teléfono que haya comprado. Por consiguiente, la demanda de aplicaciones de calidad para dispositivos móviles aumentará en forma drástica.

Para satisfacer esta demanda será necesario disponer de un framework de desarrollo específico para este tipo de aplicaciones [7, 12], por lo que Google comenzó en el año 2005 a desarrollar el proyecto Android, que se consolidó en el año 2007 cuando el Open Handset Alliance (OHA), un consorcio de varias compañías que incluyen a Google, HTC, Intel, Motorola, Qualcomm, T-Mobile, Sprint Nextel y nVidia, anunció su objetivo de crear estándares abiertos para dispositivos móviles. El primer producto de OHA fue Android, una plataforma para dispositivos móviles construida sobre el kernel GNU/Linux. Hoy en día Android incluye un sistema operativo, middleware y aplicaciones clave para la operación de los dispositivos móviles. La versión beta Android para el desarrollo de aplicaciones provee las herramientas y librerías necesarias para programar aplicaciones para esta plataforma por medio del lenguaje de programación Java. Recientemente, en octubre de 2008, la plataforma Android fue licenciada como software libre. De esta manera Google busca obtener otro elemento clave para poder satisfacer la creciente demanda de aplicaciones en dispositivos móviles: una comunidad de desarrolladores que trabaje sobre la plataforma Android y que brinde además soporte a los usuarios de estas aplicaciones.

Al estar licenciado como software libre, Android garantiza eficiencia, calidad y bajo costo [19]. El término software libre hace referencia a las cuatro libertades que están garantizadas a quien recibe este tipo de software [22]:

1. Libertad para ejecutar el programa en cualquier sitio, con cualquier propósito y para siempre.
2. Libertad para estudiarlo y adaptarlo a nuestras necesidades.
3. Libertad de redistribución, de modo que se nos permita colaborar con vecinos y amigos.
4. Libertad para mejorar el programa y publicar las mejoras.

Existen numerosos beneficios, tanto para los usuarios como para los desarrolladores, que devienen de la adopción del modelo del software libre. La principal ventaja producto de contar con las cuatro libertades es la calidad que se obtiene a partir de la colaboración voluntaria de los desarrolladores y/o usuarios que contribuye a la comunidad formada en torno al proyecto, descubriendo, reportando o incluso corrigiendo errores en entornos y situaciones de naturaleza muy diversa. Como mencionamos anteriormente, el contar con esta comunidad es seguramente una de las razones por las cuales Android fue licenciado como software libre. Otra de las consecuencias del modelo es que si un programa no ofrece la calidad suficiente, la competencia puede tomarlo y mejorarlo, partiendo del estado actual y realizando lo que se conoce como un fork, para obtener así un mejor producto. De esta forma la colaboración y la competencia, dos poderosos mecanismos de refinamiento presentes en la naturaleza, se combinan para obtener software de mejor calidad. Por otra parte los costes de desarrollo disminuyen, a partir de la reutilización de software existente que se ve posibilitada por el acceso al código fuente y a la historia de cada proyecto que nos brindan los repositorios de software libre.

Teniendo en cuenta las razones expuestas creemos que la plataforma Android revolucionará el mercado de los dispositivos móviles en un futuro cercano. Es por esto que consideramos que estamos en un momento propicio para centrar la investigación en la plataforma Android. Ya existe en la actualidad un conjunto de aplicaciones más que interesantes desarrolladas sobre esta plataforma, como por ejemplo:

Cab4me: software que utiliza la versión de google maps que corre bajo Android y permite llamar a un taxi mediante un solo clic de un botón. Cab4me usa los servicios de GPS para seleccionar en forma automática la compañía de taxi más cercana al usuario.

Locale: una aplicación que hace que el teléfono celular se pase automáticamente a modo vibrador cuando el usuario entra a la sala de conferencias, para evitar las molestias cuando el teléfono suena en forma inesperada en una reunión. Esta aplicación puede además determinar si el usuario se encuentra en su casa o en la oficina desviando las llamadas al teléfono fijo que corresponda según el caso.

Life360: una aplicación para crear una red social entre los vecinos del barrio que permite por ejemplo recibir las emergencias que sucedieron en el vecindario y enviar alertas a todas las personas en el área.

goCart: software que asiste a los compradores para que estos puedan tomar decisiones inteligentes e informadas. Utilizando el GPS y la cámara del celular para hacer un scan de los códigos de barra del producto, goCart busca en la Web de las tiendas locales para comparar los precios del producto en distintos lugares. Es también posible programar alertas para que el sistema nos avise cuando el precio de un producto que nos interesa desciende por debajo de un límite dado.

3. Líneas de investigación y desarrollo

El propósito de esta línea de trabajo es desarrollar el área de investigación en dispositivos móviles que involucra tecnología de última generación. Esta disciplina se encuentra actualmente poco avanzada en nuestra universidad en particular y en nuestro país en general. Es también nuestro objetivo servir de marco para la formación de recursos humanos mediante el desarrollo de tesis de grado y/o de posgrado en el área, tarea que ya se encuentra bajo desarrollo.

El objetivo general del proyecto es investigar como protege Android la seguridad y la privacidad de los usuarios [16, 18, 15]. Buscamos realizar un estudio exhaustivo de estos aspectos a fin de hallar los problemas que puedan existir en este sentido y proponer soluciones para los mismos.

Los objetivos específicos son:

- Crear aplicaciones y test beds para la plataforma Android
- Proponer políticas de seguridad apropiadas para entornos móviles
- Aplicar técnicas de seguridad en sistemas para implementar estas políticas y disminuir la cantidad de ataques
- Usar técnicas de inteligencia artificial para desarrollar algoritmos inteligentes que permitan detectar soluciones de seguridad y que puedan ser aplicados a este contexto (como por ejemplo algoritmos de detección de patrones).

En particular nuestra hipótesis de partida para este proyecto es que la seguridad y la privacidad en Android pueden ser mejoradas sustancialmente, estableciendo políticas de seguridad adecuadas que combinen técnicas como firewalls, sistemas de detección de intrusos y sistemas que permitan realizar el análisis de amenazas en forma automática para mejorar la usabilidad de las aplicaciones y simplificar la tarea de los usuarios, para quienes las herramientas de seguridad deberían ser transparentes. El trabajo de investigación a desarrollar se guiará por la metodología utilizada tradicionalmente en las ciencias de la computación, siguiendo los principios del método científico usados en el área de las ciencias básicas.

4. Resultados Esperados

En este momento el proyecto se encuentra dando sus primeros pasos. Para conseguir los objetivos de este proyecto comenzaremos en primer lugar por desarrollar un análisis del estado del arte de la plataforma Android, aprovechando el licenciamiento de la plataforma como software libre, que nos permitirá acceder al código fuente del proyecto y además participar de las listas de mensajería y foros asociados al mismo. De esta manera podremos acceder a información de primera mano sobre Android y su estado de desarrollo. Por otra parte esto permitirá colaborar con la construcción y depuración del software en cuestión. Una vez familiarizados con el sistema se realizarán auditorías del mismo en relación a la seguridad y la privacidad de los usuarios, siguiendo lineamientos internacionales en la disciplina, como las normas ISO 27001. Una vez analizado el sistema se espera construir soluciones para los problemas existentes y probar las soluciones diseñadas, para lo cual se desarrollarán aplicaciones bajo la plataforma Android que se harán luego disponibles a la comunidad de software libre licenciándolos adecuadamente.

Para el desarrollo de las aplicaciones se utilizará una metodología de desarrollo ágil (en particular Extreme Programming), complementada con Scrum como metodología de gestión. Durante el todo el desarrollo del proyecto se publicarán los resultados preliminares obtenidos en congresos relacionados con la especialidad. Al finalizar el proyecto se espera reportar los resultados finales en un artículo que será enviado a una revista especializada.

5. Formación de recursos humanos

El proyecto ha ya realizado un aporte significativa a la misión educativa de nuestra unidad académica. Nuestra tarea de investigación contribuye a la formación de estudiantes de grado y posgrado que se encuentran realizando sus tesis en esta área o en temas relacionados. Los temas de vanguardia que están siendo abordados en esta línea se integrarán a la brevedad el programa de materias optativas y cursos de especialización del Departamento de Ciencias e Ingeniería de la Computación. Se espera además que varios estudiantes de grado y posgrado, y en particular becarios, se sumen a este proyecto en un futuro próximo.

Cabe destacar que la temática del proyecto involucra la interrelación de distintas áreas como ingeniería de software, sistemas operativos y seguridad en sistemas. Por lo tanto crea un espacio de trabajo común para distintos grupos de trabajo establecidos dentro del departamento: el Laboratorio de Investigación en Sistemas Distribuidos, el Grupo de Investigación en Software Libre y el Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información.

Referencias

- [1] ABLESON, F., COLLINS, C., AND SEN, R. *Unlocking Android*. Manning Publications, 2009.

- [2] BELLOVIN, S. M. Security problems in the TCP/IP protocol suite. *Computer Communication Review* 19, 2 (Apr. 1989).
- [3] BURNETTE, E. *Hello Android: Introducing Google's Mobile Development Platform*. The pragmatic programmer, 2008.
- [4] DENNING, D. E. R. *Cryptography and Data Security*. Addison-Wesley, Reading, Massachusetts, 1982.
- [5] GUIMARAES, L. Confidentiality, integrity and high availability with open source IT green. *ArXiv e-prints* (2008).
- [6] HASEMAN, C. *Android Essentials (Firstpress)*. APress, 2008.
- [7] HASHIMI, S., AND KOMATINENI, S. *Pro Android: Developing Mobile Applications for G1 and Other Google Phones*. APress.
- [8] HILKER, M. Distributed self management for distributed security systems. In *Proceedings of the 2nd International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2007)* (Zhengzhou, China, 2007).
- [9] JAEGER, E., AND HARDIN, T. A few remarks about formal development of secure systems. *ArXiv e-prints* (2009).
- [10] LANDAU, S. Security and privacy landscape in emerging technologies. *IEEE Security & Privacy* 6, 4 (2008), 74–77.
- [11] MARZIO, J. D. *Android, a programmers guide*. McGraw-Hill, 2008.
- [12] MEIER, R. *Professional Android Application Development*. Wiley Publishing, Indianapolis, Indiana, 2009.
- [13] MILLEN, J. K. Models of multilevel computer security. In *Advances in Computers*, M. C. Yovits, Ed., vol. 29. Academic Press, Inc., 1988.
- [14] MURPHY, M. *The Busy Coder's Guide to Android Development*. Commonsware, 2009.
- [15] PERRY, W. E. Developing a computer security and control strategy. In *Advances in Computer System Security*, R. Turn, Ed., vol. II. Artech House, 1984, pp. 39–48.
- [16] PFLEEGER, C. P., AND PFLEEGER, S. L. *Security in computing*, fourth ed. Prentice-Hall, Inc., pub-PH:adr, 2007.
- [17] PFLEEGER, C. P., PFLEEGER, S. L., AND THEOFANOS, M. F. A methodology for penetration testing. *Computers & Security* 8, 7 (1989), 613–620.
- [18] RAMASWAMI, R. Placement of data integrity security services in open systems interconnection architecture. *Computers & Security* 8, 6 (1989), 507–516.
- [19] RAYMOND, E. *The cathedral and the bazaar*. O'Reilly publishings, 2001.
- [20] ROBSON, B. People and technology: The security management challenge. In *1990 Decus Europe Symposium* (Cannes, France, Sept. 1990).
- [21] ROGERS, R., LOMBARDO, J., MEDNIEKS, Z., AND (AUTHOR), B. M. *Android Application Development: Programming with the Google SDK*. Commonsware, O'Reilly Media.
- [22] STALLMAN, R. M. The gnu manifesto. *Dr. Dobb's Journal of Software Tools* 10, 3 (1985), 30–35.
- [23] TEDESCO, G., AND AICKELIN, U. Strategic alert throttling for intrusion detection systems. In *The 4th WSEAS International Conference on Information Security* (Tenerife, Spain, 2005).

CONVERSOR DIGITAL-RESOLVER PARA DIGITALIZACIÓN DE SERVISISTEMAS INDUCTIVOS

H. Nelson Acosta, José A. Marone y M. Carolina Tommasi

nacosta@exa.unicen.edu.ar, josemarone, carolinatommasi}@gmail.com

INCA-INTIA / Facultad de Ciencias Exactas / UNCPBA

Tandil - Provincia de Buenos Aires - Argentina

CONTEXTO

El grupo INCA trabaja en diferentes líneas de investigación tales como: procesamiento digital de señales, búsqueda de patrones en imágenes, sistemas de tiempo real, sistemas de navegación y sistemas embebidos. Este trabajo integra estos conocimientos, para el desarrollo de un sistema embebido que permita la conversión de señales digitales en analógicas, a través de algoritmos de procesamiento de señales en tiempo real.

RESUMEN

Este trabajo propone el desarrollo de un sistema embebido que permita la conversión de información digital a señales sincrónicas. El sistema se diseña manteniendo las interfaces y dimensiones de los convertidores similares ya existentes, siendo apto como reemplazo de éstos de forma transparente. Por otro lado, se establecen parámetros de configuración del convertidor, para que futuros desarrolladores puedan ajustar la calidad de la señal obtenida en función de las características de la unidad de proceso seleccionada. Con el desarrollo propuesto se busca contar con un convertidor de bajo costo que pueda ser utilizado como prototipo en la implementación de cadenas sincrónicas, logrando la síntesis digital de una señal sincrónica de alta resolución.

Palabras clave: *Sistemas Embebidos, AVR, Microcontroladores, Procesamiento de señales, Servosistemas Inductivos, Tiempo Real.*

1. INTRODUCCIÓN

La expansión tecnológica en el área industrial, militar y robótica ha crecido vertiginosamente a partir de la segunda mitad del siglo XX, dichas áreas están profundamente ligadas con la *teoría de control de sistemas* y son, en mayor o menor medida, dependientes de control de la posición mecánica de diversas piezas. Dichos sistemas son denominados “*Sistemas de control por realimentación*”, Un caso particular de estos sistemas son los “*Servosistemas*” siendo los *inductivos* los más utilizados en el área militar y robótica. Estos sistemas de control, están basados en motores de corriente alterna que transmiten la información del ángulo de rotación, modulando en amplitud una onda senoidal de periodo fijo denominada “*portadora*” [NAI06].

En la actualidad, los servosistemas están siendo digitalizados, consecuentemente ha surgido la necesidad de convertir señales sincrónicas en información digital y viceversa.

El dispositivo a desarrollar permite convertir el valor de un ángulo en formato digital, a su homólogo en formato resolver [NAI06, DDC94]. El módulo resultante puede ser utilizado como prototipo en la implementación de sistemas o para reemplazar a otros de igual funcionalidad en algún sistema de control [DOR04].

El proyecto no tiene requerimientos en cuanto a las tecnologías y/o técnicas a utilizar para el desarrollo; sin embargo, el trabajo propone la utilización de técnicas de síntesis digital para la generación de señales de control [SMT99, EAD99, MAR93], sin tener en cuenta soluciones analógicas puras.

La palabra digital de entrada, que representa el ángulo, debe estar en formato binario con una representación natural; la salida se genera en formato resolver, en fase con la señal de referencia y acondicionada para permitir el acople con los demás elementos de la cadena sincrónica.

En la Figura 1 se muestra un esquema de las entradas y salidas del sistema, viendo al convertidor como una caja negra.

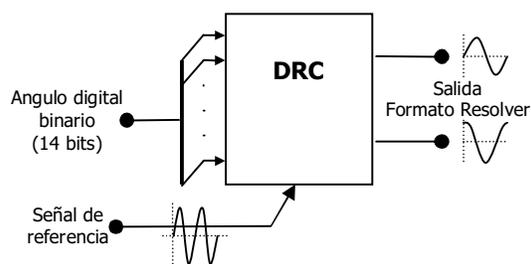


Figura 1 - Sistema DRC

Si bien existen en el mercado dispositivos que realizan tal conversión, son demasiados costosos y difíciles de conseguir, sobre todo cuando el presupuesto es escaso o se está desarrollando un nuevo prototipo. Por lo tanto, se propone como objetivo puntual, el desarrollo de dicho convertidor, con una resolución de 14 bits y una exactitud de ± 4 arc min.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Se trabaja en paralelo con varias líneas de investigación, por un lado el estudio de metodologías para el desarrollo de sistemas embebidos que permita la selección de diferentes ciclos de vida en función de los requerimientos del sistema a desarrollar, por otro lado, el desarrollo de algoritmos de síntesis digital de señales analógicas mediante unidades de proceso de bajo costo (microcontroladores de 8 bits). Por último se analizarán los medios para la digitalización de los sistemas de control inductivos y su aplicación a la robótica.

Los resultados del desarrollo del trabajo constan de dos etapas bien diferenciadas. La primera de ellas es un análisis exhaustivo del problema, que incluye la investigación sobre todos los temas relacionados al sistema (sistemas de control, sincros/resolvers, señales y sistemas embebidos), y el planteo del diseño y las arquitecturas del módulo, junto con las decisiones para la materialización de cada parte del mismo. La segunda etapa constituye la implementación en sí del sistema propuesto.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados del trabajo fueron obtenidos a través etapas bien diferenciadas por la metodología. La primera de ellas es un análisis exhaustivo del problema, que incluye la investigación sobre todos los temas relacionados al sistema (sistemas de control, sincros / resolvers, señales y sistemas embebidos), el planteo del diseño y las arquitecturas del módulo, junto con las decisiones para la materialización de cada parte del mismo. Mientras que la segunda etapa constituye la implementación en sí del sistema propuesto. Por último, se procede a la integración de los diferentes módulos desarrollados y se comprueba que los requerimientos sean satisfechos.

3.1. Análisis - Requerimientos

Los requerimientos para este trabajo, están acotados a la realización de un dispositivo cuya funcionalidad replique a la de los DRC (Convertor Digital-Resolver) comerciales, minimizando principalmente el costo por unidad. Debido a esto, luego de un relevamiento de las especificaciones de los productos del mercado, se determinan las características comunes a todos ellos que deben estar presentes para la materialización del sistema. En la Tabla 1 se listan los principales requerimientos a ser tenidos en cuenta.

Requerimientos Funcionales	
Parámetro	Valor
Entrada de Referencia:	
Frecuencia	400Hz
Tensión	115 Vrms + 20%
Entradas Digitales:	
Formato	14 bits paralelo
Tensión	5V - TTL
Salidas Analógicas:	
Formato	Resolver
Tensión	5Vrms
Tensiones de entrada:	+5v / +15v / -15v

Tabla 1 - Requerimientos Funcionales

Además de las especificaciones mencionadas, se considera, a partir del análisis del problema, que existen atributos no funcionales a tener en cuenta. Entre los mismos se considerarán la confiabilidad, la latencia y la modificabilidad.

3.2. Análisis - Arquitectura del Sistema

La arquitectura elegida para el desarrollo del problema está basada en microcontroladores, debido que es la que mejor se adapta a las necesidades del sistema. Por lo tanto, el diseño del sistema se plantea teniendo en cuenta las características propias de los sistemas embebidos [BAR06, NOE05, SAS03, CAT02].

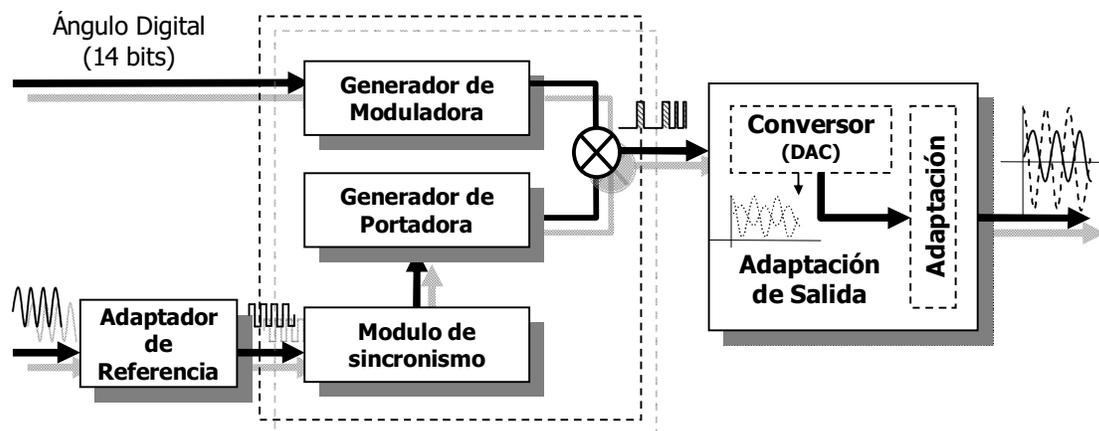


Figura 2 - Arquitectura Final del Sistema

Se debe tener en cuenta que cierta parte de la funcionalidad del módulo a desarrollar se implementa en hardware y el resto utilizando módulos de software. En la Figura 2. se muestra la Arquitectura final del Sistema, en la cual se observan los módulos necesarios para la implementación de la funcionalidad requerida, así como también las señales de entrada y salida con la adaptación necesaria que debe realizarse a las mismas.

Los módulos contenidos dentro de la línea punteada son aquellos que se implementan vía software; mientras que las adaptaciones de entradas y salidas, para lograr los formatos de señales adecuados, se llevan a cabo por medio de dispositivos de hardware. Cabe aclarar que para poder utilizar dichos dispositivos, se deben implementar los controladores de los mismos.

Una vez establecida la partición entre software y hardware de la funcionalidad a desarrollar, se toman las decisiones de los dispositivos específicos a través de los que se materializan dichos módulos. La Tabla 2 detalla los componentes de hardware elegidos.

Componentes de Hardware	
Dispositivo	Componente
Microprocesador	ATmega16 (AVR)
DAC	DAC8532 (Serie)
Adaptadores E/S	Electrónica discreta

Tabla 2 - Componentes de Hardware

Los módulos a implementar por software son, básicamente, los responsables de la modulación de las señales a través de la cual se realizará la conversión. Para cumplir con los requerimientos de calidad de señales establecidos, debe tenerse en cuenta, en particular, el análisis del tiempo de que se dispone para el proceso completo. En función de ese análisis se toman las decisiones en cuanto a los posibles algoritmos a utilizar para la generación y/o transformación de las distintas señales presentes.

3.3. Implementación – Hardware

Esta etapa supone la materialización de los módulos de hardware. Para cada uno de los módulos de hardware surgidos en la etapa de particionado, se implementa y se simula, de manera independiente, la funcionalidad del mismo. Se testea, además, la correctitud de los valores obtenidos utilizando las herramientas de medición necesarias [CAT02]. Una vez comprobada la funcionalidad de todos los componentes, se procede a diseñar y construir el circuito impreso (PCB, Printed Circuit Board) que servirá de base para la integración y testeo del sistema en conjunto. Para construir el PCB se utiliza una herramienta de CAD (Computer Aided Design),

que acompaña el desarrollo del sistema embebido, permitiendo la creación de esquemáticos para su posterior ruteo e implementación.

En la Figura 3 se observa una vista en tres dimensiones del circuito impreso desarrollado.

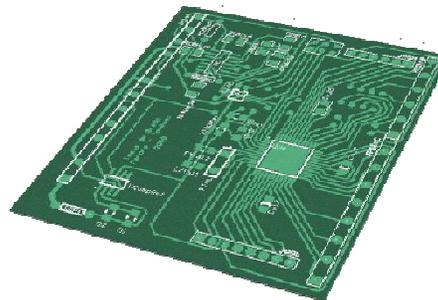


Figura 3. Vista 3D del Circuito Impreso

3.4. Implementación – Software

La implementación de software se basa en una arquitectura de capas, dividida principalmente en dos partes bien definidas; por un lado, la implementación de los controladores (*drivers*) de los dispositivos de hardware, y por otro, la de la funcionalidad específica del sistema. Para esta última, los módulos a implementar son:

Generador de Moduladora: Este módulo es el encargado de la actualización de los valores de amplitud que modulan la portadora, cuando el ángulo de entrada cambia.

Generador de Portadora: Responsable de la generación de una señal senoidal, en sincronismo con la referencia del sistema.

Generador de Sincronismo: Encargado de generar el evento necesario para el sincronismo de la señal portadora.

Las decisiones que se deben tomar en esta etapa están relacionadas, principalmente, con la obtención de señales que cuenten con la calidad requerida. Para lograr dicha calidad se debe realizar un análisis de tiempo sobre el sistema para saber cual será la máxima cantidad de muestreos por segundos que se pueden realizar.

Luego de la codificación de cada módulo, se realizan los testeos y simulaciones necesarias para verificar su correcto funcionamiento [BAR06].

3.5. Integración

Luego de la implementación y testeo de todos los componentes del sistema (Hardware y Software), se procede a la integración de todos ellos, materializando las interfaces de hardware y de software definidas y ensamblando el sistema completo.

Luego de los ajustes necesarios para que el sistema sea funcional, se verifica que la salida generada esté dentro de los valores requeridos. Para esto se realizan dos tipos de chequeos:

Verificación de la calidad de la señal: Se hace un análisis de la distorsión armónica de la señal resultante y se comprueba que la misma, no presenta distorsiones significativas para el contexto donde se utilizara el conversor [EAD99].

Verificación de la precisión de la señal: Se crea un banco de pruebas que permite introducir todos los posibles valores de entrada al conversor y compararlos con su salida, graficando la diferencia. Como resultado de este testeo se obtiene un error máximo acotado en ± 7 arco minutos, valor que se encuentra próximo a los requerimientos especificados [EAD04].

En la Figura 4 se observa la placa final desarrollada y ensamblada.

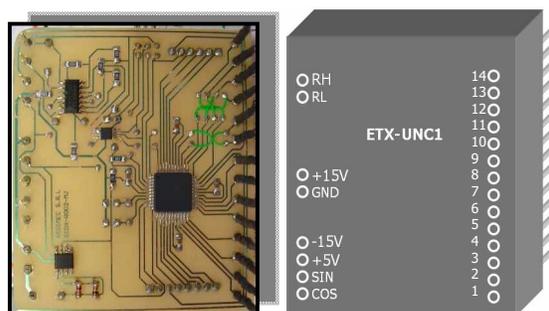


Figura 4. Conversor Terminado

4. FORMACIÓN DE RECURSOS HUMANOS

Durante el desarrollo de este trabajo participaron dos estudiantes avanzados de la carrera (ahora graduados), por el periodo de 6 meses. Del mismo proyecto se realizó la tesina de grado, además uno de los estudiantes se incorporó en una empresa del polo informático donde se aseguró el buen funcionamiento del dispositivo sobre el diseño electrónico. Por último a uno de estos estudiantes se le ha otorgado la beca de posgrado de CONICET para realizar su doctorado.

5. BIBLIOGRAFÍA

[NAI06] North Atlantic Industries Inc. - "Synchro and Resolver Conversion" REV A, edited by John Gasking and the editorial group of NAI Inc, ISBN 0-916550-06-0, 2006.

[BAR06] Michael Barr, Anthony Massa, "Programming Embedded Systems", O'Reilly & Associates, Inc., ISBN: 0-596-00983-6, 2006.

[NOE05] Tammy Noergaard, "Embedded Systems Architecture", Elsevier, ISBN: 0-7506-7792-9, 2005.

[DOR04] "Modern Control Systems (10th Edition)".- Richard C. Dorf, Robert H Bishop.- ISBN-13: 978-0131457331.-2004.

[EAD04] Engineering Staff Analog Devices - "Data Conversion Handbook", ISBN-10: 0750678410, ISBN-13: 978-0750678414, 2004.

[SAS03] Sastry, S. Sztipanovits, J. Bajcsy, R. Gill, H. "Special issue on modeling and design of embedded software", Proceedings of the IEEE, Vol. 91, No.1, January 2003 - University of California, Berkeley.

[CAT02] John Catsoulis "Designing Embedded Hardware".- O'Reilly.- ISBN: 0-596-00362-5.- November 2002

[BAL00] Ball, S.- "Embedded Microprocessor Systems: Real World Design".- ISBN-13: 978-0750672344, 2000.

[SMT99] Steven W. Smith - "Digital Signal Processing", California Technical Publishing, ISBN 0-9660176-7-6, ISBN 0-9660176-6-8, 1999.

[EAD99] Engineering Staff Analog Devices - "A Technical Tutorial on Digital Signal Synthesis", www.analog.com, 1999.

[DDC94] Data Device Corporation - "Synchro and Resolver Conversion Handbook" FOURTH EDITION - Catalog Number: 74-77038 - <http://www.ddc-web.com> 1994.

[MAR93] C. Marven, G. Wiers - "A Simple Approach to Digital Signal Processing", Texas Instruments 1993. ISBN 0-904-04700-8. Texas Instruments 1993.

[CHE88] C.H. Chen - "Signal Processing Handbook" - ISBN/EAN:978-0-8247-7956-6, 1988.

APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA LA NAVEGACIÓN DE ROBOTS MÓVILES UTILIZANDO EL SIMULADOR PLAYER/STAGE

Marcelo Marinelli, Horacio Kuna, Hugo Plenc, Fernando Korol,

Departamento de Informática, Facultad de Ciencias Exactas Químicas y Naturales.
Universidad Nacional de Misiones.

marcelomarinelli@fceqyn.unam.edu.ar, hdkuna@unam.edu.ar, hplenc@fceqyn.unam.edu.ar,
fernandokorol@fceqyn.unam.edu.ar

CONTEXTO

Este proyecto es la continuación de línea de investigación del proyecto acreditado 16Q366 “Evaluación de la Efectividad para Evadir Obstáculos de un Sistema Experto de Navegación para Robot Móvil” y el “Programa de Investigación en Computación” de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones; también vinculado con el Doctorado en Ingeniería de Sistemas y Computación que funciona en la Universidad de Málaga.

RESUMEN

En la robótica móvil hay distintos problemas típicos que se están investigando, por ejemplo el problema de seguir un contorno le permite a un robot realizar una trayectoria en la modalidad denominada “in door”, en donde se pretende que un robot domestico, de oficina o industrial se desplace llevando una carga desde un punto origen hasta un punto destino siguiendo a una pared y evitando los obstáculos que se le interpongan.

Existen distintas técnicas de Inteligencia Artificial que se aplican a estos sistemas de navegación como Redes Neuronales, Lógica Difusa, Neuro-Difusas, etc.

Por medio del simulador y seleccionando la técnica adecuada se pretende implementar un controlador para la navegación autónoma y evaluar los resultados obtenidos.

Palabras clave: Navegación de Robot, Inteligencia Artificial, Simulador de Entorno.

1. INTRODUCCION

1.1 Tipos de Control de Navegación

Los sistemas de navegación autónoma para robots móviles modernos combinan el control reactivo y el deliberativo en modelos de capas [1], [2]. Esta estratificación se diseña de manera que la capa de nivel mas bajo se encargue del control reactivo basado en los datos que recibe de los sensores, estos pueden ser grupos de radares ultrasónicos, infrarrojos o de barrido láser y permiten interactuar con el entorno encargándose de tareas como ser: seguir a un objeto a una determinada distancia o moverse en un entorno complejo siguiendo una trayectoria que mantenga una distancia fija a una pared.

Las arquitecturas híbridas suelen ser de tres capas y sobre la capa reactiva descrita en el párrafo anterior, interactúan las capas ejecutiva y deliberativa. La primera recibe la información de la capa inferior y realiza una planificación de corto plazo pudiéndose encargar de la localización de la posición y del almacenamiento de los mapas de entorno. La capa deliberativa se encarga de la planificación y utiliza distintos modelos para la toma de decisiones a mediano plazo. Pueden existir modelos con mayor número de capas, como por ejemplo una capa adicional para la interfaz de usuario.

Para la capa de control se utilizan, con muy buenos resultados, los controladores difusos combinados con técnicas de generación automática de conjunto de datos [3].

A fin de aumentar la eficiencia de estos controladores se utilizan técnicas de

optimización basadas en algoritmos genéticos [4], [5] y [6]. La metodología empleada en este trabajo combina el aprendizaje automático de la base de conocimientos utilizando una sistema híbrido, en el que se aplica la cooperación entre reglas (COR) complementado con un algoritmo de optimización mediante colonia de hormigas (OCH) [1] y [7]. Esta combinación tiene la característica de generar bases de reglas en forma rápida, con muy buena precisión.

Se utiliza el algoritmo desarrollado en [1] adaptándolo a un simulador compuesto por un servidor de simulación denominado Player (<http://playerstage.sourceforge.net>) bajo licencia GNU que permite el control de un robot que interactúa sobre los dispositivos de navegación y lee las señales provenientes de sus sensores a través de una conexión TCP/IP. Este servidor se conecta a un simulador de robots en dos dimensiones denominado Stage perteneciente al mismo proyecto. Estos dos programas en conjunto son denominados Player/Stage y pueden realizar la simulación del control de un robot sobre escenarios artificiales de manera que permiten desarrollar algoritmos de control de robots móviles en distintos lenguajes y ponerlos a prueba sin disponer de un robot real.

El proyecto Player fue desarrollado originalmente en el Robotic Research Lab de la USC (University of South California) pero actualmente es un proyecto de código abierto sustentado por un grupo de desarrolladores [9].

1.2 Técnica para generar conjuntos de datos

Para realizar la generación automática del conjunto de datos se siguen los siguientes pasos:

1. Definición de las variables de entrada y salida dependiendo de la configuración del robot: se determinan las variables de control que son enviadas al robot y los valores de las variables de los sensores que dispone el mismo para utilizar en la percepción del entorno.
2. Definición del universo de discurso: la partición de los conjuntos difusos de todas las variables con sus respectivas precisiones.

3. Definición de una función de puntuación que provee la efectividad del controlador difuso.

4. Definición de la función objetivo para evaluar la calidad global del conjunto de reglas.

5. Modelado matemático del comportamiento del robot para definir las ecuaciones que rigen la dinámica del mismo en las situaciones planteadas.

6. Construcción del conjunto de entrenamiento: a partir de la combinación de los valores que toman las variables, en función de los antecedentes, se obtienen los consecuentes combinando todos los valores posibles que pueden tomar.

1.3 Desarrollo de aplicaciones clientes con Player/Stage

En C++ la conexión entre el programa cliente y el servidor se hace operando con objetos locales denominados proxies. El programa interactúa con los métodos de estos objetos abstrayéndose de la comunicación TCP/IP por la red.

Para diseñar un programa cliente se deben cumplir tres pasos:

1. Realizar la conexión con el simulador a través de un objeto de la clase PlayerClient.
2. Interactuar con los sensores y sistemas de tracción con la clase correspondiente.
3. Establecer un bucle que periódicamente interactúe con los sensores y en función de la telemetría envíe los datos de control correspondientes a los sistemas de tracción.

Inicio de programa

```
#include <libplayerc++/playerc++.h> // se declara la librería
```

```
PlayerClient robot(gHostname, gPort); // realiza la conexión con el robot o simulador
```

```
Position2dProxy pp(&robot, gIndex); // instancia la clase del sistema de tracción
```

```
LáserProxy lp(&robot, gIndex); // instancia la clase correspondiente al barrido láser
```

Inicio de bucle

```
robot.Read(); // realiza la conexión con el robot o simulador
```

d=lp[i] // toma lectura del láser

Rutina de control

*pp.SetSpeed(vel_lineal ,vel_angular); // envía
datos de control de tracción*

fin de bucle

fin de programa

2. LÍNEAS DE INVESTIGACION Y DESARROLLO

Dentro de las líneas de investigación que contienen el “Programa de Investigación en Computación”, este proyecto se enmarca en el área de inteligencia artificial en donde se aplican técnicas de lógica difusa y algoritmos bioinspirados, estas técnicas se utilizarán en trabajos futuros para:

- Controladores difusos aplicados al proceso de elaboración de yerba mate.
- Sistemas de control de navegación para robot utilizando sensores láser.
- Aplicación de controladores difusos para el control de invernaderos con riego artificial.

3. RESULTADOS OBTENIDOS/ESPERADOS

En el Proyecto “Evaluación de la Efectividad para Evadir Obstáculos de un Sistema Experto de Navegación para Robot Móvil” se diseñó una plataforma móvil con un sistema de tracción diferencial compuesto por dos motores de corriente continua.

Para el control de velocidad de los motores se desarrolló un control PWM con un microcontrolador PIC de la empresa Microchip Technology Inc., se diseñó un software que permite entregar cinco valores de velocidad a cada uno de los motores en concordancia con los cinco valores difusos que entregará el controlador difuso.

Las señales de control se interconectan con la CPU principal del sistema por medio de un puerto RS232.



Disposición de Componentes

Fig.1 placa de control PWM

Actualmente se está trabajando en el diseño del controlador difuso para el sistema de navegación.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de investigación se prevén desarrollar dos Tesis de grado de la carrera de Licenciatura en Sistemas de Información y dos Tesis doctorales del Doctorado en Ingeniería de Sistemas y Computación del Departamento de Lenguajes y Ciencias de la Computación de la Universidad de Málaga

5. BIBLIOGRAFIA

- [1] M. Mucientes, J. Casillas. “Quick Design of Fuzzy Controllers With Good Interpretability in Mobile Robotics”. IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 15, NO. 4, pp. 636-651, 2007.
- [2] S. Russell, P. Norvig, “Inteligencia Artificial Un Enfoque Moderno”. Madrid, Pearson Educacion S.A. 2004.
- [3] M. Mucientes, R. Iglesias, C. V. Regueiro, A. Bugarini, and S. Barro, “A fuzzy temporal rule-based velocity controller for mobile robotics”. Fuzzy Sets Syst., vol. 134, pp. 83–99, 2003.

- [4] B. del Brio, A. Molina “Redes Neuronales y Sistemas Borrosos”. W. Pedrycz, Ed. Norwell, Alfaomega, pp. 289–306, 2007.
- [5] A. Bonarini, “Evolutionary learning of fuzzy rules: Competition and cooperation”. in *Fuzzy Modelling: Paradigms and Practice*, W. Pedrycz, Ed. Norwell, MA: Kluwer Academic, pp. 265–284, 1996.
- [6] M. Mucientes, D. L. Moreno, C. V. Regueiro, A. Bugarini, and S. Barro, “Design of a fuzzy controller for the wall-following behavior in mobile robotics with evolutionary algorithms”. in *Proc. Int. Conf. Inf. Process. Manage. (Uncertainty Knowledge-Based Syst. IPMU’2004)*, Perugia, Italy, pp. 175–182, 2004.
- [7] J. Casillas, O. Cordon, I. Fernandez de Viana, F. Herrera, “Learning cooperative linguistic fuzzy rules using the best-worst ant system algorithm”. *International Journal of Intelligent Systems* 20:4, pp. 433-452. ISSN 0884-8173. doi:10.1002/int.20074, 2005.
- [8] B. Gerkey, R. Vaughan, K. Stoy, A. Howard, G. Sukhatme, M. Mataric. “Most Valuable Player: A Robot Device Server for Distributed Control”. Robotics Research Labs, University of Southern California Los Angeles, CA 90089-0721, USA.
- [9] Collett, T.H., MacDonald, B.A., Gerkey, B.P.: “Player 2.0: Toward a Practical Robot Programming Framework”. *Proceedings of the Australasian Conference on Robotics and Automation (ACRA 2005)*, 2005.
- [10] J. Urzelai, J. P. Uribe, and M. Ezkerra, “Fuzzy controller for wallfollowing with a non-holonomous mobile robot”. *Proc. 6th IEEE Int. Conf. Fuzzy Syst. (Fuzz-IEEE’97)*, Barcelona, Spain, pp. 1361–1368, 1997.34
- [11] J. Casillas, O. Cordon, and F. Herrera, “COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 32, no. 4, pp. 526–537, 2002.
- [12] J. Casillas, O. Cordon, “COR methodology: A simple way to obtain linguistic fuzzy models with good interpretability and accuracy,” in *Accuracy Improvements in Linguistic Fuzzy Modeling*, J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, Eds. Heidelberg, Germany: Springer, 2003.
- [13] M. Dorigo and T. Stutzle, *Ant Colony Optimization*. Cambridge, MA: MIT Press, 2004.
- [14] O. Cordon, F. Herrera, I. F. de Viana, and L. Moreno, “A new AGO model integrating evolutionary computation concepts: The best-worst ant system,” in *Proc. 2nd Int. Workshop Ant Algorithms*, Brussels, Belgium, pp. 22–29, 2000.

Procesamiento de Ontologías Inconsistentes en la Web Semántica Usando Argumentación Rebatible: Reporte de Progreso

Sergio Alejandro Gómez, Carlos Iván Chesñevar, Guillermo Ricardo Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA)¹
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Av. Alem 1253 - (8000) Bahía Blanca - Argentina
Tel/Fax: (+54) 291-4595135/6
E-mail: {sag, cic, grs}@cs.uns.edu.ar

Resumen

La *Web Semántica* es una visión futura de la Web en la cual la información tiene un significado exacto definido en términos de *ontologías*. Una ontología define un vocabulario en términos de una jerarquía de conceptos e individuos que pertenecen a tales conceptos. Una ontología es anómala cuando es incoherente o inconsistente; estas situaciones se caracterizan por la presencia de definiciones de conceptos vacíos y por contradicciones lógicas, respectivamente. La argumentación rebatible es un tipo de razonamiento no-monótono que produce conclusiones tentativas en presencia de información incompleta y potencialmente contradictoria. En esta línea de investigación, exploramos la aplicación de la argumentación rebatible al procesamiento de ontologías inconsistentes en la Web Semántica. En este artículo, presentamos brevemente la problemática asociada, recopilamos los resultados obtenidos y discutimos el trabajo en progreso.

Palabras clave: Inteligencia Artificial, Razonamiento no-monótono, Argumentación rebatible, Programación en Lógica Rebatible, Web Semántica, Ontologías, Lógicas para la Descripción, Agentes inteligentes

Contexto: Esta línea de investigación está financiada por la Agencia Nacional de Promoción Científica y Tecnológica, por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina), por el Ministerio de Educación y Cultura (España) y por la Universidad Nacional del Sur.

1. Introducción y motivaciones

La *Web Semántica* es una visión futura de la Web donde el significado de los recursos web se define en forma precisa para permitir su procesamiento por agentes [3]. La Web actual está formada por documentos HTML formateados para su presentación a usuarios humanos. El éxito de la web a nivel de usuarios finales se debe a que la información se halla presentada en modo textual. Este modo textual no es adecuado para su entendimiento por agentes o programas inteligentes, quienes deben recurrir a la presencia o ausencia de palabras claves para determinar la relevancia de un determinado documento web. En la Web Semántica el significado de los términos presentes en una página web se halla definido precisamente por medio de una o varias ontologías. Una ontología es una formalización de una parte de un dominio de aplicación [13].

Existen varios acercamientos al problema de la definición de ontologías en la Web Semántica. El lenguaje RDF [14] es el más simple, pero al brindar flexibilidad hace que el tratamiento de su caso

¹LIDIA es un miembro del IICyTI (Instituto de Investigación en Ciencia y Tecnología Informática).

general sea indecible computacionalmente. Otro lenguaje de representación de ontologías es RDFS-schema [4], que permite definir una jerarquía de clases o conceptos junto con atributos y relaciones con dominio y codominio en tales clases. Sin embargo, la necesidad de expresar restricciones más adecuadas (como por ejemplo, expresar que dos clases son disjuntas) llevó a la necesidad del desarrollo de un nuevo lenguaje llamado OWL [15]. OWL contiene tres sublenguajes: OWL-Lite (el cual es simple y debería ser suficiente para muchos usuarios finales), OWL-DL (el cual se halla basado en las Lógicas para la Descripción) y OWL-Full (que provee los mismos constructores que OWL-DL pero permite representar situaciones más ricas que lo hacen computacionalmente indecible).

La semántica del lenguaje OWL-DL se halla basada en las Lógicas para la Descripción. Las *Lógicas para la Descripción* (DL) [1] son una familia de formalismos de representación de conocimiento basados en las nociones de *conceptos* (predicados unarios, clases) y *roles* (relaciones binarias), y están principalmente caracterizados por constructores que permiten describir conceptos complejos y roles a partir de otros atómicos usando conjunción, disyunción, negación, restricciones (cuantificaciones) existencial y de valor, entre otros. Una ontología DL consiste de dos conjuntos finitos y mutuamente disjuntos: una Tbox que introduce la *terminología* y una Abox que contiene *aserciones* acerca de objetos particulares en el dominio de aplicación. Las sentencias de la Tbox representan inclusiones o equivalencias de descripciones (posiblemente complejas) de conceptos.

Una de las tareas de razonamiento que se pueden realizar con ontologías es el *chequeo de instancia*, el cual consiste en determinar si un individuo determinado es o no instancia de un concepto dado. Las anomalías que pueden surgir a la hora de computar el chequeo de instancia son dos: la incoherencia y la inconsistencia. Una ontología es incoherente cuando contiene definiciones de conceptos vacíos. En cambio, una ontología es inconsistente cuando no tiene modelo (*i.e.*, la ontología es incoherente pero además se han declarado explícita o implícitamente individuos en las clases incoherentes).

El éxito de la iniciativa para la Web Semántica se apoya en la existencia de razonadores existentes (*e.g.*, Racer, Fact, Pellet), los que permiten razonar con ontologías coherentes y consistentes exclusivamente. Desde un punto de vista lógico, el problema con las ontologías inconsistentes estriba en que no es posible decidir si un individuo pertenece o no a una determinada clase. Sin embargo, si bien tales razonadores son capaces de reconocer ontologías incoherentes e inconsistentes, el ingeniero de conocimiento solamente es notificado de tal situación.

Hay dos maneras no excluyentes de tratar el problema de la inconsistencia. Una consiste en *reparar* la ontología; es decir, hacerla nuevamente coherente y consistente. Esta opción muchas veces es imposible de llevar a cabo por varios motivos, entre los que se cuentan que el ingeniero de conocimiento no tiene autoridad ni el conocimiento para corregir una ontología, una ontología puede estar basada en otras ontologías importadas, o el dominio a modelar puede ser intrínsecamente inconsistente. Una segunda manera de lidiar con el problema de la inconsistencia en ontologías consiste en aceptar la inconsistencia y utilizar un mecanismo de razonamiento no-estándar para obtener alguna respuesta *significativa*. En nuestro acercamiento al razonamiento con ontologías inconsistentes aplicamos la *argumentación rebatible* para llevar a cabo el segundo enfoque.

El estudio de la argumentación puede ser considerado concerniente con cómo las aserciones son propuestas, discutidas y resueltas en el contexto de asuntos en los cuales se pueden mantener varias opiniones divergentes [18, 5, 2]. Tales opiniones divergentes son llamadas *argumentos*. Dado un argumento todos los contraargumentos del mismo son considerados, los contraargumentos de estos últimos y así sucesivamente para formar un árbol de dialéctica. Cuando un contraargumento es derrotado por otro contraargumento, se dice que el argumento inicialmente atacado es *reinstaurado*. Los argumentos que son raíz de algún árbol de dialéctica en el cual todos sus derrotadores (si los hubiera) se hallan derrotados se dicen *argumentos garantizados* y los mismos son considerados las conclusiones válidas del sistema argumentativo.

En particular, la Programación en Lógica Rebatible (DeLP) [6, 19] brinda un lenguaje de representación de conocimiento y razonamiento que combina la Programación en Lógica y la argumentación rebatible. De esta manera, la codificación de conocimiento por medio de la DeLP provee un equilibrio entre expresividad e implementabilidad, que permite lidiar con conocimiento potencialmente contradictorio. Sucintamente, un programa DeLP \mathcal{P} se halla formado por un conjunto hechos, un conjunto de reglas rebatibles (o tentativas) y un conjunto de reglas estrictas. Dada una consulta con respecto a \mathcal{P} , representada por un literal lógico L , el mecanismo de razonamiento forma argumentos a favor y en contra de L , y, mediante un análisis dialéctico, se determina su estado de garantía dando una respuesta que puede ser SÍ (cuando L se halla garantizado), NO (cuando la negación de L se halla garantizada) o INDECISO (cuando no es posible llegar a una conclusión).

Por otro lado, estudios recientes [12, 20] relacionan a las DL con el formalismo de la Programación en Lógica. Esta área de investigación se ha asentado con el nombre de Programación en Lógica Descriptiva (o en inglés, *Description Logic Programming* o DLP). Básicamente, los axiomas de inclusión de clases de las Tboxes DL se corresponden con implicaciones lógicas de primer orden, las cuales, a su vez, pueden ser expresadas como reglas Prolog. Además, las aserciones de las Aboxes DL pueden ser expresadas como hechos en un programa Prolog. También, las conjunciones, disyunciones y restricciones existenciales y de valor pueden ser traducidas al lenguaje de la Programación en Lógica bajo ciertas restricciones que permiten incluir más del noventa por ciento de las ontologías existentes.

En la Sección 2 explicamos cómo esta línea de investigación integra a las Lógicas para la Descripción con la argumentación rebatible en un formalismo de razonamiento con ontologías posiblemente inconsistentes llamado δ -ontologías, repasamos los resultados obtenidos, y, además, discutimos el trabajo actual en progreso vislumbrando líneas futuras de investigación. Finalmente, en la Sección 3 detallamos la relación de esta línea de investigación con la formación de recursos humanos.

2. Resultados obtenidos / esperados

En esta línea de investigación exploramos los siguientes tópicos:

- Representación de ontologías mediante argumentación rebatible
- Razonamiento con ontologías inconsistentes utilizando argumentación rebatible
- Caracterización formal de las propiedades emergentes del formalismo de razonamiento desarrollado
- Análisis de las posibilidades de aplicación e implementación del formalismo desarrollado junto con una evaluación empírica de su aplicación a ontologías reales.

A continuación resumimos los resultados obtenidos en este respecto y discutimos los resultados esperados en el corto y mediano plazo para esta línea de investigación.

2.1. Resultados obtenidos

En trabajos previos [8, 9, 11], presentamos un formalismo de razonamiento con ontologías DL posiblemente inconsistentes llamado δ -ontologías. Una δ -ontología conserva el lenguaje de representación de conocimiento de las ontologías DL tradicionales² pero da semántica a tales ontologías como programas DeLP. En una δ -ontología Σ , si bien la caja asercional tiene el mismo significado que en

²Como ontologías DL tradicionales, nos referimos a la interpretación dada por el texto de [1].

las ontologías DL tradicionales (*i.e.*, es un conjunto de afirmaciones de pertenencia de individuos a conceptos y relaciones entre individuos), en cambio la terminología se encuentra dividida en un conjunto de axiomas terminológicos estrictos (o Sbox) y un conjunto de axiomas tentativos (o Dbox). De esta manera, las aserciones de una δ -ontología se corresponden con los hechos de un programa DeLP, mientras que la Sbox se corresponde con un conjunto de reglas estrictas y la Dbox con un conjunto de reglas rebatibles. Así, en presencia de inconsistencias en la ontología, un análisis dialéctico es llevado a cabo sobre el programa DeLP que interpreta Σ para computar las tareas de razonamiento típicas tales como el chequeo de instancia y la recuperación de individuos.

En nuestras investigaciones hemos definido formalmente el marco de razonamiento de las δ -ontologías y hemos caracterizado su comportamiento con un conjunto de ejemplos típicos. Por otro lado, hemos estudiado una buena parte de las propiedades que emergen de la aplicación de tal formalismo al razonamiento con ontologías inconsistentes. También, hemos mostrado una aplicación del formalismo introducido al problema de la integración de ontologías [9] y al problema del desarrollo de formularios web inteligentes [10].

2.2. Resultados esperados

Nuestro trabajo de investigación actual está orientado hacia dos objetivos principales. Por un lado, estamos interesados en completar la caracterización formal de las propiedades emergentes del formalismo de las δ -ontologías, entre las que se incluyen el estudio de estrategias para particionar Tboxes DL automáticamente en Sboxes y Dboxes y aplicaciones al área de mezcla de ontologías.

Por otro lado, estamos interesados en el estudio de la implementabilidad computacional del sistema planteado. En este sentido se presentan particularmente atractivas el uso de las herramientas servidor KAON y DLP [17]. Para facilitar el uso de DLP, Motik *et. al.* [17, 16] brindan la herramienta de conversión `dlpconvert` que permite convertir fragmentos DLP codificados en OWL en sintaxis de *Edinburgh Prolog*. `dlpconvert` está basado en los algoritmos para reducir Lógicas para la Descripción a Datalog implementado en KAON2. El sistema lee una ontología OWL, la reduce a Datalog disyuntivo, si es posible, y finalmente la serializa como un programa lógico. Creemos que esta herramienta puede ser combinada con el servidor DeLP [7], el cual permite realizar consultas con respecto a un programa rebatible en forma remota .

3. Formación de recursos humanos

Esta línea de investigación está relacionada con el desarrollo de una Tesis de Doctorado en Ciencias de la Computación inscripta en la Universidad Nacional del Sur.

Agradecimientos: Esta línea de investigación está financiada por la Agencia Nacional de Promoción Científica y Tecnológica (PICT 2002 No. 13.096, PICT 2003 No. 15.043, PAV 2004 076), y por los Proyectos PIP 112-200801-02798 (CONICET, Argentina), TIN2006-15662-C02-01 (MEC, Spain), PGI 24/ZN10 (SGCyT, UNS, Argentina) y la Universidad Nacional del Sur.

Referencias

- [1] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook – Theory, Implementation and Applications*. Cambridge University Press, 2003.
- [2] T. J. M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artif. Intell.*, 171(10-15):619–641, 2007.

- [3] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scient. American*, 2001.
- [4] Dan Brickley and R.V. Guha. Resource Description Framework (RDF) Schema Specification W3C Proposed Recommendation 03 March 1999. <http://www.w3.org/TR/1999/PR-rdf-schema-19990303/> (checked 2009/03/18), 1999.
- [5] C. Chesñevar, A. Maguitman, and R. Loui. Logical Models of Argument. *ACM Computing Surveys*, 32(4):337–383, December 2000.
- [6] A. García and G. Simari. Defeasible Logic Programming: An Argumentative Approach. *Theory and Prac. of Logic Program.*, 4(1):95–138, 2004.
- [7] Alejandro Javier García, Nicolás D. Rotstein, Mariano Tucát, and Guillermo Ricardo Simari. An argumentative reasoning service for deliberative agents. In Zili Zhang and Jörg H. Siekmann, editors, *KSEM*, volume 4798 of *Lecture Notes in Computer Science*, pages 128–139. Springer, 2007.
- [8] Sergio Alejandro Gómez, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. Inconsistent Ontology Handling by Translating Description Logics into Defeasible Logic Programming. *Revista Iberoamericana de Inteligencia Artificial*, 11(35):11–22, 2007.
- [9] Sergio Alejandro Gómez, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. An Argumentative Approach to Reasoning with Inconsistent Ontologies. In Thomas Meyer and Mehmet A. Orgun, editors, *Proc. of the Knowledge Representation in Ontologies Workshop (KROW 2008)*, volume CPRIT 90, pages 11–20, Sydney, Australia, 2008.
- [10] Sergio Alejandro Gómez, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. Integration of Web-based Forms with Ontologies in the Semantic Web. In Antonio Castro Lechtaler Norberto Caminoa, Fernanda Carmona, editor, *Actas del XIV Congreso Argentino de Ciencias de la Computación (CACIC 2008)*, Chilecito, Argentina, 6-10 octubre 2008, 2008.
- [11] Sergio Alejandro Gómez, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. Reasoning with inconsistent ontologies through argumentation. *Submitted to J. of Applied Artificial Intelligence*, 2008.
- [12] Benjamin Grosz, Ian Horrocks, Raphael Volz, and Stefan Decker. Description Logic Programs: Combining Logic Programs with Description Logics. *WWW2003, May 20-24, Budapest, Hungary*, 2003.
- [13] T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.
- [14] Johan Hjelm. *Creating the Semantic Web with RDF*. John Wiley & Sons, 2001.
- [15] Deborah L. McGuinness and Frank van Harmelen. OWL Web Ontology Language Overview, 2004. <http://www.w3.org/TR/owl-features/>.
- [16] B. Motik, D. Vrandečić, P. Hitzler, Y. Sure, and R. Studer. dlpconvert: Converting owl dlp statements to logic programs. In *European Semantic Web Conference 2005 (ESWC 2005) Demos and Posters*, 2005.
- [17] Boris Motik, Raphael Volz, and Sean Bechhofer. Dlp - description logic programs, 2003. <http://kaon.semanticweb.org/alphaworld/dlp> (checked 2009/03/18).
- [18] Henry Prakken and Gerard Vreeswijk. Logical Systems for Defeasible Argumentation. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic*, pages 219–318. Kluwer Academic Publishers, 2002.
- [19] G. Simari and R. Loui. A Mathematical Treatment of Defeasible Reasoning and its Implementation. *Artificial Intelligence*, 53:125–157, 1992.
- [20] Raphael Volz. *Web Ontology Reasoning with Logic Databases*. PhD thesis, Universität Fridericiana zu Karlsruhe, 2004.

Proyecto NNGen: Aplicación de controladores neuronales a la navegación autónoma

Mg. Marcelo A. Tosini
Grupo Inca/Intia
Departamento de Computación y Sistemas
Facultad de Ciencias Exactas
Universidad Nacional del Centro de la Provincia de Buenos Aires

CONTEXTO

La línea de trabajo presentada se encuentra inmersa dentro del proyecto de Incentivos *Sensores y Algoritmos Para Plataformas de Navegación* cuyo objetivo fundamental es la definición de una metodología y el desarrollo de las herramientas correspondientes, para asistir al diseño e implementación de sistemas empotrados a medida de la aplicación, que utilizan técnicas de Inteligencia Artificial para la detección de patrones en señales o imágenes, sistemas de control industrial, y navegación de vehículos autónomos. Estos sistemas pueden estar basados en plataformas tales como: microcontroladores, procesadores genéricos o circuitos a medida (en FPGA o ASIC). Como banco de prueba para validar la metodología y herramientas desarrolladas, se analizarán aplicaciones de uso industrial, tales como redes neuronales digitales, sistemas de control basados en lógica difusa, sistemas de navegación basados en esquemas deliberativos, sensores de variables imprecisas, giro-estabilización, cámaras inteligentes para reconocimiento, entre otras.

RESUMEN

El proyecto Senyal Plana llevado adelante en el seno del grupo Inca/Intia persigue como objetivos fundamentales la creación

de herramientas y metodologías para el desarrollo de agentes robóticos autónomos. En el marco de dicho proyecto, la línea orientada a robótica evolutiva y redes neuronales digitales trabaja a partir de lineamientos específicos destinados a obtener materializaciones en hardware de controladores neuronales que minimicen el área *on chip* manteniendo alta la relación tiempo/eficiencia. Además de esto, se realizan estudios formales a nivel evolutivo de controladores robóticos en busca de relaciones entre parámetros que permitan implementar controladores minimales pero con aceptable comportamiento a fin de integrarlos en controladores mas elaborados con estrategias integradas deliberativo/reactivas que soporten comportamientos mas complejos.

Palabras clave: *inteligencia artificial, redes neuronales, robótica evolutiva*

1. INTRODUCCION

La robótica evolutiva, entendida como la aplicación de técnicas de evolución genética para entrenar controladores neuronales en determinadas tareas, se ha orientado en los últimos tiempos al aspecto teórico de la disciplina dejando de lado el punto de vista de la ingeniería [1]. Esta orientación de las investigaciones a dado buenos resultados manifestados sobre todo

en estudios teóricos en la complejidad de las tareas realizadas y en el ámbito de la aproximación simbólica a los problemas [2] [3].

Aún así, es de interés el estudio de las problemáticas asociadas a controladores aplicables a dispositivos robóticos reales en hardware. Justamente en este entorno se debe atender a factores asociados colateralmente al problema central que maneja el controlador y que están relacionados con interrupciones producidas por interferencias eléctricas (o de otra índole) en los sensores, en los actuadores (motores o circuitos de control de los mismos) o del propio ambiente (diferencias entre el comportamiento del ambiente real y el modelo simulado) [4] [5].

Desde el punto de vista del diseño del controlador los diferentes ruidos se incluyen en las ecuaciones que modelan la red neuronal de control en tres puntos diferentes: 1) como componente aditiva a los valores aportados por los sensores; 2) como factor (coeficiente de fricción de los motores, ruido mecánico o ruido eléctrico amplificado por los controladores de los motores) o 3) como ruido adicionado al modelo de cada neurona.

El ruido propio del ambiente se puede definir ya sea como perturbaciones del escenario real que no se pueden modelar correctamente en el entorno simulado o como ruido propio eventual del escenario real. En este último caso el modelo debe ser capaz de generar situaciones en las que no existe perturbación y otras en las que la perturbación puede aparecer eventualmente en algún momento del experimento.

Varios autores han trabajado el tema de la influencia del ruido (u otro tipo de incertezas) en el comportamiento intrínseco de controladores neuronales genéticamente evolucionados. Di Paolo en [6] plantea básicamente que para capturar todas las características del problema en un controlador es necesario considerar todos los aspectos subyacentes – ruidos, incertezas- puesto que la simplificación del

modelo al excluir alguno de ellos aumenta el riesgo de obtener controladores buenos pero más frágiles.

En [7] Gallardo analiza el tema de la generación de trayectorias aplicando entrenamiento evolutivo y plantea soluciones al problema de la incertidumbre en el comportamiento de los sensores y actuadores reales mediante la incorporación de un factor de ruido gaussiano en las ecuaciones que manejan un controlador para la realización de trayectorias óptimas en un escenario con obstáculos. Los resultados muestran que las soluciones obtenidas en ambientes con ruido son menos óptimas que aquellas sin ruidos ya que los controladores se ven forzados a alejarse más de los obstáculos a fin de esquivarlos con seguridad. Aún así todos los controladores cumplen su objetivo primario de recorrer el escenario en un tiempo cercano al mínimo.

Beers [3] hace un extenso análisis del ruido en sistemas neuronales que reciben información sensorial para la toma de decisiones y la coordinación sensorimotora. En particular analiza la influencia de ruido en la interacción visual y motora (movimiento de la mano) para llegar al objetivo de tomar un objeto. Para esta problemática considera la existencia de varias fuentes de ruido o incertezas. Por un lado aquellas en el sistema de visión respecto de la posición real del objeto y por otro las propias del sistema propioceptivo que introducen errores propios del sistema motor. Estas últimas fuentes de incertezas deben entenderse como la diferencia entre la posición de la mano deseada (ordenada) por el sistema nervioso central (SNC) y la posición real de la mano debido a cualquier influencia que afecte al sistema muscular, sistema nervioso periférico, etc. Beers también analiza la influencia de un tercer tipo de incerteza que es la que surge de la captación que el sistema visual (con ruido) realiza de la posición de la mano (también con ruido). Esto genera un tipo de incerteza de mas alto nivel en el que la posición

exacta de la mano no se conoce exactamente debido a no se puede inferir directamente si los errores de posición de la misma provienen del sistema de visión o del sistema de propiocepción o, en último caso, de ambos.

En todos los casos el ruido está presente en el nivel inferior (nivel neuronal), mientras que la manera en que las tareas son realizadas se ve a alto nivel (nivel de comportamiento). Esto indica que es siempre necesario considerar los efectos a nivel neuronal a fin de entender su influencia sobre el nivel comportamental.

Por otro lado, La utilización de redes neuronales para la resolución de problemas complejos o no lineales ha inspirado la creación de varias soluciones tanto en software –sobre computadoras de propósito general-, como en hardware; con arquitecturas (en este último caso) que van desde la implementación de las redes en base a procesadores dedicados programados hasta materializaciones de alto rendimiento usando técnicas de segmentación de circuitos en configuraciones sistólicas de procesadores elementales simples [8].

En el caso particular de las redes neuronales multi etapa muchas propuestas en hardware implementan solamente la fase de prueba de la red usando pesos sinápticos obtenidos por entrenamiento *off line* realizado en computadora. Esto responde principalmente a dos razones: por un lado, la aplicación final de una red determinada en hardware no varía, por lo cual el entrenamiento *on line* no es necesario. Por otro lado, el entrenamiento *on chip* supone un aumento de la complejidad de los circuitos de la red neuronal que perjudican el rendimiento final o que en algunos casos son imposibles de materializar en determinados dispositivos de hardware.

La sub-línea de trabajo en este sentido busca analizar diversas alternativas en hardware dedicado para redes de varias capas ocultas orientado a su implementación final en FPGA. Las arquitecturas propuestas son básicamente

sistólicas con un diseño orientado a la interconexión de diversos componentes segmentados que implementan las distintas ecuaciones del algoritmo. Estas decisiones de diseño buscan lograr la generalidad necesaria para poder incorporar a futuro nuevas características a los circuitos que amplíen sus capacidades operativas al procesamiento de otras implementaciones de redes neuronales.

2. LINEAS DE INVESTIGACION Y DESARROLLO

El proyecto se puede dividir esencialmente en dos líneas a saber: Robótica evolutiva, que analiza comportamientos de controladores robóticos, su interacción con diversos escenarios experimentales y las estrategias para la obtención de controladores específicos para realizar determinadas tareas reactivas (reflejas). Arquitecturas hardware de controladores neuronales, que busca maneras de implementar circuitos de redes neuronales en distintas plataformas hardware como FPGAs o microcontroladores. Esta línea también analiza distintas variantes de implementación de los operadores necesarios para implementar la funcionalidad de la red con variantes como aritméticas de números con residuos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Algunos resultados parciales del trabajo realizado se plasmaron en publicaciones en congresos como [9] y [10].

Los objetivos de la línea plantean la obtención de los siguientes resultados: a) cuantificación de perturbaciones (ruido eléctrico) en el funcionamiento de un controlador neuronal, ya sea en fase de entrenamiento o de prueba del mismo; b)

desarrollo de arquitecturas neuronales de distintas topologías de redes en circuitos electrónicos programables como por ejemplo FPGAs o microcontroladores; c) aplicación de aritméticas no usuales para la obtención de circuitos de cálculo mas rápidos como por ejemplo RNS (Residue Number System) o aritmética decimal; d) estudio de comportamientos complejos basados en esquemas deliberativos y reactivos que permitan abstraer conocimientos de más alto nivel en capas superiores de decisión; e) combinación de diferentes estrategias reactivas tales como redes neuronales, lógica difusa o control humano en esquemas deliberativos controlados por algoritmos a tal fin; f) implementación de redes de sensores inteligentes partiendo de grupos de sensores de distintas características que permita la inferencia de conocimiento de nivel más abstracto que la información concreta que reporta cada sensor en particular.

4. FORMACION DE RECURSOS HUMANOS

En el área de robótica evolutiva se está trabajando con un becario de esta casa de estudios con residencia en la Universidad de Sussex (Inglaterra) en el estudio de comportamientos simples y complejos de agentes genéticamente evolucionados. Esta relación dio como fruto reciente la presentación de una tesina de grado que aborda el estudio del control deliberativo/reactivo en agentes simulados probados en escenarios simples como fototaxis o evasión [11].

Paralelamente se trabaja con integrantes del grupo de la línea de desarrollo de unidades aritméticas en la elaboración de módulos aritméticos para operaciones matriciales en RNS y decimal. En este sentido, el grupo tiene una estrecha vinculación con investigadores de la Universidad Autónoma de Madrid que trabajan en el área de

aritméticas en FPGA y con los cuales se han realizado varias publicaciones conjuntas.

5. BIBLIOGRAFIA

- [1] Nolfi, S.; Adaptation as a more powerful than decomposition and integration: Experimental evidences from evolutionary robotics. In P.K. Simpson (Ed.), Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 98), New York: IEEE Press, 141-146, 1998.
- [2] Nolfi, S., Floreano, D.; Learning and evolution. *Autonomous Robots*, 7(1): 89-113, 1999
- [3] Van Beers, R. J., Baraduc, P. and Wolpert, D.; Role of uncertainty in sensorimotor control. *Philosophy Transaction, The Royal Society*, 357, 1137-1145, July 2002.
- [4] Ott, H.; Noise reduction techniques in electronic systems, John Wiley & Sons, 1998.
- [5] Balcells, J. et al.; Interferencias electromagnéticas en sistemas electrónicos, Macrombo, 1992.
- [6] Di Paolo, E. and Harvey, I.; Decisions and noise: The scope of evolutionary synthesis and dynamical analysis, *International society for Adaptive Behavior*, Vol 11 (4): 284-288, 2003.
- [7] Gallardo, D., Colomina, O., Flórez, F., Rizo, R.; Generación de trayectorias robustas mediante computación evolutiva. SCETA 97, Seminario sobre computación evolutiva: teoría y aplicaciones. Torremolinos (Malaga), Noviembre 1997.
- [8] Sundarajan, N., Saratchandran, P., "Parallel Architectures for Artificial Neural Networks", *IEEE Computer*

- Society Press, ISBN: 0-8186-8399-6, 1998.
- [9] Tosini, M., “Análisis de comportamiento de agentes neuronales evolutivos ante perturbaciones externas e internas”, XIV Congreso Argentino de Ciencias de la Computación, Chilesito, La Rioja, 6 al 10 de octubre de 2008
 - [10] Tosini, M., “Diseño de un procesador neuronal orientado a redes multi-etapa entrenado con backpropagation”, XV Workshop Iberchip, Buenos Aires, Argentina, 25 al 27 de marzo de 2009.
 - [11] Goñi, E., “Estrategias de navegación evolutiva para robots autónomos”, tesina de grado para la obtención del título de Ingeniero de Sistemas, Director: Mg. Marcelo Tosini, codirector: Mg. José Fernández León

Algoritmos Metaheurísticos para Optimización y Aplicación a Problemas NP Completos

Hugo Alfonso¹, Carolina Salto¹, Gabriela Minetti¹, Natalia Stark¹

Carlos Bermúdez², Alina Orellana³, Patricia Graglia⁴

Laboratorio de Investigación en Sistemas Inteligentes (LISI)

Facultad de Ingeniería - Universidad Nacional de La Pampa

Calle 110 Esq. 9 (6360) General Pico – La Pampa – Rep. Argentina

Te. / Fax: (02302) 422780/422372, Int. 6302

e-mail: ¹{alfonsoh, saltoc, minettig, nstark@ing.unlpam.edu.ar },

²bermudezc@yahoo.com, ³orellanaalina@gmail.com, ⁴pmg_xxi@yahoo.com.ar

CONTEXTO

El Proyecto de Investigación “*Sistemas Metaheurísticos para Resolver Problemas de Optimización*” es llevado a cabo en el Laboratorio de Investigación de Sistemas Inteligentes (LISI), de la Facultad de Ingeniería de la Universidad Nacional de La Pampa. El mismo surge como consecuencia de un trabajo sostenido que se comenzó en 1997 a través del estudio de Algoritmos Genéticos.

La profundización en el tema y la propia evolución, tanto de la línea de investigación como del grupo de investigación, han llevado a extender el estudio a otras metaheurísticas que han permitido abordar problemas de mayor complejidad o resolver instancias de mayor dimensión a las ya abordadas.

RESUMEN

En la actualidad las empresas deben enfrentar un conjunto de problemas logístico-operativos, de alta complejidad, conocidos en la comunidad científica como problemas de optimización combinatoria. Actualmente, en esta comunidad se observa una importante tendencia a resolver dichos problemas con la utilización de algoritmos heurísticos y metaheurísticos.

Nuestro grupo está abocado al diseño y desarrollo de algoritmos heurísticos y metaheurísticos que resuelvan problemas de optimización. En particular se ha puesto especial énfasis en: el problema de corte y empaquetado, y en el de planificación y

programación de recursos y en el ensamblado de fragmentos de ADN. Tanto la optimización de la planificación de recursos como la de generación de patrones de cortes, reducen significativamente los costos de los distintos recursos involucrados. Esto se debe a la mejor utilización que se hace de los mismos, lograda por medio de la aplicación de metaheurísticas. Por otro lado, las metaheurísticas también permiten resolver problemas de optimización en el área de la bioinformática; la cual se beneficia con la capacidad de hacer búsquedas en el espacio de problemas realmente grandes, en un tiempo razonable sin necesidad del uso de información extra. Estas son ventajas que no ofrecen los algoritmos específicos de esta área.

Ya sea en el contexto industrial como en el bioinformático, las metaheurísticas han sido juzgadas o evaluadas como beneficiosas, ya que con un esfuerzo limitado se pueden alcanzar buenos resultados con gran versatilidad. Actualmente dos de las ramas con más éxito para diseñar metaheurísticas eficientes, y dar solución a estos problemas, son la hibridación y el paralelismo.

Palabras clave: *metaheurísticas, computación paralela, optimización combinatoria, métodos de búsqueda híbrida.*

1. INTRODUCCION

Las Metaheurísticas [7] emergieron como una nueva clase de algoritmos de búsqueda aproximados que, básicamente, combinan las características de los métodos heurísticos

básicos en estructuras de mayor nivel, intentando con ello alcanzar una mayor eficiencia y efectividad en la exploración del espacio de búsqueda. El término “*metaheurística*” fue introducido en 1986 por Glover y deriva de la composición de dos palabras griegas *Heuristic*, que significa encontrar, y *meta*, que significa "más allá, en un nivel superior". Luego, este término fue ampliamente adoptado y también frecuentemente es denominado *heurísticas* modernas [16].

Muchos problemas de optimización, de importancia tanto teórica como práctica, buscan la mejor configuración de un conjunto de variables para alcanzar el objetivo planteado por el mismo. Ellos en forma natural se dividen en dos categorías: aquellos cuyas soluciones están codificadas con variables reales y aquellos que necesitan variables discretas.

Dentro de esta última clase de problemas podemos identificar a los bien conocidos *Problemas de Optimización Combinatoria (POC)*. Estos problemas requieren de la búsqueda de la mejor solución dentro de un determinado espacio de soluciones, para ello se desarrollan diferentes algoritmos que pueden evidenciar distinto grado de *eficiencia* y *eficacia* en alcanzar tal objetivo.

Los algoritmos que atacan a estos problemas pueden ser a su vez clasificados como *completos* o *aproximados*. Los algoritmos completos están garantizando encontrar para cada instancia del problema de tamaño finito una solución óptima en tiempo limitado (ver [13 y 15]). Para POCs que sean **NP-duros** [6], no existen algoritmos que los resuelvan en tiempo polinomial, asumiendo que $P \neq NP$. Por lo tanto, los métodos completos deben necesitar tiempo de computación de orden exponencial en el peor de los casos. Esto conlleva a que para algunas instancias grandes se requiera un tiempo computacional demasiado grande para el propósito práctico perseguido. Por ello, el uso de los métodos aproximados para la resolución de POCs ha recibido mucha atención en las últimas tres

décadas. *Estos métodos sacrifican la garantía de alcanzar soluciones óptimas por la de encontrar buenas soluciones en un tiempo significativamente reducido.*

Resumiendo, se puede decir que las metaheurísticas son estrategias de alto nivel para la exploración de espacios de búsqueda usando diferentes métodos. Es por ello muy importante, que permitan un balance dinámico entre *diversificación* e *intensificación* de la búsqueda. El primero de los términos está asociado al concepto de exploración de diversas regiones del espacio de búsqueda y el segundo, a la explotación del espacio de búsqueda perteneciente al vecindario de una solución de calidad.

Hay diferentes formas de clasificar y describir los algoritmos metaheurísticos.

Nuestro proyecto se ha centrado en el estudio de un conjunto de metaheurísticas que responden a una de las clasificaciones posibles. Ésta distingue entre búsqueda basada en población versus soluciones simples, la cual divide las metaheurísticas en métodos basados en trayectoria y métodos basados en población, ya que la misma permite un mejor encuadre de los algoritmos a utilizar. Sin embargo, es necesario destacar que una tendencia actual es la de hibridar los distintos métodos ya sea para obtener mejores soluciones o soluciones más rápidamente. En esa hibridación es común integrar algoritmos de búsqueda de soluciones simples en algoritmos basados en población.

Métodos Basados en Trayectoria

El término Método Basado en Trayectoria es usado porque el proceso de búsqueda llevado a cabo se caracteriza por determinar una trayectoria en el espacio de soluciones. Por lo tanto, una solución sucesora puede o no pertenecer al vecindario de la actual. Entre estos métodos se destacan: Búsqueda Local Básica: Mejoramiento Iterativo, Recocido Simulado (Simulated Annealing - SA), Búsqueda Tabú (Tabu Search - TS), Métodos de Búsqueda Local Explorativa, Búsqueda Adaptativa Aleatoriamente Voraz (Greedy

Randomized Adaptive Search Procedure - GRASP), Búsqueda en Vecindario Variable (Variable Neighborhood Search - VNS), Búsqueda Local Guiada (Guided Local Search - GLS) y la Búsqueda Local Iterada (Iterated Local Search - ILS).

Métodos Basados en Población

Los métodos basados en población operan en cada iteración del algoritmo con un conjunto de soluciones, denominada *población*. Así ellos, en forma natural, exploran el espacio de búsqueda desde varias regiones a la vez. El resultado final depende de la forma en que la población es manipulada. Los métodos basados en población más ampliamente estudiados son los Algoritmos Evolutivos (Evolutionary Algorithms - EAs), la optimización de Colonia de Hormigas (Ant Colony Optimization - ACO), Optimización por Cúmulo de Partículas (Particle Swarm Optimization - PSO) y Búsqueda Dispersa (Scatter Search - SS).

2. LINEAS DE INVESTIGACION Y DESARROLLO

En la actualidad estamos trabajando con dos áreas muy prometedoras, una de ellas es la hibridación de distintas metaheurísticas y la otra es la paralelización de las mismas.

Hibridación de Metaheurísticas

Es importante destacar que existen distintas formas de hibridación:

- Intercambio de Componentes entre Metaheurísticas. Una de las formas más populares de hibridación es la incorporación de métodos de trayectoria en métodos basados en población. La combinación más popularmente utilizada son EAs que usan procedimientos de búsqueda local para intensificar la búsqueda en áreas más prometedoras.
- Búsqueda Cooperativa. Una forma más amplia de hibridación es provista por la búsqueda cooperativa (ver [2 y 9]) que consiste de una búsqueda realizada por

distintos algoritmos que intercambian información acerca de estados, modelos, subproblemas íntegros, soluciones u otras características del espacio de búsqueda. Generalmente, son algoritmos de búsqueda que se ejecutan en paralelo con un nivel variable de comunicación. Pueden ser que los algoritmos que están cooperando respondan a un modelo idéntico (enfoque homogéneo) o bien que respondan a distintos modelos (enfoque heterogéneo). Este enfoque, en la actualidad, recibe mucha atención debido a la creciente investigación sobre implementaciones paralelas de metaheurísticas.

- Integrando Metaheurísticas y Métodos Sistemáticos. Este enfoque recientemente ha producido muchos algoritmos efectivos, especialmente aplicados a problemas del mundo real. Diversas formas de integración se resumen en Freuder et al. [5]. La hibridación es frecuentemente llevada a cabo integrando conceptos desarrollados para metaheurísticas (por ejemplo, elecciones probabilísticas, criterios de aspiración, construcción de heurísticas) en un método de búsqueda de árbol.

Paralelización de Metaheurísticas

Desde mediados de la década de los 80's se han propuesto versiones paralelas de los métodos metaheurísticos con una frecuencia que va en aumento. Los objetivos que generalmente se usan para justificar la realización del cómputo en paralelo son dos. En primer lugar, el tiempo de cómputo más razonable para resolver instancias del problema de mayor tamaño o formuladas en forma más cercanas a la realidad. En segundo lugar, las metaheurísticas paralelas pueden ser más robustas que las versiones secuenciales relativas a diferencias en tipos y características del problema y la adecuada calibración paramétrica.

Las metaheurísticas más frecuentemente paralelizadas son las evolutivas, ya sean

algoritmos genéticos o sistemas de colonia de hormigas, y métodos de recocido simulado o procedimientos de búsqueda tabú. Independientemente de la característica metodológica, Crainic et al. [4] han propuesto una clasificación de acuerdo al nivel de impacto que la estrategia de paralelización tiene sobre el diseño algorítmico de la metaheurística. Las tres estrategias son:

- Paralelización de operaciones dentro de una iteración del método. Esta estrategia, también llamada estrategia de bajo nivel, es bastante sencilla y ayuda a agilizar los cómputos. Alcanza los mismos resultados que el método secuencial sólo que más rápido.
- Descomposición del dominio del problema o espacio de búsqueda. Esta estrategia está basada en el principio que el poder computacional puede ser dedicado a resolver una gran cantidad de pequeñas partes del problema, sobre las cuales una solución global puede ser construida o extraída. Aquí la trayectoria de la búsqueda resultante en el enfoque paralelo es diferente a la seguida en el enfoque secuencial.
- Caminos de búsqueda múltiples con varios grados de sincronización y cooperación. Esta estrategia intenta realizar una mayor exploración del espacio de búsqueda iniciando varios procesos de búsqueda simultánea en el mismo espacio.

3. RESULTADOS OBTENIDOS/ESPERADOS

Durante los últimos años el grupo ha profundizado en el estudio e implementación de diversas metaheurísticas, cuyo rendimiento se ha evaluado sobre un conjunto de funciones que responden a diversas características [14]. En todos los casos, los resultados alcanzados se han contrastado con los publicados en la literatura de referencia. Dependiendo de las características de los problemas abordados se han incorporado mecanismos de búsqueda

local en los EAs, introduciendo la hibridación [17, 10 y 12].

Las metaheurísticas se han implementado usando un software libre, desarrollado en un marco de cooperación entre tres universidades españolas: Málaga, La Laguna y Barcelona, denominado MALLBA [1].

Los estudios realizados sobre las distintas metaheurísticas consisten en: evaluar la calidad de los resultados obtenidos, medir el rendimiento al paralelizar el proceso usando un modelo isla (que permite realizar la búsqueda independiente de buenas soluciones, para luego migrar los resultados a las restantes islas, y así, cooperar en el proceso de búsqueda global). En este último caso se han realizado estudios para determinar la topología y la política de migración que resulten más adecuadas [3].

También se está trabajando en la búsqueda de soluciones a problemas de alta complejidad mediante el diseño de metaheurísticas para los siguientes problemas: de corte y empaquetado de dos dimensiones [17, 18 y 19], el problema de ensamblado de fragmentos de ADN [10, 11, 12] y el problema de ruteo de vehículos con capacidad limitada (CVRP) [8]. Algunos de los avances logrados ya han sido comunicados en revistas específicas y se continúa trabajando en estas líneas de investigación. Los primeros dos problemas son los abordados en las tesis doctorales de dos integrantes del equipo.

4. FORMACION DE RECURSOS HUMANOS

Desde su origen, el LISI ha mantenido un estrecho contacto con grupos de UNSL y de UNPA. Desde 2002 se comenzó a trabajar en forma conjunta con el grupo NEO de la Universidad de Málaga (España), lo que ha permitido realizar un trabajo cooperativo muy rico. Actualmente, dos integrantes del LISI se encuentran escribiendo sus tesis doctorales con dirección de docentes investigadores de UMA y UNSL.

Por otra parte, a lo largo del año 2008 otros dos integrantes del grupo han presentado tesis

para alcanzar el título de Ingeniero en Sistemas, a partir de las actividades concretas que desarrollaron en el LISI, y se estima que para el corriente año otro integrante realice la defensa de su tesis.

Desde el inicio del LISI, se han formado varios becarios en iniciación de la investigación, algunos de los cuales son miembros activos del actual proyecto.

5. BIBLIOGRAFIA

- [1] E. Alba, F. Almeida, M. Blesa, C. Cotta, M. Díaz, I. Dorta, J. Gabarro, J. Gonzalez, C. León, L. Moreno, J. Petit, J. Roda, A. Rojas y F. Xhafa. MALLBA: Towards a Combinatorial Optimization Library for Geographically Distributed Systems. Proc. of the XII Jornadas de Paralelismo. Pag. 105-110. [Http://citeseer.nj.nec.com/alba01mallba.html](http://citeseer.nj.nec.com/alba01mallba.html). 2001.
- [2] V. Bachelet, E-G. Talbi. Cosearch: a co-evolutionary metaheuristic. In Congress on Evolutionary Computation CEC'2000, pag. 1550-1557, 2000.
- [3] C. Bermúdez, H. Alfonso y C. Salto. Analizando el Desempeño de Distintas Topologías en Algoritmos Evolutivos Distribuidos. En los anales del XIII Congreso Argentino de Ciencias de la Computación (CACIC 2008), 2008.
- [4] T. Crainic y M. Toulouse. Parallel metaheuristics. Fleet management and logidtics, pag. 205-251, 1998.
- [5] E.C. Freuder, R. Dechter, M.L. Ginsberg, B. Selman, y E.P.K. Tsang. Systematic versus stochastic constraint satisfaction. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95, vol 2, pag. 2027-2032, Montreal, Quebec, Canada. Morgan Kaufmann. 1995.
- [6] M. Garey y D. Johnson. Computers and Intractability: a Guide to the Theory of NP-completeness. Freeman, San Francisco, California, 1979.
- [7] F. Glover. Future paths for integer programming and links to artificial intelligence. Computers and Operations Research, vol. 13, pag. 533--549, 1986.
- [8] P. Graglia, N. Stark, C. Salto y H. Alfonso. A Comparison of Recombination Operators for Capacitate Vehicle Routing Problem. En los anales del XIII Congreso Argentino de Ciencias de la Computación (CACIC 2008), 2008.
- [9] T. Hogg y C.P. Williams. Solving the really hard problems with cooperative search. In Proceedings of the 11th National Conference on Artificial Intelligence, AAAI, pag. 231-236, Washington, DC, USA. The AAAI Press/The MIT Press. 1993.
- [10] G. Minetti, E. Alba y G. Luque. Variable Neighborhood Search for Solving the DNA Fragment Assembly Problem. En los anales del XIII Congreso Argentino de Ciencias de la Computación (CACIC 2007), 2007.
- [11] G. Minetti, E. Alba y G. Luque. Seeding Strategies and Recombination Operators for Solving the DNA Fragment Assembly Problem. Information Processing Letters 108, pag 94-100. ISSN: 0020-0190. 2008
- [12] G. Minetti, G. Luque y E. Alba. Variable Neighborhood Search as Genetic Algorithm Operator for DNA Fragment Assembling Problem. Anales de la Eighth International Conference on Hybrid Intelligent Systems, HIS 2008, pag. 714-719. ISBN: 978-0-7695-3326-1, 2008.
- [13] G. Nemhauser y L. Wolsey. Integer and Combinatorial Optimization. John Wiley, New York, 1998.
- [14] A. Orellana y G. Minetti. A Phenotypic Analysis of Three Population-based Metaheuristics. Anales del IX Congreso Argentino de Ciencias de la Computación (CACIC 2008), 2008.
- [15] C.H. Papadimitriou y I.C. Steiglitz. Combinatorial Optimization: Algorithms and Complexity. Prentice-Hall, Englewood Cliffs edition, 1982.
- [16] C.R. Reeves. Modern Heuristic Techniques for Combinatorial Problems. Advanced Topics in Computer Science. Mc Graw-Hill, 1995.
- [17] C. Salto, G. Leguizamón, E. Alba y J.M. Molina. Hybrid Ant Colony System to Solve a 2-Dimensional Strip Packing Problem. En los anales de Internacional Conference on Hibrid Intelligent Systems (HIS08), pag 708-713. Publicado en CD-ROM. ISBN: 978-0-7695-3326-1, 2008.
- [18] C. Salto, J.M. Molina y E. Alba. Greedy Seeding and Problem Specific Operators for GAs Solving Strip Packing Problems. In "Optimization Techniques for Solving Complex Problems", E. Alba, C. Blum, P. Isasi, C. León, and J.A. Gómez (Editors), A Wiley Interscience Publication, JOHN WILEY & SONS, (in Press). ISBN: 978-0-470-29332-4. 2008
- [19] C. Salto, G. Leguizamón, E. Alba, y J.M. Molina. Evolutionary and ACO Based Approaches for Two-dimensional Strip Packing Problem. In "Natural Intelligence for Scheduling, Planning, and Packing Problems", Springer-Verlag in the series "Studies in Computational Intelligence". Raymond Chiong (editor). In press. 2008

NUEVOS DESARROLLOS PARA SISTEMAS ADAPTATIVOS INTELIGENTES

DURAN, Elena, MALDONADO, Marilena, UNZAGA, Silvina, COSTAGUTA, Rosanna
Departamento de Informática

Facultad de Ciencias Exactas y Tecnologías
Universidad Nacional de Santiago del Estero

Tel: 03854509560 – E-mails: {eduran, marilena, sunzaga, rosanna }@unse.edu.ar
Avda. Belgrano (S) 1912 - 4200 Santiago del Estero, Argentina

CONTEXTO

En este artículo se presenta el subproyecto “Sistemas Adaptativos Inteligentes”, que es parte del proyecto de investigación “Herramientas conceptuales, metodológicas y técnicas de la informática teórica y aplicada”, aprobado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SICYT- UNSE), Código 23/C062, en ejecución desde el año 2006. Este proyecto, es a su vez una continuación de la línea de investigación iniciada con el proyecto “Estudio Sistemático de impactos y Derivaciones Metodológicas – Técnicas de la Informática Aplicada (BIO-PSICO-SOCIO-TECNO-CULTURAL)” (SICYT- UNSE); Código 23/C044.

RESUMEN

Como la Web continua ganando popularidad, son numerosos los servicios que operan en este ambiente, aunque muchos de ellos no satisfacen a los usuarios, quienes día a día se vuelven más exigentes y demandan un mejor servicio. Atendiendo a esta problemática, se ha desarrollado el subproyecto de investigación “Sistemas Adaptativos Inteligentes”, cuyo objetivo principal es mejorar el nivel de aceptación de los sistemas informáticos por parte de los usuarios, satisfaciendo sus demandas, reduciendo sus temores y aumentando el atractivo y la flexibilidad. Para ello, se han diseñado, desarrollado y evaluado sistemas adaptativos en diferentes áreas de aplicación. En este artículo se presenta el estado de avance alcanzado en las tres líneas de investigación de este subproyecto: la educación a distancia soportada por computadora (e-learning), los sistemas de gestión y el comercio electrónico.

Se presentan también, los resultados alcanzados y la formación de recursos humanos concretada en el marco del subproyecto.

Palabras Claves: Sistemas adaptativos, Modelo de usuario, Comercio Electrónico, Sistemas de aprendizaje, Interfaces adaptativas.

1. INTRODUCCIÓN

Internet es un medio de comunicación que permite el intercambio de información entre los usuarios conectados a la red, ofreciendo una oportunidad única, especial y decisiva a diversas organizaciones; aunque no siempre el servicio que brindan satisface a los usuarios en sus necesidades. Esto sucede porque en la mayoría de las aplicaciones no se consideran las características particulares de los usuarios y no se construyen relaciones que lo motiven a usar nuevamente el servicio seleccionado. La respuesta a esta problemática ha sido el desarrollo de sistemas adaptativos o personalizados. Estos sistemas aprenden automáticamente las características del usuario, y sobre la base de ese conocimiento alteran aspectos de su funcionalidad e interacción adaptándola a las diferentes preferencias de los usuarios. La personalización puede traducirse en la construcción de distintas interfaces adaptadas a cada aplicación y cliente en particular, en proporcionar caminos de navegación y/ o información personalizados según las preferencias del usuario, entre otras facilidades.

Los Sistemas Adaptativos Inteligentes (SAI) incorporan nueva tecnología que involucra el

uso de software que aprende patrones, hábitos y preferencias de los usuarios. Esto se logra mediante la incorporación al sistema de un modelo de usuario, que es una representación de algunas características del mismo, necesarias para concretar una interacción individualizada. Durante la interacción, el sistema construye un modelo para cada usuario y lo actualiza constantemente para tener siempre disponible el estado actual del usuario (Durán y Amandi, 2009).

En los SAI, las técnicas de inteligencia Artificial (IA) juegan un rol preponderante, ya que el sistema tiene que realizar diversas tareas del proceso de adaptación que deben minimizar el esfuerzo del usuario, buscando maximizar su grado de satisfacción. La tecnología de agentes de software es un paradigma emergente en el área de la IA. Particularmente los agentes de interfaz, agentes personales o agentes asistentes, se han propuesto especialmente para colaborar con el usuario aprendiendo acerca de sus intereses, hábitos y preferencias.

Atendiendo a la problemática descrita, en este subproyecto se plantean como objetivos:

- Mejorar el nivel de aceptación de los sistemas proporcionando a los usuarios ayudas adaptadas a sus características y preferencias.
- Realizar propuestas metodológicas para el diseño y construcción de Modelos de Usuario, sobre la base de aprendizaje de máquina.
- Diseñar, desarrollar y evaluar modelos de usuario para adaptar sistemas en diferentes áreas de aplicación, utilizando la tecnología de agentes inteligentes.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Atendiendo la finalidad plantada en la sección anterior, en el presente subproyecto se aborda el diseño, construcción y evaluación de SAIs, en diferentes dominios. Por lo tanto las líneas de investigación que de él se desprenden son las que se describen a continuación:

Desarrollos de SAIs en ambientes de aprendizaje virtual o e-learning. Estos ambientes brindan diferentes ventajas en el proceso educativo, entre las cuales está la posibilidad que dicho proceso se adapte al ritmo de aprendizaje del alumno. Pero a pesar del auge que está teniendo hoy en día esta modalidad, no garantiza una mayor calidad, ni un aprendizaje más rápido ni más eficaz por sí solo. La principal causa de este problema es que los sistemas de e-learning se diseñan sin considerar las características particulares de cada estudiante. Por lo tanto, en esta línea de investigación se están realizando estudios tendientes a detectar automáticamente las características y preferencias de los estudiantes para adaptar los cursos de e-learning.

Puntualmente se están desarrollando dos trabajos. Uno que tiene por objetivo construir un sistema de detección del estilo de aprendizaje de cada estudiante por medio de la aplicación de técnicas de análisis de Cluster. En este sentido, una de las características más relevantes de los estudiantes es su estilo de aprendizaje, ya que no todos los estudiantes aprenden del mismo modo sino que existen diferentes estilos para acceder al conocimiento (Felder y Silverman, 1988). El análisis de Cluster, es empleado en este trabajo para agrupar comportamientos similares mientras el alumno interactúa con el sistema e-learning. Una vez obtenido los clusters, se espera que el cluster dominante determine el estilo de aprendizaje del estudiante el cual será enmarcado dentro de las cuatro dimensiones propuestas por el modelo de aprendizaje de Felder y Silverman (1988).

El otro trabajo abordado en esta línea se orienta a definir un modelo de alumno basado en Redes de Bayes para detectar automáticamente el nivel de conocimiento del estudiante en un curso de e-learning, implementado en la plataforma Moodle. Disponer de información sobre el nivel de conocimiento del estudiante sobre un dominio

en particular, permitirá proveerle material de estudio o problemas ajustados a dichos conocimientos; así como ofrecerle consejos que lo orienten durante el proceso de aprendizaje.

Desarrollos de SAIs en sistemas de gestión. Debido a la gran diversidad de usuarios que manejan las aplicaciones hoy en día, resulta indispensable que los sistemas posean interfaces que se adapten a las necesidades y preferencias de estos de manera natural y progresiva. Es por esto que, en esta línea, se están realizando estudios tendientes desarrollar interfaces adaptativas para aplicaciones de gestión administrativa. Para ello, se ha diseñado un modelo de usuario basado en estereotipos, para personalizar las interfaces sobre la base del conocimiento semántico y sintáctico, propuesto en el modelo SSOA (Shneiderman y Plaisant, 1998). También se está trabajando en el diseño y desarrollo de bibliotecas digitales, basadas en la tecnología de agentes inteligentes, aplicadas a entornos académicos y universitarios que cuenten con un servicio de consultas personalizadas acordes a los interés y preferencias demostradas por los usuarios.

Desarrollos de SAIs en comercio electrónico. En este dominio, la mayoría de los sitios web no proporcionan asistencia para ayudar a los consumidores a seleccionar el producto más apropiado. Por lo tanto, en esta línea de investigación se busca personalizar paso a paso la interacción con los distintos usuarios, con el objetivo de incrementar la utilidad del servicio de oferta en comercio electrónico; pues la mayoría de los desarrollos en estos sitios web están orientados a la compra-venta de productos o servicios y son pocos los que introducen solo la oferta en comercio electrónico. Para lograr este cometido se diseñó un modelo de usuario que permite aprender y actualizar los intereses, necesidades y preferencia de los clientes, basado en una estructura de metadatos multinivel (Liu et al., 2001). En este caso se representó el modelo en tres niveles, con dos

categorías de información; actualmente se trabaja en la selección de una técnica de inteligencia artificial que sea la más adecuada para el aprendizaje del modelo propuesto. Además se está trabajando sobre la personalización de sitios web de oferta electrónica, de forma tal que, a través de un sistema de agentes inteligentes sea posible generar y mantener el modelo de usuario y la gestión de la oferta adaptativa.

3. RESULTADOS OBTENIDOS/ ESPERADOS

En la Tabla 1 se presentan los objetivos del subproyecto, los resultados obtenidos en relación a cada objetivo y los indicadores que permiten verificar la consecución de tales resultados.

Tabla1. Objetivos, resultados e indicadores verificables del subproyecto.

OBJETIVOS	RESULTADOS OBTENIDOS	INDICADOR VERIFICABLE
Identificar los rasgos culturales a incluir en los modelos de usuario	Rasgos culturales de los modelos de usuario	(Costaguta, et al, 2006a)
Determinar el estado actual de conocimiento y desarrollo sobre modelos de usuario para SAI	Estado del arte sobre aprendizaje colaborativo	(Costaguta, 2006a)
Diseñar modelos de usuario que contemplen los rasgos culturales identificados, en los diferentes dominios de aplicación	Modelos de usuario basados en rasgos culturales	(Costaguta, et al, 2006b) (Durán, et al, 2006)
Analizar las técnicas de Inteligencia Artificial para llevar a cabo el aprendizaje de los perfiles de usuario.	Estudio comparativo de técnicas de Inteligencia Artificial para el aprendizaje de perfiles de usuario en el ámbito de la educación.	(Duran et al, 2007a) (Costaguta et al, 2007a)
	Estudio comparativo de técnicas de Inteligencia Artificial para el aprendizaje de perfiles de usuario en el ámbito del comercio electrónico.	(Maldonado et al, 2007)

OBJETIVOS	RESULTADOS OBTENIDOS	INDICADOR VERIFICABLE
Construir Sistemas Adaptativos Inteligentes basados en la tecnología de agentes incorporando los modelos de usuario diseñados.	Sistema de detección automática de habilidades colaborativas en ambientes de aprendizaje a distancia.	(Durán et al, 2007b)
	Sistema para entrenamiento personalizado de habilidades colaborativas en ambientes de aprendizaje a distancia.	(Costaguta y Durán, 2006) (Costaguta, 2006b) (Costaguta, 2007)
	Sistema adaptativo de gestión administrativa basado en estereotipos	(Salazar et al, 2008)
	Sistema inteligente para gestión de biblioteca	(Cáceres et al, 2008)
	Sistema adaptativo para ofertas electrónicas	(Maldonado et al, 2008)
	Sistema de generación y mantenimiento de un modelo de estudiante basado en estilos de aprendizaje para entornos de e-learning	(Duran et al, 2008a)
	Diseñar modelos de usuario que contemplen los rasgos culturales (meméticos) identificados, para los diferentes dominios de aplicación.	Modelo de estudiante para entornos de aprendizaje colaborativo.
Modelo de usuario para sistemas de comercio electrónico.		(Maldonado et al, 2007) (Durán et al, 2007c)
Validar los sistemas construidos	Evaluación del sistema de detección automática de habilidades colaborativas en ambientes de aprendizaje a distancia	(Duran 2008a) (Duran et al, 2008b) (Duran, 2008b) (Duran, 2008c)
	Evaluación del sistema para entrenamiento personalizado de habilidades colaborativas en ambientes de aprendizaje a distancia	(Costaguta, 2008a) (Costaguta et al, 2008) (Costaguta, 2008b) (Costaguta, 2008c)
	Evaluación del sistema de generación y mantenimiento de un modelo de estudiante basado en estilos de aprendizaje para entornos de e-learning	(Farias et al, 2008a)
		(Duran et al, 2008c) (Farias et al, 2008b)

4. FORMACIÓN DE RECURSOS HUMANOS

En la Tabla 2 se presenta la formación de recursos humanos en las tres líneas de investigación del subproyecto. En todos los casos se trata de Tesis de grado de la carrera Licenciatura en Sistemas de Información de la facultad de Ciencias Exactas y tecnologías de la Universidad Nacional de Santiago del Estero.

Tabla 2. Tesis de grado elaboradas en el marco del subproyecto.

TÍTULO DE LA TESIS	ALUMNOS	ESTADO DE LA TESIS
Detección de estilos de aprendizaje mediante técnicas de análisis de cluster	Roberto Adrián Farias	Tesis aprobada.
Modelo de Usuario basado en estereotipos para interfaces adaptativas	Ivana Salazar y Carlos Galván.	Anteproyecto de Tesis aprobado. En ejecución.
Modelo de Usuario para ofertas adaptativas en sitios Web de Comercio electrónico	Carlos Fabian Cancino, Gabriel Gustavo Marcos y Carlos Alberto Rodríguez.	Anteproyecto de Tesis aprobado. En ejecución.
Las técnicas de simulación en la validación del software	Cecilia Cristina Lara.	Anteproyecto de Tesis aprobado. En ejecución.
Prototipo de una Biblioteca Académico Digital Inteligente	Berat Cáceres y Ramiro Gurmendi	Anteproyecto de Tesis aprobado. En ejecución.
Detección automática del nivel de conocimiento de estudiantes en entornos de e-learning	Carlos Gustavo Paladea	Anteproyecto de Tesis presentado en evaluación.

5. BIBLIOGRAFÍA

Caceres B., Gurmendi R, Costaguta R. y Ludueña M. (2008), "Un Prototipo de Biblioteca Digital Inteligente Aplicado al Ámbito Académico Universitario", X Workshop de Investigadores en Cs. de la Computación

Costaguta R. (2006a), "Una Revisión de Desarrollos Inteligentes para Aprendizaje Colaborativo Soportado por Computadora" Revista Ingeniería Informática, ISSN-0717-4195, Edición N° 13.

Costaguta R. y Durán E. (2007a), "Minería de Datos para Descubrir Estilos de Aprendizaje", Revista Iberoamericana de Educación, ISSN-1681-5653, N° 42.

Costaguta R. y Maldonado M. (2006a) "Rasgos Culturales Compartidos por Usuarios de Sistemas Informáticos". Journal of Information Systems and Technology

- Costaguta R. (2006a), Management, ISSN- 1807-1775, Vol. 3, No. 1, 2006, p. 68-81, "Modelo Multiagente para Entrenamiento de Habilidades Colaborativas", 2º Simposio Internacional "La Investigación en la Univ.: Experiencias Innovadoras de Investigación Articulada a la Docencia y a la Extensión", Jujuy, Argentina.
- Costaguta R y Durán E., (2006b) "Personalización Basada en Agentes para Sistemas de Comercio Electrónico". Memorias de las II Jor. Reg. de Ciencia y Tecnología de las Fac. de Ingeniería del NOA (CT 06), Vol II, pag.548 – 554.
- Costaguta R. (2008a), "Habilidades de Colaboración Manifestadas por los Estudiantes de Ciencias de la Computación", Revista Nuevas Propuestas Vol. 43-44, Ediciones UCSE., Santiago del Estero, Argentina.
- Costaguta R. y Amandi A. (2008), "Training Collaboration Skills to Improve Group Dynamics", Euroamerican Conference on Telematics and Information Systems. Aracaju, Sergipe, Brasil.
- Costaguta R. (2008b), "Conflictos y habilidades de colaboración en grupos presenciales de estudiantes de Informática", IV Jornadas de Ciencia y Tecnología de las Facultades de Ingeniería del NOA, Sgo. del Estero.
- Costaguta R. (2008c), "Agentes inteligentes para el entrenamiento de habilidades de colaboración en estudiantes", Workshop de Inteligencia Artificial, 3ª Escuela de Postgrado PAV, Mendoza.
- Durán E. y Costaguta R., (2006) "Descubrimiento de Conocimiento para Personalización de Sistemas de Aprendizaje Colaborativo". III Congreso Online. Observatorio para la Cibernética.
- Durán E. y Amandi A. (2007a), "Student Models in Intelligent Learning Systems", ACM Computing Surveys, ISSN 0360-0300. Manuscript ID: CSUR-2007-0051.
- Durán E. y Amandi A. (2007b), "Detecting Collaborative Skills in a Distance Learning Environment", Journal Interactive Learning Environment, ISSN 1049-4820. Ed. Taylor and Francis Group.
- Durán E., Costaguta R., Maldonado M. y Unzaga S. (2007c), "Sistemas Adaptativos Inteligentes", IX Workshop de Investigadores en Ciencias de la Computación, Trelew, Chubut.
- Durán E. (2007), "Modelo de Estudiante para Personalización de Sistemas de Aprendizaje Colaborativo", 2º Simposio Internacional "La Investigación en la Univ., Jujuy, Argentina.
- Durán E. y Costaguta R. (2008a), "Una experiencia de enseñanza de la Simulación adaptada al Estilo de Aprendizaje de los Estudiantes", Rev. Internacional Formación Universitaria, Año 1, Edición 1, pag. 19-28, ISSN-0718-5006.
- Durán E. y Amandi A. (2008b), "Collaborative Profile to Support Assistance in CSCL Environment", Euro-American Conference on Telematics and Information Systems (EATIS-2008), Aracaju, Sergipe, Brasil.
- Durán E., Figueroa S., Farias R. (2008c), "Estilos de Aprendizaje en Entornos Virtuales: Descripción de una Experiencia", 1º Jornadas de Educación a Distancia del NOA, San Fernando del Valle de Catamarca.
- Durán E. (2008a), "Evaluación de Perfiles de Estudiante con Técnicas de Simulación", Revista Nuevas Propuestas, Vol. 43-44, 31-54, ISSN 0327-7437. Ediciones UCSE, Santiago del Estero, Argentina.
- Durán E. (2008b), "Las técnicas de simulación en la evaluación de sistemas de aprendizaje adaptativos", IV Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, Santiago del Estero.
- Durán E. (2008c), "Personalización en Sistemas de Aprendizaje Colaborativo", Workshop de Inteligencia Artificial, la 3ª Escuela de Postgrado PAV, Mendoza.
- Durán E. y Amandi A. (2009). Personalised collaborative skills for student models. *Interactive Learning Environments*, 1-20, iFirst Article, Taylor & Francis Group.
- Farias R., Durán E., Figueroa S. (2008a), "Las técnicas de clustering en la personalización de sistemas de e-learning", X Workshop de Investigadores en Cs de la Computación, General Pico, La Pampa.
- Farias R., Durán E., Figueroa S., (2008b), "Detección de Estilos de Aprendizaje mediante Técnicas de Clustering", X Workshop de Investigadores en Cs. de la Computación, General Pico, La Pampa.
- Felder R.M. AND Silverman L.K., (1988). Learning and teaching styles in engineering education. *Journal of Engineering Education* 78(7), 674-681.
- Liu D., Lin Y., Chen Ch.y Huang Y.,(2001), "Deployment of personalized e-catalogues: An agent-based framework integrated with XML metadata and user models", *Journal of Network Computer Applications*, Vol 24, 201-228.
- Maldonado M., Unzaga S., (2007), "Un Modelo de Usuario Basado en Metadatos Multinivel para Sistemas de Comercio Electrónico", 2º Simposio Internacional "La Investigación en la Universidad, Jujuy, Argentina
- Maldonado M., Unzaga S., (2008), "Un Modelo de Usuario para personalizar la oferta en Comercio Electrónico", X Workshop de Investigadores en Cs de la Computación (WICC). Gral Pico, La Pampa,
- Salazar N., Carlos Galvan, Elena Durán, Ivana Harari, (2008), "Modelos de Usuario Basados en estereotipos para Interfaces Adaptativas", X Workshop de Investigadores en Cs. de la Computación, La Pampa.
- Schneiderman Ben, (1998), "Designing the User Interface: Strategies for Effective Human-Computer Interaction". 3ra Edición. Addison-Wesley.

Desarrollo de Agentes Inteligentes Basados en Argumentación para Robótica Móvil

Andrea Cohen **Mariano Tucac** † **Sebastian Gottifredi** †
Alejandro J. Garcia † **Guillermo R. Simari**

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,
† Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina
Tel: (0291) 459-5135 / Fax: (0291) 459-5136
e-mail: {ac, mt, sg, ajg, grs}@cs.uns.edu.ar

Contexto

Esta línea de investigación se realizará dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial, y está asociada a los siguientes proyectos de investigación:

“Programación en lógica distribuida con primitivas de sincronización, comunicación y ejecución en paralelo de agentes, para desarrollo de sistemas multi-agente en ambientes dinámicos” Código: PIP 5050. Director: Alejandro Javier García. Financiamiento: CONICET (Proyecto de Investigación Plurianual).

“Representación de conocimiento, argumentación y apoyo a la toma de decisiones: Herramientas inteligentes para la web semántica” (código 24/N023). Inicio 01/01/2007 Finalización: 31/12/2009. Acreditado en el Programa de Incentivos. Director: Simari, Guillermo Ricardo.

Resumen

El objetivo principal de esta línea de investigación es el estudio de métodos de representación de conocimiento para agentes de software inteligentes deliberativos. En particular, este trabajo se enfocará en la implementación de una arquitectura de agentes BDI basada en mecanismos argumentativos. Por lo tanto, los componentes mentales de estos agentes se especificarán mediante Programación en Lógica Rebatible. El propósito de esta arquitectura es el desarrollo de agentes con capacidades de razonamiento de alto nivel, para controlar robots móviles que juegan al fútbol.

Palabras clave: Argumentación, Sistemas Multi-Agente, Robótica Móvil

1. Introducción

Un agente es una entidad computacional autónoma que obtiene percepciones de su entorno a través de sensores y actúa en dicho entorno utilizando efectores. Decir que es autónoma implica que tiene algún tipo de control sobre su propio comportamiento y que puede actuar sin la intervención de otros agentes o humanos. Actualmente los agentes tienen un campo de aplicación muy amplio,

y existe una gran variedad de tipos de agentes con distintos dominios de aplicación. En particular, los agentes inteligentes están reconocidos en la literatura por su importancia en la resolución de problemas complejos [20]. Este tipo de agentes suelen estar caracterizados a través de componentes mentales como creencias, deseos, compromisos o intenciones.

En la mayoría de los casos los agentes no existen por sí solos, sino que participan de un Sistema Multi-Agente (SMA) o Multi-Robot (SMR). En cualquiera de estos casos, varios agentes interactúan para conseguir algún objetivo o realizar alguna tarea común [19]. En estos sistemas cada agente tiene información incompleta y capacidades limitadas, el control del sistema es distribuido, los datos están descentralizados y la computación es asincrónica. Además los agentes se desenvuelven en un entorno dinámico, el cual no puede predecirse y se ve afectado por las acciones llevadas a cabo por los agentes y por humanos. Otro aspecto relevante de estos sistemas es que, por lo general, la interacción entre los agentes para fines colaborativos, es efectuada mediante mecanismos de comunicación.

Cada agente inteligente generalmente necesita contar con una base de conocimiento para almacenar la información referida al entorno y a sus componentes mentales. El conocimiento allí almacenado debe ser actualizado periódicamente. Asimismo, debe respetar ciertas restricciones de integridad para garantizar el correcto razonamiento del agente. En un entorno multi-agente el conocimiento puede estar distribuido, fragmentado, y/o replicado. Existen en la literatura numerosos formalismos de representación de conocimiento y razonamiento. En particular, en nuestro laboratorio (LIDIA) se han desarrollado técnicas que permiten representar conocimiento tentativo y posiblemente contradictorio, y métodos basados en argumentación para razonar acerca de él. Puntualmente, se cuenta con un formalismo de Programación en Lógica Rebatible [8] que permitiría implementar agentes inteligentes deliberativos para aplicaciones concretas.

Dentro del conjunto de posibles aplicaciones, el fútbol de robots representa una alternativa interesante para poner en práctica teorías desarrolladas para agentes inteligentes [13]. Este dominio incluye agentes con características autónomas, cooperativas y competitivas, los cuales pueden razonar, aprender y revisar sus conocimientos.

En la actualidad, aún es necesario contar con herramientas declarativas para la implementación de agentes inteligentes [6, 9, 10, 11, 3]. Estas herramientas deben proporcionar mecanismos que permitan especificar los agentes a través de sus componentes mentales. Además, deben proveer formalismos sofisticados de alto nivel para representación de conocimiento y razonamiento.

El objetivo de esta línea de investigación es desarrollar agentes inteligentes que controlan robots que juegan al fútbol. Estos pueden interactuar con otros agentes en forma colaborativa, y de esta manera llevar a cabo tareas en forma cooperativa. Dada la complejidad de este dominio de aplicación, es necesario que los agentes consideren diversos factores al momento de decidir sus acciones. Por ello, es importante contar con un mecanismo que permita especificar de manera declarativa estas decisiones. Además, tanto las creencias como las decisiones potenciales podrían ser conflictivas. En consecuencia, en este trabajo se buscará una aproximación argumentativa para implementar este tipo de agentes.

2. Líneas de Investigación y Desarrollo

Esta línea de investigación busca combinar resultados teóricos obtenidos en las áreas de arquitecturas de agentes y mecanismos de representación de conocimiento. El objetivo es aplicarlos en el desarrollo de agentes para controlar robots que juegan al fútbol. En particular, se empleará una arquitectura de agentes basada en las teorías Belief-Desire-Intentions (BDI). Como mecanismo de representación de conocimiento de dichos agentes se utilizará Defeasible Logic Programming (DeLP), una herramienta de argumentación rebatible concreta.

2.1. Agentes Inteligentes

Las teorías BDI proporcionan arquitecturas para el desarrollo de agentes inteligentes [4, 16]. En la actualidad, el formalismo BDI es uno de los más promisorios [14]. Este framework combina dos aspectos importantes del área de agentes inteligentes: contempla teoría de agentes, y representa una arquitectura práctica de agentes. Es decir, además de ser una reconocida alternativa de teoría de razonamiento para agentes basados en lógica, representa una arquitectura lógica para la construcción de agentes capaces de razonar acerca de su entorno y deliberar acerca de sus acciones. Un agente BDI posee componentes mentales para representar Creencias (Beliefs), Deseos (Desires) e Intenciones (Intentions). Estos componentes son empleados para determinar el estado mental del agente. Por lo tanto, un agente BDI está compuesto por un conjunto de creencias que contiene información acerca del entorno en cual se desenvuelve, un conjunto de deseos que representan las motivaciones u objetivos que el agente utilizará para actuar proactivamente, y un conjunto de intenciones que contiene todas las posibles acciones que el agente puede realizar para llevar a cabo sus deseos.

2.2. Sistemas Argumentativos

Dentro de la teoría de representación de conocimiento, los formalismos argumentativos son reconocidos por la construcción y comparación de argumentos para razonar con información inconsistente [5]. En este tipo de formalismos, una conclusión es aceptada o rechazada según el análisis de los argumentos a su favor y en su contra. Varios trabajos en la literatura reconocen la importancia del uso de argumentación como un mecanismo de razonamiento en los sistemas de agentes inteligentes [2],[15],[1],[17]. En particular, esta línea está vinculada con un sistema argumentativo concreto, Defeasible Logic Programming (DeLP) [8]. DeLP es un formalismo que combina los resultados de programación en lógica y argumentación rebatible. Utilizando DeLP el desarrollador de agentes podrá representar información tentativa de forma declarativa por medio de reglas “débiles”. Dado que es posible utilizar negación estricta en la cabeza de estas reglas, será posible representar información contradictoria. En este formalismo se identifican aquellos elementos en contradicción y se lleva a cabo un proceso argumentativo de dialéctica para decidir cuál de estos elementos prevalecerá.

3. Resultados Obtenidos/Esperados

Anteriormente, en esta línea de investigación se ha desarrollado una arquitectura de agentes para controlar robots que juegan al fútbol [12, 18]. Esta arquitectura fue construida en base a un sistema de capas. Este sistema provee el nivel de abstracción necesario para que en los niveles superiores el agente pueda efectuar razonamiento cognitivo. El diseño por capas permite la abstracción y modularización de los diferentes aspectos de un dominio de aplicación complejo, tal como lo es el fútbol de robots. En particular, el último nivel de abstracción fue implementado mediante una arquitectura BDI. El mecanismo de representación de conocimiento y razonamiento está basado en el modelo de programación en lógica. La programación en lógica representa una herramienta flexible para implementar agentes, brindando un diseño declarativo y posibilitando un rápido desarrollo de los mismos utilizando sistemas de razonamiento. Este diseño declarativo permite especificar las estrategias de los agentes de forma tal que pueda seleccionarse una de ellas de forma elegante, para luego ser ejecutada.

También se ha desarrollado DeLP-Server [7] como parte de esta línea de investigación. Esto se debe a las ventajas de DeLP y a la necesidad de una herramienta declarativa flexible de razonamiento en sistemas multi-agente. Un DeLP-Server es una implementación concreta de DeLP que brinda servicios de razonamiento argumentativo en sistemas multi-agente. Cada servidor es un programa stand-alone que puede interactuar con múltiples agentes. Estos servidores pueden almacenar pro-

gramas DeLP conocidos como públicos (o comunes). Un agente cliente puede enviar consultas a algún servidor y recibir la respuesta correspondiente. Para cada respuesta recibirá explicaciones, las cuales consisten en el análisis argumentativo de la respuesta. Un DeLP-Server puede ser consultado por varios agentes, y un agente en particular puede consultar varios servers. Las respuestas enviadas por los DeLP-Servers siguen el formato de DeLP clásico [8]: “yes” si la información consultada está garantizada, “no” si el complemento de la información consultada está garantizado, o “undecided” en cualquier otro caso.

En la actualidad, esta línea de investigación tiene como principal objetivo el desarrollo de una implementación concreta de sistemas multi-agente basados en razonamiento argumentativo para controlar robots móviles. Se partirá de la arquitectura presentada en [12, 18] y se buscará integrar DeLP-Servers a los componentes mentales de cada agente. Esto es, cada componente mental Creencias, Deseos, Intenciones estará conectado a un servidor específico que le proveerá soporte de razonamiento argumentativo. El objetivo de tal integración, es brindar a los agentes que controlan los robots una herramienta más sofisticada de razonamiento y representación de conocimiento. Se deberán considerar mecanismos formales para actualizar los diferentes componentes mentales y establecer claramente el vínculo entre ellos. Además, se intentará añadir formalismos de comunicación a la arquitectura. El propósito es que tales formalismos se encuentren directamente vinculados con los componentes mentales, aprovechando las ventajas provistas por los DeLP-Servers. Esta implementación se llevará a cabo en el dominio de fútbol de robots presentado en [12, 18].

Referencias

- [1] Leila Amgoud, Caroline Devred, and Marie-Christine Lagasquie. A constrained argumentation system for practical reasoning. In *In 7th International Joint Conference on Autonomous Agents and Multi-Agents Systems, AAMAS, 2008*.
- [2] Trevor J. M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artif. Intell.*, 171(10-15):619–641, 2007.
- [3] Rafael Bordini, Lars Braubach, Mehdi Dastani, Amal El Fallah Seghrouchni, Jorge Gomez-Sanz, Joao Leite, Gregory O’Hare, Alexander Pokahr, and Alessandro Ricci. A survey of programming languages and platforms for multi-agent systems. In *Informatica 30*, pages 33–44, 2006.
- [4] M. Bratman, D. Israel, and M. Pollack. Plans and resource-bounded practical reasoning. In *Philosophy and AI: Essays at the Interface*.
- [5] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.
- [6] Michael Fisher, Rafael Bordini, Benjamin Hirsch, and Paolo Torroni. Computational logics and agents: A road map of current technologies and future trends. *Computational Intelligence*, 23(1):61–91, February 2007.
- [7] A. García, N. Rotstein, M. Tucacat, and G. Simari. An argumentative reasoning service for deliberative agents. In *KSEM 2007*.
- [8] A. Garcia and G. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4(1-2):95–138, 2004.

- [9] S. Gottifredi and A. Garcia. Análisis de lenguajes de implementación de agente. In *Workshop de Investigadores en Ciencias de la Computación (WICC 2006)*, 2006.
- [10] Sebastian Gottifredi and Alejandro J. Garcia. Análisis lenguajes de especificación de agente en robótica móvil. In *9vo. Workshop de Investigadores en Ciencias de la Computación (WICC 2007)*, pages 27–31, 2007.
- [11] Sebastian Gottifredi, Alejandro J. Garcia, and Guillermo R. Simari. Agent programming using defeasible argumentation for knowledge representation and reasoning. In *13vo. Congreso Argentino de Ciencias de las Computación (CACIC 2007)*, pages 1464–1475, 2007.
- [12] Sebastian Gottifredi, Mariano Tucát, Daniel Corbatta, Alejandro J. García, and Guillermo R. Simari. A BDI Architecture for High Level Robot Deliberation. In *Proceedings of XIV Argentine Congress of Computer Science*, La Rioja, Argentina, 2008. Universidad Nacional de Chilecito.
- [13] Kwun Han and Manuela Veloso. Automated robot behavior recognition applied to robotic soccer. *Robotics Research: the Ninth International Symposium*, pages 199–204. Springer-Verlag, London, 2000. Also in the Proceedings of IJCAI-99 Workshop on Team Behaviors and Plan Recognition.
- [14] V. Mascardi, D. Demergasso, and D. Ancona. Languages for programming BDI-style agents: an overview. In F. Corradini, F. De Paoli, E. Merelli, and A. Omicini, editors, *Proceedings of WOA 2005: Dagli Oggetti agli Agenti. 6th AI*IA/TABOO Joint Workshop “From Objects to Agents”*, pages 9–15. Pitagora Editrice Bologna, 2005.
- [15] Iyad Rahwan and Leila Amgoud. An argumentation based approach for practical reasoning. In *AAMAS*, 2006.
- [16] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In James Allen, Richard Fikes, and Erik Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR’91)*, pages 473–484. Morgan Kaufmann publishers Inc.: San Mateo, CA, USA, 1991.
- [17] N. D. Rotstein, A. J. Garcia, and G. R. Simari. Reasoning from desires to intentions: A dialectical framework. In *AAAI*, 2007.
- [18] Mariano Tucát, Sebastian Gottifredi, Federico Vidaurreta, Alejandro J. García, and Guillermo R. Simari. A Layered Architecture using Schematic Plans for Controlling Mobile Robots. In *Proceedings of XIV Argentine Congress of Computer Science*, La Rioja, Argentina, 2008. Universidad Nacional de Chilecito.
- [19] Michael Wooldridge. Intelligent agents. In Gerhard Weiss, editor, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, chapter 1, pages 27–78. The MIT Press, Cambridge, MA, USA, 1999.
- [20] Michael Wooldridge and Nicholas R. Jennings. Intelligent agents: Theory and practice. *Knowledge Engineering Review*, 10(2):115–152, 1995.

Tópicos avanzados en categorización de textos

Marcelo Errecalde,[†] Diego Ingaramo,[†] M. Verónica Rosas,[†] Amparito Asensio

[†]Laboratorio de Investigación y Desarrollo en
Inteligencia Computacional (LIDIC)¹
Departamento de Informática
Universidad Nacional de San Luis
Ejército de los Andes 950 - Local 106
(D5700HHW) - San Luis - Argentina
Tel: (02652) 420823 / Fax: (02652) 430224
e-mail: {merreca, daingara, mvrosas,
aaasensi}@unsl.edu.ar

Resumen

Contexto: Este artículo describe, en forma resumida, los trabajos de investigación y desarrollo que se están llevando a cabo en la línea “Agentes y Sistemas Inteligentes” del LIDIC, en el área de *categorización de textos*. Otras líneas de investigación del LIDIC, también abordan problemas de categorización pero, en nuestro caso, nos centramos en problemas que involucran *documentos*. Por este motivo, en nuestra línea se presta especial atención a técnicas vinculadas al procesamiento del lenguaje natural, la lingüística computacional y la recuperación de la información. En este sentido, buena parte de los desarrollos en estos temas, se han realizado en forma conjunta con grupos de investigación con una experiencia considerable en el procesamiento del lenguaje natural, como por ejemplo, el NLEL² de la Universidad Politécnica de Valencia, España.

Los enfoques utilizados en nuestra línea de trabajo, buscan mejorar los procesos de categorización automática de textos en base a dos mecanismos principales: 1) el uso de técnicas de representación de textos más elaboradas, 2) el uso de algoritmos de categorización más eficientes y efectivos. Respecto al primer punto, nuestros trabajos incluyen el uso de representaciones que incorporan información semántica (conceptos) a los

métodos tradicionales basados en términos y representaciones basadas en LSI (Latent Semantic Indexing). Las soluciones algorítmicas por su parte, incluyen el ensamblaje de clasificadores y los métodos de optimización bio-inspirados.

1. Introducción

Las tecnologías de la información y la comunicación han hecho disponible hoy en día, una gran cantidad de información. Gracias a Internet, es posible conseguir, sin demasiado esfuerzo, casi cualquier información que se desee en pocos segundos. Sin embargo, también es común recibir mucha más información que la que se desea (o se puede procesar), y somos inundados diariamente por una cantidad creciente de mensajes de e-mail, SMSs, reportes internos, faxes, llamadas telefónicas, noticias, artículos de revistas, páginas Web, etc [12]. Un aspecto importante a ser observado en este contexto, es que gran parte de esa información es de tipo textual, razón por la cual, todos los aspectos vinculados al procesamiento y organización automática de documentos (recuperación, categorización, agrupamiento, etc) adquieren, día a día, una relevancia creciente.

En el caso particular de la categorización automática de textos, ésta ha despertado un interés significativo, debido a la creciente disponibilidad de documentos en formato digital no sólo a través de Internet y librerías digitales sino también por el hecho de que las empresas organizan mayoritariamente sus escritos y correspondencia a partir de estos formatos. Dada la necesidad imperiosa de su organización y mantenimiento, es común hoy en día, encontrar innumerables aplicaciones de categorización de documentos que abarcan diversas áreas como la detección de “spams” [11, 22], filtrado de noticias [1], detección de plagios e identificación de autores [19, 24] análisis de opinión [2, 6], organización de patentes en categorías [18], y clasificación y organización de páginas Web [21] entre otras.

A partir del auge de esta línea de investigación,

¹Las investigaciones realizadas en el LIDIC son financiadas por la Universidad Nacional de San Luis y por la Agencia Nacional de Promoción Científica y Tecnológica.

²Natural Language Engineering Lab., Departamento de Sistemas Informáticos y Computación (DSIC). Universidad Politécnica de Valencia, Valencia, España.

como una de las tareas del procesamiento del lenguaje natural, surge la necesidad de incorporar y plantear nuevas mejoras a dicho proceso, que permitan mayor exactitud al momento de seleccionar la categoría a la que pertenece un determinado texto.

Los enfoques para la construcción automática de clasificadores, pueden ser agrupados en dos grandes áreas: la categorización *supervisada* y la *no supervisada*. La categorización supervisada utiliza un proceso inductivo general basado en el conocimiento que se tiene de: a) las categorías y b) ejemplos de documentos categorizados por un experto. En la categorización no supervisada (agrupamiento) en cambio, no se conocen a priori las categorías ni asignaciones correctas de categorías (sólo se conoce alguna medida de similitud).

A la hora de realizar mejoras en cualquiera de estos procesos de categorización, los enfoques pueden ser agrupados en dos grandes categorías: a) aquellos que buscan mejorar la *representación* de los documentos, y b) los que se centran en obtener mejoras en los *algoritmos* de categorización. Los primeros, se centran en tratar de capturar en las estructuras utilizadas para representar los documentos, tanta información relevante como sea posible, que pueda ser útil para determinar la categoría de los documentos. Los segundos en cambio, intentan desarrollar algoritmos que hagan un uso eficiente de esa información, tratando de encontrar soluciones de calidad y en forma eficiente.

Actualmente, en nuestra línea de investigación se están abordando cuatro temas principales: 1) agrupamiento de documentos cortos sobre temáticas relacionadas, 2) agrupamiento de documentos cortos multilingüe, 3) categorización semántica de documentos y 4) ensamblaje de clasificadores.

En estos temas, en algunos casos se hace hincapié en los aspectos de representación y en otros casos en los aspectos algorítmicos, como se puede observar en la Tabla 1. A modo de ejemplo, mientras que en la categorización semántica se ha puesto especial énfasis en enriquecer los enfoques tradicionales basados en el modelo vector (*representación*), en el agrupamiento de textos cortos los esfuerzos se han concentrado en la definición de métodos de clustering bio-inspirados (algoritmos).

El resto de este trabajo se organiza de la siguiente manera. En la sección 2, se describen brevemente los cuatro temas de investigación referi-

dos previamente. En la sección 3 se describen algunos de los resultados obtenidos en estos temas y posibles trabajos futuros. Finalmente, la sección 4 menciona los distintos trabajos de tesis que se han desarrollado o están en ejecución, referidos a estos temas.

2. Líneas de Investigación y desarrollo

A continuación, se detallan los ejes de los temas que se están investigando, de acuerdo a los criterios especificados en la sección previa.

2.1. Agrupamiento de documentos muy cortos

El agrupamiento de documentos es la asignación de documentos a categorías desconocidas. Esta tarea es más difícil que la categorización supervisada debido a que no se dispone con anticipación, de ninguna información sobre las categorías y clasificaciones correctas de los documentos. Cuando el agrupamiento involucra textos muy cortos, la tarea es aún más difícil debido a las bajas frecuencias de ocurrencia de los términos en los documentos. No obstante esto, el trabajo de investigación relacionado al agrupamiento de textos cortos es muy importante, en especial si consideramos la tendencia actual/futura de la gente a usar “lenguajes pequeños”, por ejemplo en blogs, mensajes de textos, e-mails, snippets, etc. Las aplicaciones potenciales en diferentes áreas del procesamiento del lenguaje natural incluyen el re-ranking de snippets en recuperación de información y el agrupamiento automático de textos científicos disponibles en la Web.

Nosotros estamos interesados en analizar este tipo de colecciones para desarrollar técnicas novedosas que puedan ser usadas para mejorar los resultados obtenidos con técnicas de agrupamiento clásicas. Este análisis ha incluido el estudio de la correlación entre medidas de validez de clustering internas y externas, la relación de estas medidas con la dificultad inherente de las colecciones, y el desarrollo de técnicas de clustering basadas en optimización mediante enfoques bio-inspirados.

	Representación	Algoritmos
Agrupamiento textos cortos - temas relacionados		X
Agrupamiento textos cortos - multilingüe	X	X
Categorización semántica de documentos	X	
Ensamblaje de clasificadores		X

Tabla 1: Temas abordados - Aspecto enfatizado

2.2. Agrupamiento de documentos cortos multilingüe

En el clustering multilingüe, se tiene una colección con documentos en distintos lenguajes y la idea es lograr agrupamientos donde en cada grupo puede haber documentos relacionados pero de distintos lenguajes. El principal problema en este caso, es la definición de una medida de similitud entre documentos escritos en distintos lenguajes. Esto implica en general, que los documentos deban ser preprocesados para hacerlos comparables. En general existen dos formas básicas para lograr esto: 1) mediante tecnologías de traducción, 2) mediante la traducción de los documentos a una representación independiente del lenguaje.

En nuestro caso, no estamos realizando agrupamiento multilingüe estricto, sino que realizamos agrupamiento *asistido* por información multilingüe. La idea intuitiva es que si se dispone de corpus de textos paralelos en lenguaje dual (por ejemplo, en inglés y español) esta información puede servir para lograr un agrupamiento más efectivo que en el caso en que se trabaja con los documentos en cada uno de estos lenguajes por separado

2.3. Categorización semántica de documentos

La incorporación de información semántica en el indexado de documentos, es una idea que ha sido considerada en distintas áreas del Procesamiento del Lenguaje Natural, reportándose buenos resultados con este enfoque en tareas de recuperación de la información, agrupamiento y categorización de documentos. La idea principal en estos enfoques, es enriquecer las tradicionales representaciones basadas en el modelo de espacio vector, donde cada texto es representado por un vector de n -términos, siendo n el número de términos que aparecen en la colección de documentos. En los enfoques basados en información semántica en cambio, la idea es tomar en cuenta un indexado basado en los *conceptos* o *significados* de los términos que aparecen en los documentos, y reducir así los problemas introducidos por los

fenómenos de *sinonimia* y *polisemia*.

Una componente fundamental en estos enfoques semánticos, es el proceso encargado de determinar en forma no ambigua, cual es el sentido (o significado) de una palabra en un contexto particular, denominado *Word Sense Disambiguation* (WSD). Si bien se han propuesto las más diversas técnicas para la tarea de WSD, podemos diferenciar en este sentido dos grandes enfoques: WSD supervisada y WSD no supervisada. Otro aspecto fundamental, es el impacto que tienen los procesos de reducción de vocabulario necesarios para reducir la dimensionalidad de los vectores usados para representar los documentos, y de que manera esta reducción se relaciona con los nuevos conceptos introducidos en la representación.

Este último aspecto es el eje principal de nuestra investigación en este tema, el cual intenta realizar aportes en el área de la categorización de textos cortos, planteando posibles mejoras a partir de la incorporación de información semántica y el uso de técnicas de reducción de vocabulario. En este sentido, se hace especial hincapié en el uso de técnicas de desambiguación no supervisadas para la obtención de los conceptos. También se analizan las distintas herramientas de la Ingeniería de Software que pueden ser útiles en el desarrollo de este tipo de aplicaciones y cual es la factibilidad de la aplicación de este tipo de técnicas en problemas complejos de la vida real, como la categorización automática de noticias.

2.4. Ensamblaje de clasificadores

La idea de considerar todas las hipótesis disponibles para resolver un problema y construir una hipótesis de más alto nivel, es un enfoque conocido en los sistemas de decisión donde es común tomar en cuenta las opiniones de varios asesores en lugar de restringirse a uno solo. Este principio ha sido aplicado en distintos métodos de minería de datos que combinan varias hipótesis en un único modelo. La idea en este caso, se basa en combinar las predicciones realizadas por las distintas hipótesis disponibles (normalmente

por votación) para clasificar ejemplos o hacer regresión sobre los ejemplos. La precisión obtenida de esta manera supera, generalmente, la precisión de cada componente individual del conjunto. A estas técnicas de combinación de hipótesis se las suele referenciar como *modelos combinados* o *métodos de ensamblaje de modelos* [5] e incluyen técnicas como *bagging* [3] y *boosting* [10].

Los sistemas de ensamblaje están fundamentados matemáticamente pero además existe una fuerte componente psicológica surgida de nuestra experiencia de la vida diaria: los usamos todo el tiempo, cada vez que recurrimos a las opiniones de otros individuos (o *expertos* en una materia) para tomar una decisión. Nuestro objetivo en este caso, es mejorar la confianza en que estamos tomando la decisión correcta ponderando las distintas opiniones y combinándolas a través de algún proceso de pensamiento para alcanzar una decisión final.

En nuestro caso, el rol del experto es cumplido por un *clasificador*, por lo que el problema puede ser replanteado como un problema de clasificación donde cada clasificador hace una hipótesis sobre la clasificación de una instancia de datos dada, dentro de un conjunto predefinido de categorías que representan distintas decisiones. La decisión estará basada en el entrenamiento previo del clasificador usando un conjunto representativo de datos de entrenamiento para el cual las decisiones correctas son conocidas a priori.

Si bien los métodos de ensamblaje ya tienen una trayectoria considerable en tareas de clasificación general dentro de la minería de datos, su uso particular en tareas de categorización de textos es relativamente reciente [23, 20]. Este último aspecto es el eje principal de nuestro trabajo, cuyo objetivo principal es realizar una primera aproximación al problema general del uso de técnicas de ensamblaje de clasificadores en la categorización de textos. En este contexto, un objetivo parcial a cumplir será el estudio y análisis de las distintas estrategias para lograr una *diversidad* adecuada en los clasificadores individuales, un aspecto que suele ser clave para el éxito de la aplicación de estas técnicas. En este sentido, nuestro enfoque para lograr la diversidad de clasificadores estará dado por el uso de distintas codificaciones de documentos y distintas técnicas para seleccionar las palabras claves utilizadas en la codificación de los documentos.

3. Resultados esperados/obtenidos

Los principales resultados en nuestra línea de trabajo se han realizado en el agrupamiento de documentos muy cortos [14, 17, 7, 4, 15, 13] y la categorización semántica de documentos [8, 9]. En el primer caso, los trabajos futuros están orientados a extender los buenos resultados obtenidos con enfoques PSO [4, 13], a otros enfoques bio-inspirados, como por ejemplo el propuesto en [16]. En el segundo caso, se pretende realizar un estudio similar al realizado en [8, 9] pero utilizando ahora distintos enfoques de WSD no supervisada.

Para el agrupamiento de documentos cortos multilingüe, se prevee comparar enfoques basados en LSI para construir un espacio semántico multilingual, con enfoques que trabajan con los documentos separados y que utilizan medidas de similitud basadas en la correspondencia entre los documentos en el corpus paralelo. En el ensamblaje de clasificadores por su parte, los resultados preliminares han permitido observar resultados competitivos con los obtenidos por métodos clásicos como SVM, Naive Bayes y *k*-nn en experimentos con las colecciones de documentos más conocidas. Actualmente, nuestros experimentos se han basado en los algoritmos AdaBoost y Bagging, proyectándose en el futuro hacer uso de enfoques basados en la teoría de Dempster-Shafer.

4. Formación de recursos humanos

Trabajos de tesis vinculados con las temáticas descriptas previamente:

- 1 tesis doctoral en ejecución (co-dirección con investigador del NLEL (UPV))
- 1 tesis de maestría en ejecución (co-dirección con investigador del NLEL (UPV))
- 1 tesis de Licenciatura aprobada.
- 1 tesis de Licenciatura en ejecución.

Referencias

- [1] G. Amati, D. D'Áloisi, V. Giannini, and F. Ubal dini. A framework for filtering news and managing distributed data. *Journal of Universal Computer Science*, pages 1007–1021, 1997.
- [2] A. Balahur and A. Montoyo. Determining the semantic orientation of opinions on products- a comparative analysis. *Procesamiento del Lenguaje Natural*, pages 201–208, 2008.

- [3] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–146, 1996.
- [4] L. Cagnina, M. Errecalde, D. Ingaramo, and P. Rosso. A discrete particle swarm optimizer for clustering short-text corpora. In *Proc. of the 3rd International Conference on Bioinspired Optimization Methods and their Applications (BIOMA08)*, pages 93–103. 2008.
- [5] T. G. Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2):139–157, 2000.
- [6] X. Ding, B. Liu, and P. Yu. A holistic lexicon-based approach to opinion mining. In *Proceedings of the international conference on Web search and web data mining*, pages 231–240, Palo Alto, California, USA, 2008.
- [7] M. Errecalde, D. Ingaramo, and P. Rosso. Proximity estimation and hardness of short-text corpora. In *Proc. 5th International Workshop on Text-based Information Retrieval (TIR-2008)-Dexa 2008*, pages 15–19. IEEE Computer Society, 2008.
- [8] E. Ferretti, M. Errecalde, and P. Rosso. The influence of semantics in text categorisation: A comparative study using the k nearest neighbours method. In *Proc. of the 2nd Indian International Conference on Artificial Intelligence (IICAI)*, 2005.
- [9] E. Ferretti, M. Errecalde, and P. Rosso. Does semantic information help in the text categorisation task? *Journal of Intelligent Systems*, 17(1–3), 2008.
- [10] Y. Freund and R. E. Schapire. Experiments with a new boosting algorithm. In *Proceedings of the 13th International Conference on Machine Learning, ICML'96*, pages 148–156, 1996.
- [11] D. Heckerman, M. Sahami, S. Dumais, and E. Horvitz. A bayesian approach to filtering junk e-mail. In *Proceeding of AAAI-98 Workshop on Learning for Text Categorization*, pages 55–62, Madison, Wisconsin, USA, 1998.
- [12] F. Heylighen. Information overload, complexity and information overload in society: Why increasing efficiency leads to decreasing control. <http://pespmc1.vub.ac.be/papers/info-overload.pdf>, CLEA, Free University of Brussels, Pleinlaan, 2002.
- [13] D. Ingaramo, M. Errecalde, L. Cagnina, and P. Rosso. *Computational Intelligence and Bioengineering*, chapter Particle Swarm Optimization for clustering short-text corpora. KBIES. IOS press, 2009.
- [14] D. Ingaramo, M. Errecalde, and P. Rosso. Medidas internas y externas en el agrupamiento de resúmenes científicos de dominios reducidos. *Procesamiento del Lenguaje Natural*, 39, 2007.
- [15] D. Ingaramo, M. Errecalde, and P. Rosso. Density-based clustering of short-text corpora. *Procesamiento del Lenguaje Natural*, 41:81–87, 2008.
- [16] D. Ingaramo, G. Leguizamón, and M. Errecalde. Adaptive clustering with artificial ants. *Journal of Computer Science & Technology*, 5(4):264–271, 2005.
- [17] D. Ingaramo, D. Pinto, P. Rosso, and M. Errecalde. Evaluation of internal validity measures in short-text corpora. *LNCS*, 4919:555–567, 2008.
- [18] L. Larkey. A patent search and classification system. In *Proceedings of DL-99, 4th ACM Conference on Digital Libraries*, pages 179–187, Berkeley, CA, USA, 1999.
- [19] R. Lukashenko, V. Graudina, and V. Grundspenkis. Computer-based plagiarism detection methods and tools: an overview. In ACM, editor, *Proceedings of the international conference on Computer systems and technologies, CompSysTech '07*, pages 1–6, New York, NY, USA, 2007.
- [20] A. Montejo and L. A. Ureña. Binary classifiers versus adaboost for labeling of digital documents. In *Procesamiento del Lenguaje Natural*, pages 319–326, 2006.
- [21] H. Oh, S.H. Myaeng, and M.H. Lee. A practical hypertext categorization method using links and incrementally available class information. In *Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval*, pages 264–271, Athens, Greece, 2000.
- [22] G. Sakkis and I. Androutsopoulos. A memory-based approach to anti-spam filtering for mailing lists. *Information Retrieval*, pages 49–73, 2003.
- [23] Robert E. Schapire and Yoram Singer. BoosTexter: A boosting-based system for text categorization. *Machine Learning*, 39(2/3):135–168, 2000.
- [24] B. Stein, M. Koppel, and E. Stamatatos. Plagiarism analysis, authorship identification, and near-duplicate detection. *SIGIR Forum*, 2007.

Integrando capacidades de alto nivel en la toma de decisiones de agentes inteligentes

Marcelo Errecalde[†], Guillermo Aguirre[†], Edgardo Ferretti[†], Federico Schlesinger

[†]Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)¹
Departamento de Informática
Universidad Nacional de San Luis
Ejército de los Andes 950 - Local 106
(D5700HHW) - San Luis - Argentina
Tel: (02652) 420823 / Fax: (02652) 430224
e-mail: {merreca, gaguirre, ferretti}@unsl.edu.ar, fedest@gmail.com

Resumen

Contexto: Este artículo describe, en forma resumida, los trabajos de investigación y desarrollo que se están llevando a cabo en la línea “Agentes y Sistemas Inteligentes” del LIDIC, en el área de agentes cognitivos. A diferencia de otras líneas de investigación del LIDIC, que se centran en problemas de optimización y el uso de enfoques de *soft computing*, nuestra línea se centra en la formalización, diseño y desarrollo de agentes computacionales con capacidades cognitivas de alto nivel, y en especial en enfoques basados en lógica, razonamiento argumentativo y teoría de decisión cualitativa. En este sentido, gran parte de los desarrollos realizados en estos temas se han realizado en forma conjunta con investigadores del LIDIA² que trabajan en temáticas afines.

Los enfoques utilizados en nuestra línea de trabajo, buscan dotar a los agentes inteligentes con las capacidades necesarias para enfrentar problemas complejos de la vida real, que involucran ambientes dinámicos, inciertos, limitados en recursos y parcialmente observables. Incluyen el uso de arquitecturas BDI, la integración con Web Services, formalismos de comunicación de alto

nivel (speech acts) y la toma de decisiones mono y multi-atributo cualitativa. Estos trabajos se basan, mayoritariamente, en procesos de razonamiento argumentativo que permiten enfrentar de manera elegante, situaciones con información incompleta y contradictoria, como por ejemplo, las que se presentan en distintas aplicaciones de robótica y en la comunicación de agentes de software.

1. Introducción

La capacidad de los *agentes inteligentes* [26] de comportarse de manera *flexible (reactiva, proactiva y social)*, los han convertido en herramientas muy interesantes para abordar problemas complejos del mundo real. Estas habilidades, suelen ser fundamentales en el tipo de ambientes en que estamos interesados, donde se debe tratar con información incierta, incompleta, contradictoria y cambiante y donde, además, se debe considerar un contexto con recursos computacionales limitados.

Proveer con estas capacidades a un agente no es una tarea sencilla. Ciertos enfoques, enfatizan las capacidades *reactivas* [8] de los agentes basadas en la información disponible en el ambiente, pero este enfoque suele exhibir serias falencias en ambientes parcialmente observables. En otros casos, la toma de decisiones del agente se basa en enfoques más formales como la teoría de decisión clásica [5], aunque no es claro en estos casos como tratar con el cambio o la incorporación de nuevos criterios en las preferencias del agente, considerando la naturaleza estática de la función de utilidad determinada en la etapa de diseño del agente [14].

Si bien es difícil encontrar soluciones integrales a todos los problemas que debe enfrentar un agente inteligente, se pueden identificar ciertos aspectos que jugarán un rol principal en el buen funcionamiento de cualquier agente exitoso. En primer lugar, la *información relevante* que éste tenga disponible respecto al problema bajo consideración, es un factor importante a la hora de

¹Las investigaciones realizadas en el LIDIC son financiadas por la Universidad Nacional de San Luis y por la Agencia Nacional de Promoción Científica y Tecnológica.

²Laboratorio de Investigación y Desarrollo en Inteligencia Artificial. Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina.

tomar una decisión “correcta”. En consecuencia, un aspecto fundamental es ampliar las capacidades perceptivas del agente para obtener esta información. Por ejemplo, en el caso de los agentes de software, esto puede involucrar el acceso a diversas fuentes de información o la comunicación con otros agentes para intercambiar información que pueda ser útil para resolver su problema.

El acceso a la información relevante, no es condición suficiente para garantizar su uso efectivo por parte del agente. Esta información podrá ser incompleta y contradictoria, y la decisión de cuando adoptarla como nueva creencia justificada, podrá involucrar alguna forma de *razonamiento orientado a las creencias* (o *razonamiento teórico* según [26]), que permita contrastar la nueva información disponible con las creencias previas del agente.

Los pasos previos, sirven de soporte para el proceso esencial de un agente, la *toma de decisiones*, encargado de determinar cual de las acciones disponibles, es la “mejor” acción a realizar en esas circunstancias. Este proceso podrá incluir la ponderación de distintos caminos alternativos de acción, la determinación de cuándo una línea de acción ya no tiene sentido de ser continuada o bien, compatibilizar los distintos criterios que deben ser considerados respecto a las alternativas disponibles. Estas actividades, involucran en general alguna forma de *razonamiento orientado a la acción*, conocido popularmente como *razonamiento práctico* [26].

En nuestra línea de investigación, se atacan estos 3 aspectos fundamentales referidos al diseño de un agente inteligente: 1) el acceso a la *información relevante*, 2) el *razonamiento teórico* con información incompleta y contradictoria y 3) el *razonamiento práctico* necesario para la toma de decisiones en ambientes complejos con recursos computacionales acotados. Respecto al primer punto, en la sección 2.1.1 se describirán brevemente nuestros trabajos para proveer de información a los agentes de software mediante el uso de Web services. La sección 2.1.2 por su parte, explicará nuestra propuesta para aquellos casos en que la información es obtenida mediante la comunicación con otros agentes utilizando *speech acts*. El mecanismo utilizado para razonar con información incompleta y contradictoria es brevemente introducido en 2.2.1. Por último, los aspectos vinculados con el razonamiento práctico basado en BDI son presentados en 2.3.1, y la toma de decisiones basada en múltiples criterios se des-

cribe en 2.3.2. El trabajo concluye con una breve descripción de los resultados esperados y obtenidos en esta línea de investigación (sección 3) y una síntesis de los trabajos de tesis realizados en estos temas (sección 4).

2. Líneas de Investigación y desarrollo

A continuación, se describen las ideas generales de los temas que se están investigando en nuestra línea, de acuerdo a los criterios especificados en la sección previa.

2.1. Accediendo a información desde agentes de software

2.1.1. Agentes y Web services

Las fuentes de información a las que puede recurrir un agente son numerosas y variadas. Entre estas fuentes están los *Web services* (WS) ya que responden a un comportamiento bien definido, ideal para ser utilizado por agentes computacionales actuando como clientes. De esta manera, se combinan dos dominios de aplicación diferentes; por un lado los WS que pueden invocar los servicios de algunos agentes y vice versa, los agentes pueden invocar a los WS.

Intuitivamente, y según la definición provista por WIKIPEDIA, un *servicio web* (en inglés *Web service*) “es un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Distintas aplicaciones de software desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los servicios web para intercambiar datos en redes de ordenadores como Internet. La interoperabilidad se consigue mediante la adopción de estándares abiertos.”

La plataforma JADE [4] provee una buena interconectividad entre ambos dominios (agentes y WS), mediante el WSIG (Web Service Integration Getway). Esto se consigue de una manera casi transparente para el programador de agentes ya que en primer lugar, la interacción se logra mediante el uso de solamente dos clases de mensajes ACL: de información y de requerimiento. En segundo lugar, se pueden utilizar los directorios establecidos como repositorios, a fin de facilitar acciones como registrarse, desregistrarse y modificaciones. Finalmente, quizás el aspecto más interesante, es la posibilidad de presentar

los WS como si fuesen facilidades brindadas por agentes convencionales y, en sentido contrario, los servicios de los agentes como si fuesen WS. Esto se consigue mediante la facilidad de *páginas amarillas* de JADE, que permite que los agentes puedan requerir los servicios de un agente o WS sin notar la diferencia entre ambos.

Más allá de utilizar distintos WS para proveer de información a los agentes, nuestros trabajos también incluyen mejoras en la *identificación* de los WS relevantes. A fin de reducir los inconvenientes que surgen en la especificación de los servicios debido a la falta de un estándar, nuestra línea está trabajando en el uso de *búsquedas por similitud* que permiten acceder a los servicios, de acuerdo a ciertas *características generales* especificadas por los agentes.

2.1.2. Comunicación entre agentes

Uno de los temas de interés en las actividades de investigación y desarrollo de nuestra línea, es el uso de los actos del habla (speech acts [3]) en la comunicación entre agentes. Desde las primeras consideraciones de los actos del habla en el procesamiento del lenguaje natural [9], se observó la conveniencia de considerar a estos actos como acciones que forman parte de planes. ¿Por qué una persona elige comunicar una idea en lugar de otra? Para encontrar las razones por las que se genera una frase, es mucho más simple tomarla como una acción antes que intentar encontrar esas razones a partir del análisis de las palabras que componen una frase. La comunicación es una actividad netamente social que los agentes usan como parte de un plan para conseguir sus metas. Cada acto comunicativo tiene su razón de ser, por eso se puede llegar a comprender mejor un enunciado al considerarlo como un operador que modifica el estado actual del mundo.

Durante su accionar, los agentes deben escoger el acto del habla más adecuado para cada situación, lo que es una tarea compleja y puede llegar a requerir analizar distintas variables y contextos. En este sentido, para que los agentes puedan proceder de una manera aceptable, investigamos el uso de un formalismo de argumentación rebatible [23] para la elección de las próximas acciones o actos del habla a ser aplicados. Este formalismo tiene en cuenta las condiciones necesarias para aplicar un operador y los cambios producidos [1].

2.2. Razonando con información incompleta y contradictoria

2.2.1. Programación en Lógica Rebatible

La argumentación es un modelo de razonamiento basado en la construcción y evaluación de argumentos que interactúan entre sí. Estos argumentos tienen como fin, soportar, explicar o atacar enunciados que pueden ser creencias, opiniones, decisiones, etc. Existen numerosas propuestas de sistemas de argumentación abstractos y concretos [10], donde el sistema abstracto propuesto por Dung [15], ha sido un trabajo seminal sobre el cual se han desarrollado otros sistemas argumentativos. De los sistemas de argumentación concretos, nosotros utilizamos para la representación de conocimiento y razonamiento a DeLP (*Defeasible Logic Programming*) [23], un formalismo de argumentación rebatible basado en programación lógica. Este formalismo ha sido exitosamente integrado en aplicaciones del mundo real tales como búsqueda inteligente en la web [11, 13], clustering [24], procesamiento del lenguaje natural [12] y knowledge management [7], entre otras.

En particular, para realizar los trabajos de investigación y desarrollo, en la línea se usa DeLP-Server [22], una implementación del intérprete de DeLP que provee un servicio de razonamiento argumentativo para agentes deliberativos. Este servicio se ejecuta como una aplicación stand-alone que interactúa con los agentes deliberativos usando el protocolo TCP/IP. De esta manera, los agentes son vistos como clientes que envían consultas al servidor (DeLP-Server) el cual, luego de evaluarlas, retorna las respuestas a los clientes. Un cliente puede conectarse a múltiples DeLP-Servers y un DeLP-Server puede recibir consultas de múltiples clientes. Esto hace que DeLP-Server sea muy adecuado para la programación multi-agente.

Los trabajos de investigación realizados en esta línea, se llevan a cabo siguiendo un enfoque teórico-práctico. En algunos casos, la confrontación con el mundo real se hace utilizando un grupo de tres robots móviles Khepera 2 [25], que pueden manipular y transportar objetos y poseen además distintas capacidades de sensado. Como muchos de los aspectos relacionados con el comportamiento de los robots (agentes), requieren un lenguaje de representación expresivo que refleje fácilmente los procesos de decisión realizados por los robots, se ha desarrollado en Prolog, un frame-

work [16, 20] que les permite a los robots Khepera 2 razonar y representar conocimiento en DeLP, a través de la interacción con un DeLP-Server.

2.3. Aspectos del razonamiento práctico

2.3.1. Integrando agentes BDI y argumentación

Uno de los objetivos principales de esta línea, es el estudio y desarrollo de sistemas con agentes basados en el modelo BDI. Las arquitecturas (y modelos) BDI proponen a la trinidad **BDI (Beliefs, Desires e Intentions)** como los elementos claves del estado mental de un agente para tomar las decisiones acerca de cuándo y cómo actuar. Este tipo de enfoque ha demostrado una gran flexibilidad y efectividad en diversos problemas de gran complejidad del mundo real, lo que ha llevado a un creciente interés en la investigación de sus aspectos teóricos pero también de las plataformas que soportan el desarrollo de este tipo de agentes. En este sentido, un objetivo general de nuestro trabajo es el estudio de frameworks específicos para el desarrollo de agentes BDI como, por ejemplo, el framework de distribución gratuita Jadex.

Actualmente, nuestro trabajo principal en este área, está orientado a integrar planes escritos en Jadex con WS que proveen información relevante para distintos problemas. En particular, uno de los servicios más importantes que puede acceder un agente Jadex es el de razonamiento rebatible provisto por el DeLP-Server. De esta forma, se está logrando la integración en un mismo marco de trabajo, de capacidades de razonamiento teórico y razonamiento práctico.

2.3.2. Toma de decisión multicriterio basada en argumentación

La idea de articular decisiones en base a argumentos es relevante para varios enfoques diferentes de decisión, tales como decisión bajo incertidumbre, decisión multi-criterio y decisiones basadas en reglas. Teniendo en cuenta la experiencia de la línea en el desarrollo de un framework de decisión individual basado en argumentación rebatible [19] (sistema de decisión basado en reglas), actualmente se está redefiniendo este framework para que el tomador de decisiones pueda ponderar múltiples criterios de elección al momento de tomar una decisión.

DeLP es un lenguaje expresivo que nos permitió reflejar fácilmente los procesos de decisión de los agentes que tomaban decisiones usando

el framework propuesto en [19]. Sin embargo, no puede tratar con incertidumbre explícita dado que la rebatibilidad está codificada directamente en el lenguaje objeto. Es por eso, que al momento de extender el framework de [19] para que considerara múltiples criterios, se utilizó P-DeLP [2] en lugar de DeLP. P-DeLP extiende las capacidades de razonamiento de DeLP al permitir la incorporación de incertidumbre explícita, y esta característica primordial nos facilita la integración de un método existente de agregación multi-criterio [6] al framework de decisión.

Si bien el framework de decisión multi-criterio basado en argumentación aún no está completamente formalizado, la idea conceptual consiste en utilizar el mecanismo de inferencia de P-DeLP, generando todos los argumentos posibles a favor y en contra de cada acción (comparándolas de a pares) considerando en cada argumento un único criterio de preferencia del agente. Luego, se realiza una agregación de aquellos argumentos que soporten la misma conclusión (por ejemplo, que la acción a_1 es mejor que la a_2) y el procedimiento de garantía se aplica sobre los argumentos resultantes de esa agregación.

3. Resultados esperados/obtenidos

Los resultados obtenidos respecto a la toma de decisiones basadas en argumentación han sido presentados en las siguientes publicaciones científicas [19, 21, 18, 17]. Estos trabajos muestran la evolución de una propuesta original para resolver un problema de aplicación de toma de decisiones basadas en argumentación [17], hasta la presentación de un modelo teórico general para la toma de decisiones ([19]). Este último, combina reglas de decisión y argumentos, y se formaliza su comportamiento de elección comparándolo con el basado en los principios de la Teoría de Decisión Clásica. Los resultados esperados en esta línea de trabajo, se orientan a lograr una propuesta más acorde a la forma en que los seres humanos deliberan y finalmente toman o entienden una elección. Uno de nuestros trabajos futuros que tiende a lograr este objetivo, es la reformulación de este framework general para que permita la agregación dinámica de múltiples criterios de preferencia.

En el área de integración de agentes BDI, WS y razonamiento rebatible, se está desarrollando actualmente un sistema que permite acceder (vía WS) a información meteorológica actualizada, que

es transformada y combinada con información sobre las preferencias de un usuario especificadas en DeLP. Luego, un proceso de razonamiento rebatible es utilizado para determinar cual es la mejor acción a tomar en cada circunstancia.

En lo referente a la comunicación de agentes basada en speech acts y argumentación, nuestra idea es extender el trabajo previo realizado en [1] para considerar protocolos de interacción más generales.

4. Formación de recursos humanos

Trabajos de tesis vinculados con las temáticas descriptas previamente:

- 2 tesis doctorales en ejecución (co-dirección con investigadores del LIDIA (UNS))
- 1 tesis de maestría aprobada (co-dirección con investigadores del LIDIA (UNS))
- 1 tesis de maestría en ejecución (co-dirección con investigadores del LIDIA (UNS))
- 2 tesis de Licenciatura aprobadas.
- 5 tesis de Licenciatura en ejecución.

Referencias

- [1] G. Aguirre, M. Errecalde, and G. Simari. Actos del habla en la comunicación. una propuesta usando programación lógica rebatible. In *Proceedings del XIV Congreso Argentino de Ciencias de la Computación (CACIC 2008)*, pages 2009–2020, 2008.
- [2] Teresa Alsinet, Carlos I. Chesñevar, Lluís Godo, and Guillermo R. Simari. A logic programming framework for possibilistic argumentation: Formalization and logical properties. *Fuzzy Sets and Systems*, 159(10):1208–1228, 2008.
- [3] J. L. Austin. *How to Do Things with Words*. Harvard University Press, Cambridge, MA, Cambridge, MA, 1962.
- [4] F. Bellifemine, G. Caire, and D. Greenwood. *Developing Multi-Agent Systems with JADE (Wiley Series in Agent Technology)*. John Wiley & Sons, 2007.
- [5] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, 2nd edition, 1985.
- [6] R. Bisdorff. Preference aggregation with multiple criteria of ordinal significance. In D. Bouyssou, M. Janowitz, F. Roberts, and A. Tsouki, editors, *Annales du LAMSADE*, pages 25–44, October 2004.
- [7] Ramón F. Brena, José Luis Aguirre, Carlos Iván Chesñevar, Eduardo H. Ramírez, and Leonardo Garrido. Knowledge and information distribution leveraged by intelligent agents. *Knowl. Inf. Syst.*, 12(2):203–227, 2007.
- [8] Rodney A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation-Computational Intelligence*, 2(1):14–23, 1986.
- [9] B. Bruce and D. Newman. Interacting plans. *Cognitive Science*, pages 195–233, 1978.
- [10] C. Chesñevar, A. Maguitman, and R. P. Loui. Logical models of argument. *ACM Computing Surveys*, 32(4):337–383, 2000.
- [11] C. I. Chesñevar and A. G. Maguitman. Arguenet: an argument-based recommender system for solving web search queries. In *Intelligent Systems, Proc. of 2nd International IEEE Conference*, volume 1, pages 282–287, June 2004.
- [12] Carlos I. Chesñevar and Ana G. Maguitman. An argumentative approach to assessing natural language usage based on the web corpus. In *Proc. of the 16th ECAI Conference. Valencia, Spain*, pages 581–585. IOS Press, August 2004.
- [13] Carlos Iván Chesñevar, Ana Gabriela Maguitman, and Guillermo Ricardo Simari. A first approach to argument-based recommender systems based on defeasible logic programming. In *NMR*, pages 109–117, 2004.
- [14] J. Doyle and T. Thomason. Background to qualitative decision theory. *AI Magazine*, 1999.
- [15] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
- [16] E. Ferretti, M. Errecalde, A. García, and G. Simari. Khedelp: A framework to support defeasible logic programming for the khepera robots. In *ISRA*, pages 98–103, San Miguel Regla, Hidalgo, México, August 2006.
- [17] E. Ferretti, M. Errecalde, A. García, and G. Simari. An application of defeasible logic programming to decision making in a robotic environment. In *LPNMR*, volume 4483 of *LNAI*, pages 297–302, AZ, USA, May 2007. Springer.
- [18] Edgardo Ferretti, Marcelo Errecalde, Alejandro García, and Guillermo Simari. Defeasible decision making in a robotic environment. In *Anales del XIII CACIC*, pages 1335–1446, Octubre 2-5 2007.
- [19] Edgardo Ferretti, Marcelo Errecalde, Alejandro García, and Guillermo Simari. Decision rules and arguments in defeasible decision making. In *COMMA*, volume 172, pages 171–182, Toulouse, France, May 2008. IOS Press.
- [20] Edgardo Ferretti, Marcelo Luis Errecalde, Alejandro Javier García, and Guillermo Ricardo Simari. Khepera robots with argumentative reasoning. In *AMIRE Symposium*, pages 199–206, Bs.As., Argentina, October 2-5 2007.
- [21] Edgardo Ferretti, Nicolás Rotstein, Marcelo Errecalde, Alejandro García, and Guillermo Simari. Defeasible decision making in a multi-robot environment. *Research in Computing Science*, 32:150–160, September 2007. Special Issue: Advances in Artificial Intelligence and Applications.
- [22] Alejandro Javier García, Nicolás D. Rotstein, Mariano Tucat, and Guillermo Ricardo Simari. An argumentative reasoning service for deliberative agents. In *KSEM*, volume 4798 of *LNCS*, pages 128–139. Springer, 2007.
- [23] Alejandro Javier García and Guillermo Ricardo Simari. Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming*, 4(2):95–138, 2004.
- [24] Sergio A. Gómez and Carlos I. Chesñevar. A hybrid approach to pattern classification using neural networks and defeasible argumentation. In *Proc. of the Intl. FLAIRS Conference. Miami, USA*, pages 393–398, May 2004.
- [25] K-Team. Khepera 2. <http://www.k-team.com>. A miniature mobile robot designed as a research and teaching tool.
- [26] Michael Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley & Sons, Chichester, England, 2002.

VALIDACION DE FIRMAS OLOGRAFAS

Tamara D. Blum, Ariel J. A. Sztern, Daniela López De Luise
AI-Group
Facultad de Ingeniería de la Universidad de Palermo
Ciudad de Buenos Aires, Argentina. aigroup@palermo.edu

CONTEXTO

Dentro del ámbito de investigación que se da en el laboratorio de Inteligencia Artificial de la Universidad de Palermo, se viene trabajando sobre un proyecto para la validación de firmas ológrafas o manuscritas.

Este trabajo presenta un prototipo funcional en desarrollo cuyo propósito es facilitar la validación de firmas ológrafas. El sistema realiza la verificación contra una base de datos enriquecida con el agregado de vectores numéricos que sintetizan la imagen de la firma capturada según cierta red neuronal. Además se presenta el diseño del esquema de seguridad y se justifican las decisiones arquitectónicas desde un punto de vista teórico.

RESUMEN

Existen, al día de hoy, una cantidad de alternativas en lo que hace a la validación de individuos. La mayoría de éstas presentan una serie de inconvenientes que las hace difícilmente masivas y/o suficientemente seguras. El presente trabajo profundiza acerca de los conceptos involucrados en dichas alternativas, y presenta la implementación práctica de una Solución Informática para validación de firmas manuscritas, denominada proyecto PI-1. El prototipo obtenido responde a una arquitectura original e introduce variantes que han surgido como producto de la investigación realizada en la combinación y optimización de ciertas técnicas. Se han concentrado los esfuerzos en obtener un procesamiento de tipo off-line de las firmas (denominado estático) a partir de imágenes capturadas desde un dispositivo celular móvil.

Palabras clave: Validación de firmas, Firmas manuscritas, Firmas ológrafas, Biometría, Criptografía

1. INTRODUCCION

El prototipo PI-1 permite calcular el grado de cercanía que tiene una firma capturada respecto de otras previamente registradas para un determinado firmante, a la vez que valida otros datos identificatorios. Ello ayudará a validar la identidad de un individuo, preservando el sentido legal que tiene la firma manuscrita.

Algunas de las razones que motivan esta investigación son:

- Los dispositivos biométricos que existen actualmente deben ser complementados para cumplir con los requisitos esperados en cuanto a la seguridad o son métodos invasivos y de difícil aplicación.
- Aún existiendo la Ley de Firma Digital [1][2][3][4][5][6][7], la firma manuscrita es un requisito ineludible en gran cantidad de trámites, como transacciones inmobiliarias, cesiones de poder, autorizaciones o documentos jurídicos [8][12].
- Es un método en el que se puede obtener un bajo índice de error.
- No se requiere que el firmante sepa leer y escribir, dado que las firmas son simplemente garabatos.
- Es una forma natural de identificarse, ya que es costumbre complementar ciertos trámites con la rúbrica personal en la documentación.
- No se ve afectada por implicancias externas como puede ser el tipo de trabajo que realice una persona, como si sucede con la huella dactilar [9], donde puede deteriorarse el dibujo de la misma.
- No requiere de costoso equipamiento.

Entre las características establecidas para el

prototipo se han definido las siguientes:

- Debe ser portable.
- Capacidad de ser usado en forma masiva.
- Debe ser accesible vía Internet.
- No debe tener costos adicionales por licenciamientos de terceras partes, como motores de bases de datos, sistemas operativos, etc.
- Debe poder ser ejecutable desde dispositivos móviles, como también en servidores o computadoras personales.
- No puede depender de la existencia de hardware de alto costo como una limitación funcional.
- Debe proveer un nivel de seguridad aceptable para el entorno de uso, ya que los datos que se manejan pueden llegar a ser sensibles.

El prototipo consta de una base de datos que debe ser alimentada con la información de los firmantes y una muestra de firmas (se recomienda no menos de 5 firmas por firmante para garantizar la calidad de los resultados). Dichas muestras son las que permiten extraer las características de las firmas de cada firmante por una red MultiLayer Perceptron (MLP), que se usará luego para poder validar las futuras muestras y devolver un grado de coincidencia respecto de las muestras existentes.

Si bien la propuesta originalmente contempla la captura de firmas utilizando como fuente de información diferentes alternativas y dispositivos, la implementación realizada se centró en el uso de dispositivos celulares móviles. Sin embargo debe destacarse que la forma en que se provee la entrada influye en la calidad de resultados que se obtienen.

2. PROTOTIPO PI-1

A continuación, se describen brevemente los puntos más importantes que se han tomado en cuenta para el prototipo.

Arquitectura general

Como se ha mencionado anteriormente, la solución que se plantea permitiría que se incorporen y validen firmas a partir de

diferentes formatos y dispositivos: dispositivo de lectura on-line, archivo de firmas (escaneado o fotocopiado), o Webcams. El desarrollo admite ser invocado a través de Internet como un servicio Web.

Las posibilidades que brindan los dispositivos de lectura on-line son mayores, motivo por el cual se puede ofrecer un mejor resultado mediante este tipo de procesamiento. No obstante, mediante la lectura off-line también se puede llegar a obtener resultados con un bajo índice de error.

El esquema de funcionamiento se muestra en la figura 1.

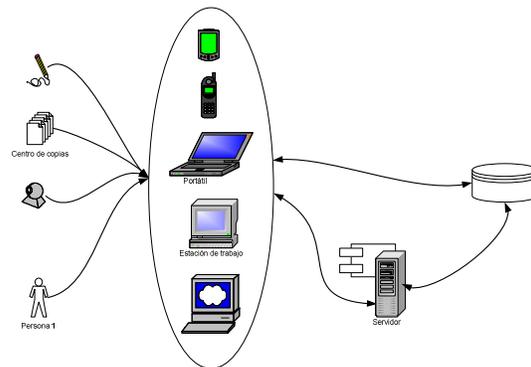


Fig. 1. Arquitectura general del sistema de validación.

Previo a todo proceso de validación es imprescindible que se cargue una serie de muestras de cada individuo contra las cuales validar. A fin de contemplar los cambios naturales que se van dando en las firmas de los individuos es recomendable que las mismas sean actualizadas a lo largo de los años.

PI-1 analiza la firma y la registra en la base de datos de firmas; o bien sólo la valida dependiendo de los requerimientos eventuales.

Base de Datos

La base de datos se almacena en formato XML (Extensible Markup Language). Esto permite tener un mejor control a nivel seguridad, y la hace totalmente portable ya que no requiere motores específicos para el acceso.

Ya se ha mencionado anteriormente que dentro de los objetivos de la propuesta se encuentra el obtener un producto capaz de funcionar en diferentes dispositivos y/o plataformas. Por lo tanto, no se puede dejar de tener en cuenta que

motores de bases de datos existentes podrían llegar a ser un inconveniente en este aspecto. La situación se agrava aún más si requiere de costos adicionales en concepto de licencia. Es por ese motivo que se ha decidido trabajar con una base de datos XML.

Adicionalmente, este tipo de bases de datos suponen un acceso a la información más performante que las bases de datos relacionales, dado que guardan los documentos enteros físicamente juntos o utilizan punteros físicos en lugar de lógicos entre las partes de los documentos [10].

Las bases de datos XML son del tipo árbol, donde los nodos son los atributos y cada una de las subdivisiones son las raíces jerárquicas.

Existen actualmente diferentes alternativas para acceder a la información desde distintos lenguajes de programación. Y no necesariamente hay una forma que sea la más conveniente. Los requisitos para validar no son los mismos que para la registración.

Se optó por desarrollar una interfaz de acceso a los datos e implementarla utilizando SAX. Esto permite intercambiar la forma de acceder a los datos en un futuro, si se encontrara una mejor alternativa, con sólo implementar en una nueva clase la misma interfaz.

Seguridad

Los datos se guardan encriptados para dificultar la interpretación de los mismos e impedir el “robo” de información.

Sin embargo, teniendo en cuenta que pueden producirse ataques de distintos tipos, dentro del presente esquema, se trabaja la seguridad en todos los niveles de la arquitectura. Esta se brinda a partir de:

- Encriptación GSM (Global System for Mobile Communications)
- Páginas Web Seguras
- Encriptación en Servicios Web
- Seguridad en Threads

El procesamiento de los datos contempla altos niveles de seguridad, ya que se utilizan los propios de las conexiones Web o móviles y también métodos adicionales para dar mayor seguridad al proceso y prevenir robo de información en el momento del procesamiento

de los datos.

Interfaces gráficas

Las interfaces gráficas son simples e intuitivas, tratando que los datos sean visibles en su conjunto dentro de la misma ventana gráfica. Ello agiliza y facilita el uso de la solución.

A continuación se grafican las interfaces principales del sistema.

En las figuras 2.1, 2.2 y 2.3 se muestran las pantallas de registración de un nuevo firmante, con los datos básicos necesarios. El primer caso grafica como sería desde un teléfono celular y el segundo desde una computadora personal.

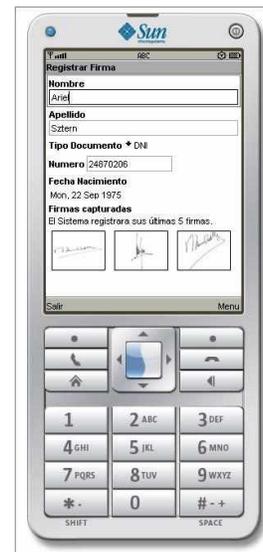


Fig. 2.1. Formulario para captura de datos del firmante y muestras de firmas desde teléfono celular.



Fig. 2.2. Formulario para registrar un nuevo firmante desde una computadora personal.

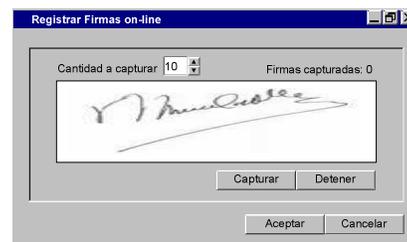


Fig. 2.3. Formulario para registrar firmas utilizando un dispositivo de captura (on-line).

Procesamiento de imágenes

El análisis de las características y patrones comprendidos en las firmas ológrafas, propone un desafío complejo, dado que el mismo, varía dependiendo de la calidad de la muestra obtenida.

Teniendo en cuenta que los dispositivos utilizados para la captura de firmas, son variados y con características diferentes (celulares, PDAs, archivos, etc.), es necesario realizar un proceso que permita normalizar y eliminar aspectos que dificulten el procesamiento de las imágenes en la red neuronal.

Se han analizado distintas alternativas que tienden a mejorar la calidad de los registros de firmas obtenidos, enfocándose principalmente en los siguientes aspectos:

- **Ruido:** el ruido comprende cualquier alteración producida antes o durante la captura, que incorpore trazos, marcas, manchas o sombras, externas a la firma. Esta deformación de la firma, puede generar problemas en la identificación de sus características.
- **Centrado:** teniendo en cuenta que el método para captura de las firmas no se encuentra estandarizado, es altamente probable que las imágenes obtenidas no se encuentren centradas. Esto distorsiona el reconocimiento de los patrones, entre firmas de una misma persona.
- **Tamaño:** las muestras obtenidas, al no ser uniformes, pueden responder a distintas escalas. Esto dificulta no sólo la comprensión de los patrones de la firma, sino también ocasiona un mayor tiempo de procesamiento por parte de la red neuronal, ya que las imágenes contienen mayor información. Para ello, es fundamental reducir el tamaño de la muestra, generando condiciones similares para el procesamiento de todas las firmas.
- **Color:** se ha considerado para el presente trabajo, la estandarización de los colores correspondientes a la firma, con el fin de

normalizarla y reducir su volumen. Por ello, se determinó utilizar imágenes en blanco y negro, lo que permite simplificar la obtención del vector de características y agilizar el procesamiento.

Identificación de patrones

El punto más crítico de la investigación, se ha focalizado en la identificación de los patrones característicos de una firma ológrafa. Estos patrones son los que identifican aquellas particularidades que vuelven única a la firma. Se han estudiado distintas aproximaciones utilizadas en campos similares:

- **Reconocimiento de rostros:** en este punto se ha analizado la propuesta introducida por Sirovich y M. Kirby, conocida como Eigenfaces [11]. A diferencia de un rostro, la identificación de firmas, es una tarea que requiere mayor exactitud, ya que al validar una firma, se debe desestimar aquellas que correspondan a falsificaciones. En el caso de los rostros, si dos personas poseen facciones muy similares, es probable se deseen obtener ambos resultados como válidos.
- **Redes neuronales de Clasificación:** estas redes responden a patrones más específicos, pudiendo clasificar una determinada entrada, en un conjunto de salidas ya conocidas. En este caso, el resultado obtenido no es un vector de características, sino una clase específica, previamente establecida durante el entrenamiento de la red. La complejidad presentada por las firmas, hace que la identificación concreta de una salida, puede desembocar en falso positivo, al catalogar como válida una falsificación, por poseer mayor afinidad con la firma real que con el resto del universo de firmas.
- **Redes neuronales de Retropropagación:** hemos considerado este tipo de redes, como la aproximación más concreta sobre la problemática planteada, dado que es capaz de ajustar sus pesos en forma automática a partir de los valores obtenidos. Esto contribuye a mejorar la

comprensión de los rasgos que componen la firma, obteniéndose como resultado un vector que representa los patrones significativos de la firma procesada.

3. RESULTADOS Y TRABAJO A FUTURO

Al momento se han desarrollado los prototipos para captura de imágenes y el almacenamiento de las firmas y demás información acerca del firmante en la Base de Datos.

Se ha implementado la seguridad en los procesos de comunicación y resguardo. Se está trabajando en las alternativas de procesamiento con redes neuronales.

Queda como trabajo futuro finalizar el estudio de las redes neuronales y las alternativas de implementación. También se realizará un estudio de eficiencia y confiabilidad.

Asimismo, se realizará el análisis estadístico de confiabilidad en validaciones on-line (dinámico) de firmas manuscritas.

Respecto al diseño de la base de datos, queda pendiente terminar de definir si la mejor alternativa es que los datos se guarden en su conjunto dentro del mismo archivo o guardarlos estilo “cabecera-detalle”.

Las ventajas de guardar el conjunto de datos en diferentes archivos es que ante el hecho de corromperse alguno de ellos, se puede rearmar la base de datos con sólo tomar nuevamente las muestras que corresponden al firmante que tiene sus firmas corruptas. Sin embargo, el resguardo de la información puede verse afectado si el archivo que se corrompe es el que actúa de “cabecera”. El mismo grado de dificultad existe ante la corrupción del archivo que contendría a la base de datos completa,

Aunque el almacenamiento de la información utilizando varios archivos es ventajoso por su disposición, este sistema podría traer como consecuencia alguna desventaja en la seguridad, ya que permitiría un rastreo más simple del contenido y las muestras de las firmas por parte de usuarios locales del sistema.

4. FORMACION DE RECURSOS HUMANOS

En el marco de la red de colaboración PI, dedicada a la implementación de sistemas inteligentes, suman sus aportes profesionales de las siguientes instituciones:

- Por la Universidad de Palermo intervienen 1 doctora y 2 licenciados en informática.
- Por la UADER intervienen 1 doctor y 2 doctorandos en roles de investigación.
- Por la UNLP interviene 1 doctor con rol de investigador.
- Por la UTN Facultad Regional Concepción del Uruguay intervienen 3 graduados como investigadores.

5. BIBLIOGRAFIA

- [1] Información consultada en <http://www.saij.jus.gov.ar>, Sistema Argentino de Informática Jurídica - Ministerio de Justicia, Seguridad y Derechos Humanos
- [2] Ley 25.506, Ley de Firma Digital, sancionada el 14 de Noviembre de 2001. <http://www.saij.jus.gov.ar>
- [3] La Ley Provincial N° 12.491, del 24 de Noviembre de 2005 dispone de un plazo de hasta 5 (cinco) años para la implementación de la tecnología necesaria para la Firma Digital en el ámbito normativo.
- [4] Ley Provincial N° 633, del 6 de Julio de 2004.
- [5] Ley Provincial N° 2.073, del 9 de Octubre de 2003.
- [6] Ley Provincial N° 5.366, del 5 de Julio de 2005.
- [7] Ley Provincial N° 13.666, del 12 de Abril de 2007.
- [8] Trevathan, J. and McCabe, A. Remote Handwritten Signature Authentication. In Proceedings of the 2nd International Conference on eBusiness and Telecommunication Networks (ICETE '05), pp 335--339, 2005. <http://citeseer.ist.psu.edu/trevathan05remote.html>.
- [9] Gupta, J and McCabe, A. A Review of Dynamic Handwritten Signature Verification. James Cook University, Australia (1997). <http://citeseer.ist.psu.edu/gupta97review.html>
- [10] Ronald Bourret, “XML and Databases”, Septiembre 2005, <http://www.rpbouret.com/xml/XMLAndDatabases.htm>.
- [11] L. Sirovich and M. Kirby, “Low-dimensional procedure for the characterization of human faces”, Journal of Optical Society of America. 1987
- [12] G. Rodríguez, “De la firma autógrafa a la firma digital”, Facultad de ciencias jurídicas y políticas, Universidad de Zulia [Disertación sobre firma ológrafa y digital; y la implementación de ésta última como medio de autenticación y confirmación para documentos electrónicos].

Manejo de Restricciones y Mejora en los Procesos de Búsqueda en Algoritmos Evolutivos aplicados a Problemas Dinámicos y Scheduling

Villagra Andrea, Lasso Marta, Pandolfi Daniel, de San Pedro Eugenia, Villagra Silvia, Rasjido José, Seron Natalia, Orozco Sergio

LabTEm - Unidad Académica Caleta Olivia
Universidad Nacional de la Patagonia Austral

{avillagra, mlasso, dpandolfi, edesanpedro, svillagra, jrasjido, nseron, sorozco}@uaco.unpa.edu.ar

CONTEXTO

Esta línea de Investigación forma parte del proyecto “Manejo de Restricciones con algoritmos evolutivos aplicados a la Planificación de Locaciones Petroleras” del Laboratorio de Tecnologías Emergentes (LabTEm) en el marco del programa de Investigación en Ciencia y Tecnología de la Universidad Nacional de la Patagonia Austral.

RESUMEN

En este trabajo se presenta un proyecto de investigación, como continuación del proyecto de Investigación sobre Metaheurísticas desarrollado en años anteriores. Durante este año se pretende focalizar en las siguientes dos líneas. La primera relacionada con la investigación experimental de tres aspectos fundamentales de la herramienta: Manejo de restricciones, Scheduling dinámico y Mejora de los procesos de búsqueda. La segunda línea constituye el desarrollo de un prototipo operacional denominado PAE. PAE, es una herramienta para la planificación del mantenimiento de locaciones petroleras sujeto a restricciones. El motor de búsqueda de esta herramienta está basada en un algoritmo evolutivo multirecombinativo conocido como SRI-MCMP (Stud and Random Immigrants - Multiple Crossover on Multiple Parents) y fue desarrollado por investigadores del LabTEm.

Palabras clave: Planificación, Mantenimiento, Algoritmo Evolutivo, Manejo de Restricciones.

1. INTRODUCCION

Las actividades petroleras de explotación y transporte del petróleo son muy importantes para el desarrollo económico de la sociedad industrial moderna. Sin embargo, estas actividades son generadoras de riesgos que se traducen en contaminaciones accidentales que afectan directamente al ecosistema. El impacto ambiental ocasionado por la industria petrolera comprende los efectos de todas y cada una de las fases involucradas en las etapas tanto de exploración, explotación, transformación, distribución y comercialización. En muchas de las regiones donde se ha explotado el

petróleo, se han generado contingencias de diferentes magnitudes como consecuencia de las operaciones desarrolladas en: tanques de almacenamiento, oleoductos, equipos de perforación, buques petroleros, etc. Además de los impactos generados en el medio como consecuencia de la exploración y explotación del petróleo, existen también daños ocasionados a las personas que lo emplean o se encuentran en contacto con él dentro de su ámbito laboral.

En la última década, las Empresas dedicadas a la explotación, producción y transporte del recurso, han visto la necesidad de implementar medidas de prevención a fin de evitar y / o minimizar los daños ocasionados al medio ambiente, personas y bienes materiales.

La mejor manera de atacar el problema de contaminación por petróleo es prevenir el incidente. Normalmente éstos se producen por fallas de equipos o del material y fallas humanas. Los primeros pueden ser subsanados mediante inspecciones periódicas y un mantenimiento adecuado; y los segundos, mediante la instrucción y el entrenamiento del personal en forma permanente. La minimización del tiempo entre cada período de inspección de las locaciones se obtiene con una mejora en la planificación del mantenimiento y esto puede ser abordado como un problema de scheduling. Se ha demostrado, que muchos problemas de scheduling pertenecen a la clase NP-hard [1] reflejando así la relevancia industrial de este tipo de problemas.

Los Algoritmos Evolutivos (AEs) han sido aplicados exitosamente en la resolución de diversos tipos de problemas de planificación tales como scheduling o routing [2], [3]. Los AEs son metaheurísticas que comparten un concepto base común que es simular la evolución de los individuos que forman la población usando un conjunto de operadores predefinidos de selección y de búsqueda. Existe una gran variedad de AEs, dentro de ellos se incluyen los Algoritmos Genéticos [4] que son los utilizados en este problema. Tendencias actuales en AEs hacen uso de enfoques con multirecombinación [5], [6], [7] y enfoques con múltiples padres conocidos como MCMP (Multiple Crossover Multiple Parent) [8], [9] [10].

La mayoría de los problemas de optimización incluyen restricciones de cierta clase, que constituyen grandes desafíos a la resolución de problemas de optimización. Las restricciones son límites impuestos a las variables de decisión y en general las restricciones son una parte integral de la formulación de cualquier problema [11].

En el dominio específico de problemas discretos, tales como Knapsack Problem, Set Covering Problem, Vehicle Routing Problem, y todos los tipos de Scheduling Problems, todos están sujetos a restricciones. El espacio de búsqueda de estos problemas se subdivide en dos subconjuntos, el espacio de soluciones factibles y el espacio de soluciones no factibles. Estos subespacios no necesariamente son convexos y no necesariamente están conectados. Para la resolución del problema de optimización sujeto a restricciones los algoritmos de búsqueda, tal como lo son AEs, promueven soluciones que navegan en ambos subespacios. El problema consistirá entonces en cómo tratar soluciones no factibles para explorar el subespacio de soluciones factibles [12].

Es importante destacar que el grupo viene desarrollando tareas de investigación en el problema de planificación del mantenimiento de locaciones petroleras en aspectos de planificación estática, dinámica y manejo de restricciones [13], [14], [15].

En la siguiente sección se presentará las líneas de investigación y desarrollo vinculadas con este proyecto.

2. LINEAS DE INVESTIGACION y DESARROLLO

- Manejo de Restricciones

El manejo de restricciones en algoritmos evolutivos puede abordarse desde distintas técnicas tales como: Funciones de penalidad, Decodificadores, Algoritmos de reparación, y operadores de preservación de restricciones.

Las funciones de penalización han sido ampliamente discutidas por la literatura en las últimas décadas [16]. Existen enfoques que penalizan soluciones no factibles como nuevos y promisorios enfoques que también penalizan las soluciones factibles. Las técnicas de penalización pueden subdividirse como estáticas, dinámicas, adaptativas y mortales.

Las técnicas basadas en algoritmos de reparación mapean una solución desde el espacio no factible al espacio factible. Los procesos de reparación que modifican genéticamente la solución no factible son conocidos como Lamarkianos y transforman una solución no factible en factible la que es evolucionada por el algoritmo. Un enfoque menos destructivo de la solución no factible, permite combinar el aprendizaje y la evolución, estrategia conocida como Baldwiniana. En este último enfoque las soluciones son reparadas solo para su evaluación. Estudios analíticos y empíricos indican que esta técnica reduce la velocidad de convergencia del

algoritmo evolutivo y permite converger a óptimos globales [17].

- Problemas dinámicos y scheduling

Las siguientes son algunas de las situaciones a considerar cuando se produce un cambio en el entorno. Ellas son tomadas principalmente de [18] y [19].

- Reusar información siempre que se pueda para evitar partir desde el inicio en el nuevo contexto del problema.

- Descubrimiento de los cambios.

- La solución al problema modificado puede ser similar al anterior.

- Cambiar una solución por cambios pequeños en el problema puede ser costoso, lo cual podría implicar el mantenimiento de la solución tal como está.

Asimismo, se plantean algunos principios básicos que un algoritmo dinámico debería cumplir: capacidad de adaptación a los cambios; capacidad de discernir, ante un cambio, entre el costo de mantener la solución como está y aquel asociado al proceso de cambio; apuntar a encontrar soluciones robustas de manera que sigan manteniendo su calidad cuando se produzcan cambios muy pequeños. Las situaciones y principios anteriores son considerados por [18] en el contexto de AEs. Sin embargo, son igualmente aplicables a otro tipo de metaheurísticas usadas en el contexto de problemas dinámicos.

Lo que hay que tener en cuenta es el tipo o grado de cambio producido en el entorno. Si el cambio es muy severo, es decir, que el problema después del cambio es completamente distinto, puede no tener sentido la aplicación de ciertos métodos dado que el conocimiento adquirido difícilmente se pueda aplicar en el nuevo contexto. En síntesis, el objetivo debería estar centrado en aquellos problemas dinámicos que exhiben un dinamismo razonable de manera tal que se pueda explotar la experiencia adquirida por el método usado.

- Mejora en el proceso de búsqueda

Las metaheurísticas (MHS) [20] son métodos que integran de diversas maneras, procedimientos de mejora local y estrategias de alto nivel para crear un proceso capaz de escapar de óptimos locales y realizar una búsqueda robusta en el espacio de búsqueda. En su evolución, estos métodos han incorporado diferentes estrategias para evitar la convergencia a óptimos locales, especialmente en espacios de búsqueda complejos. Aunque las metaheurísticas en general presentan un bajo nivel de desarrollo desde el punto de vista matemático, tales procesos son válidos en muchas situaciones prácticas para las cuales se han obtenido resultados muy interesantes para problemas del mundo real, en particular en el ámbito de la industria.

Resta mencionar como punto adicional que la garantía de un adecuado desempeño de las MHS en general depende en gran medida del desempeño de los operadores o mecanismos de exploración propios

de cada una de ellas. En este sentido, la utilización de enfoques avanzados para mejorar el proceso de exploración es un tópico de interés dado su inmediata aplicación y beneficio directo respecto a posibles mejoras en el diseño de MHs. Por ejemplo, el concepto de sociobiología [21] es mecanismo evolutivo exitoso que podría ser usado como mecanismo alternativo en los modelos algorítmicos o metaheurísticas que imitan el proceso evolutivo. En este sentido, la perspectiva Darwinista tradicional explica la regulación de la descendencia a través de las restricciones del medio ambiente: alimentos, clima y depredación, las que en su conjunto regulan los tamaños y adaptabilidad de las poblaciones. Por el contrario, según [22], los animales pueden autorregular su propia población estimando restricciones del ambiente y por lo tanto variar sus tasas de reproducción. Por otro lado, la teoría de [23] sobre la evolución de genética del comportamiento social, propone que el valor adaptativo de un individuo se podría medir no solamente por su éxito personal en materia de reproducción, sino a través del beneficio de la reproducción de otros individuos (parientes). La suma de estos dos conceptos fue denominado como “inclusive fitness” o adecuación adaptativa global. Basado en estos conceptos previos y en el constante desarrollo e investigación de conceptos útiles para mejorar el desempeño de los enfoques metaheurísticos, se plantea como una alternativa, el desarrollo de algoritmos y/o mejoras en los enfoques metaheurísticos a través de la incorporación de elementos vinculados al comportamiento social de determinadas especies del mundo animal.

3. RESULTADOS OBTENIDOS/ESPERADOS

A continuación se mencionan los resultados esperados del proyecto:

- Investigación en tendencias actuales en scheduling dinámico.
- Estudio y desarrollo de métodos para el manejo de restricciones
- Investigación en tendencias actuales respecto a la eficiencia de las distintas Metaheurísticas en la exploración del espacio de búsqueda.
- Determinación de nuevos conceptos a ser incluidos en la fase explorativa.
- Diseño e implementación de las distintas Metaheurísticas consideradas aplicando los conceptos alternativos o mejorados.
- Desarrollo del Prototipo operable para la Planificación Dinámica Sujeta a restricciones.

4. FORMACION DE RECURSOS HUMANOS

Un integrante de esta línea de investigación está desarrollando su Tesis de Doctorado en Mejoras en el proceso de Búsqueda.

Un integrante esta desarrollando su Tesis de Maestría en Scheduling Dinámico.

Hay dos Becarios alumnos y dos pasantes en el LabTEm trabajando en el proyecto y se dirigen Trabajos Finales de Ingeniería en Sistemas en temas relacionados con este proyecto.

5. BIBLIOGRAFIA

1. Brucker P., Scheduling Algorithms, 3rd ed. Springer-Verlag New York, (2004).
2. Chang P., Hsieh J., and Wang Y. Knowledge Incorporation in Evolutionary Computation, chapter F. Divina and E. Marchiori. Springer-Verlag, 2005.
3. Jaskowski P. y Sobotka A. Multicriteria construction project scheduling method using evolutionary algorithm. Operational Research an International Journal, 6(3). (2006).
4. Goldberg, D.E.; Genetic Algorithms in Search, Optimization and Machine Learning; Addison-Wesley, Reading, MA, 1989.
5. Eiben A.E. y Bäck Th. “An Empirical Investigation of Multi-Parent Recombination Operators in Evolution Strategies”, Evolutionary Computation, 5(3):347-365, (1997).
6. Eiben A.E., Raué P.E. y Ruttkay Z., “Genetic algorithms with multi-parent recombination”. Proc. 3rd Conference on Parallel Problem Solving from Nature, Springer-Verlag, number 866 in LNCS, pp. 78-87, (1994).
7. Eiben A.E., Van Kemenade C.H.M., y Kok J.N. “Orgy in the Computer: Multi-Parent Reproduction in Genetic Algorithms”. Proc. 3rd European Conference on Artificial Life, Springer-Verlag, number 929 in LNAI, pages 934-945, (1995).
8. Esquivel S., Leiva A., Gallard R. “Multiple Crossover per Couple in Genetic Algorithms”. Proc. Fourth IEEE Conference on Evolutionary Computation (ICEC'97), pp 103-106, Indianapolis, USA April (1997).
9. Esquivel S., Leiva A., Gallard R. “Couple Fitness Based Selection with Multiple Crossover per Couple in Genetic Algorithms”. Proc. International Symposium on Engineering of Intelligent Systems (EIS'98), pp 235-241, La Laguna, Tenerife, Spain, February (1998).
10. Esquivel S., Leiva H., Gallard R. “Multiple Crossovers between Multiple Parents to

- Improve Search in Evolutionary Algorithms”. Proc. Congreso on Evolutionary Computation (IEEE), pp 1589-1594, Washington DC, (1999).
11. Dhar, V., and Ranganathan, N. Integer Programming versus Expert Systems: An Experimental Comparison. Communications of the ACM 33:323–336. (1990).
 12. Michalewicz Z.; Genetic Algorithms + Data Structures = Evolutions Programs; Springer-Verlag, third, revised edition, 1997.
 13. Villagra A., de San Pedro E., Lasso M., Pandolfi D.; “Algoritmo multirecombinativo para la planificación dinámica del mantenimiento de locaciones petroleras”; Revista Internacional “INFORMACIÓN TECNOLÓGICA”; La Serena, Chile; Agosto 2008. Vol. 19 Numero 4 pp. 63-70.
 14. Villagra A., Montenegro C., de San Pedro M., Lasso M., y Pandolfi D. Planificación con restricciones del mantenimiento de locaciones petroleras. En *XII RPIC - Reunión de Trabajo en Procesamiento de la Información y Control*, 2007.
 15. Villagra A., Montenegro C., de San Pedro M., Lasso M., Vidal P., Pandolfi D.; “Mantenimiento de locaciones petroleras mediante un Algoritmo Multirecombinativo”; 8° Congreso Interamericano de Computación Aplicada a la Industria de Procesos (CAIP), Centro de Información Tecnológica (CIT). Asunción, Paraguay. Julio 2007. Cap.11 pp 319-322.
 16. Smith A. y Coit D. Constraint Handling Techniques—Penalty Functions. In Thomas Back, David B. Fogel, and Zbigniew Michalewicz, editors, Handbook of Evolutionary Computation, chapter C 5.2. Oxford University Press and Institute of Physics Publishing. (1997).
 17. Whitley, D., Gordon, S., and Mathias, K.: Lamarckian evolution, the Baldwin effect and function optimization, Parallel Problem Solving from Nature - PPSN III, pp. 6EŽ15“ú (1994).
 18. Branke J., Evolutionary Optimization in Dynamic Environments (Genetic Algorithms and Evolutionary Computation), Springer, 2001.
 19. Morrison R. W., Designing Evolutionary Algorithms for Dynamic Problems, Springer, 2004.
 20. Glover F., Kochenberger G.H. (editors)- Handbook of Metaheuristics, Kluwer Academic Publishers, 2003.
 21. Wilson E. O., Sociobiology: a new synthesis; Harvard University Press, 1975.
 22. Wynne-Edwards, V. C.; Animal dispersion in relation to social behavior; Oliver y Boyd, Edimburgo, UK, 1962.
 23. Hammilton W. D., The genetical evolution of social behaviour. Journal of Theoretical Biology, num 7, 1964, pp. 1-52.

Programación Orientada a Agentes en el Marco de Lenguajes Multiparadigmas

Maximiliano Klemen Claudio Vaucheret

Departamento de Ciencias de la Computación
Facultad de Economía y Administración
UNIVERSIDAD NACIONAL DEL COMAHUE
{maximilianoklemen,vaucheret}@gmail.com

Resumen

La presente línea de investigación surge de dos observaciones. Por un lado, observamos que el paradigma orientado a agentes aún hoy no se encuentra consensuadamente definido dentro de las diferentes áreas de las ciencias de la computación. Por el otro lado, la combinación de conceptos y herramientas de los diferentes paradigmas de programación en lenguajes actuales demostraron ser la vía más adecuada para la elaboración de herramientas de desarrollo de software que permitan representar adecuadamente los problemas y soluciones de dominios de la vida real. A partir de estas observaciones comenzamos una investigación cuyo objetivo es establecer las bases para dar origen a una definición definitiva del paradigma orientado a agentes, mostrando que constituye la convergencia entre los demás paradigmas. Con este fin nos proponemos desarrollar un lenguaje de programación que incorpore sus ideas elementales de manera armónica. Para su desarrollo escogimos un lenguaje multiparadigma basado en la programación lógica, por sus numerosas extensiones y la facilidad con la que permite definir nuevas estructuras sintácticas, semánticas y de control. Como conclusión de este trabajo utilizaremos el nuevo lenguaje para desarrollar aplicaciones prácticas en diversos campos como la robótica, la domótica, el desarrollo de aplicaciones móviles, etc.

1. Contexto

Presentamos aquí una línea de investigación contenida dentro del proyecto de investigación *Técnicas Avanzadas y Análisis para el Desarrollo Multiparadigma* de la Secretaría de Investigación de la Universidad Nacional del Comahue, código 04/E073. Entre los principales objetivos de esta investigación se encuentran establecer las bases para dar origen a una definición definitiva del paradigma orientado a agentes y su relación con los demás paradigmas de programación.

2. Introducción

Según el paradigma orientado a objetos el mundo puede ser representado como un conjunto de objetos que interactúan entre ellos a través del envío de mensajes. Estos mensajes provocan

la ejecución de porciones de código que pueden modificar el estado interno del objeto que lo recibe o enviar a su vez nuevos mensajes a otros objetos. La adhesión estricta a este paradigma demostró ser viable sobre todo con la aparición de Smalltalk [14], un lenguaje orientado a objetos puro que aún hoy es utilizado en parte de la industria del desarrollo de software.

Pero fue debido a la amplia utilización de lenguajes basados en el paradigma estructurado que la programación orientada puramente a objetos no gobernó la industria, y en su lugar lo hizo una fusión entre ambos paradigmas. Ejemplos claros de lenguajes ampliamente difundidos en la actualidad que adoptaron este modelo son Java [15] y C++ [6]. Fue esta combinación de conceptos lo que hizo de ambos paradigmas una base sólida sobre la cual pudieron sostenerse y crecer las herramientas de diseño y desarrollo de software actuales.

Las capacidades cada vez mayores de hardware y el veloz avance de las redes hicieron absolutamente necesario que los programas y sistemas sean capaces de ejecutarse distribuídamente en distintas plataformas, interoperar con otros componentes y aprovechar múltiples unidades de procesamiento. Por este motivo los lenguajes de programación debieron incorporar elementos que les permitiesen modelar esta nueva situación. Fue así como se hicieron comunes conceptos como *procesos livianos* o *lightweight processes*, *hilos de ejecución* o *threads*, *sincronización de procesos*, etc. Se denominó a la programación que se enfocaba en este modelo *programación concurrente* [1] o paralela, dependiendo si se compite por los recursos subyacentes o si pueden utilizarse simultáneamente. Si bien esta manera de programar no fue sustentada por un paradigma bien definido, su importancia actual es definitiva. Poco a poco los sistemas comenzaron a estructurarse a través de entidades autónomas que realizaban sus propios cómputos y se comunicaban entre sí a través de algún tipo de protocolo de intercambio de mensajes. Estas entidades no necesariamente estaban escritas en un mismo lenguaje de programación, o incluso se ejecutaban en plataformas muy distintas. Más tarde, la ingeniería del software denominó a estas entidades autónomas como *agentes*.

Paralelamente, el área de la Inteligencia Artificial adoptó el concepto de agente [12, 17] como una entidad activa que actúa percibiendo distintos estímulos de su entorno y actuando en consecuencia de estas percepciones, con el propósito de alcanzar una o varias metas.

Para el desarrollo de agentes racionales [12] resultó muy adecuado el paradigma lógico, por su claridad para representar estados mentales gracias a su naturaleza declarativa. Así, el comportamiento de un agente podía representarse estableciendo los hechos de su entorno, y codificando las reglas mediante las cuáles podría inferir nuevo conocimiento. Variantes y extensiones interesantes de este paradigma fueron surgiendo [13, 11, 8] para acercarse cada vez más a las estructuras y mecanismos de pensamiento de los seres humanos.

Si bien han habido intentos de acercamiento de estas dos concepciones de *agencia* [7, 5, 10], no creemos que una combinación definitiva haya surgido aún. Esto se demuestra claramente con la falta de acuerdo respecto a cuáles son las características básicas que un agente debe poseer. Como consecuencia de esta falta de claridad, todavía no es común oír hablar del diseño de sistemas en función de agentes, y sí lo es en función de threads, componentes, objetos distribuidos, etc. En esta línea de investigación tenemos como objetivo hallar una combinación adecuada de estos conceptos que de origen al paradigma orientado a agentes definitivo. Consideramos que la mejor manera de encontrar esta fusión es la implementación de un lenguaje de programación que incorpore estos conceptos de manera nativa y clara, permitiendo su utilización en el desarrollo de sistemas inteligentes distribuidos.

3. Líneas de Investigación y Desarrollo

Para alcanzar los objetivos propuestos por esta línea de investigación, se abordarán los ejes principales detallados a continuación.

Características esenciales y deseables de un lenguaje orientado a agentes Determinaremos cuál es el conjunto minimal de características que debe poseer un lenguaje para cumplir con los lineamientos de la orientación a agentes, clasificándolas entre características *inter-agentes* o *exógenas* e *intra-agentes* o *endógenas*. Para esto, analizaremos varias plataformas y lenguajes que actualmente incorporan en alguna medida conceptos de agencia.

Representación del conocimiento y razonamiento mediante *FLUX* Desarrollaremos las estructuras mentales que podrán incorporar los agentes de nuestro lenguaje adaptando el método de programación lógica *FLUX* basado en el *cálculo fluent* [16]. Éste método utiliza *Constraint Handling Rules* [9] para implementar el razonamiento en dominios parcialmente observables, por lo que será necesario su estudio y la eventual colaboración con el grupo de investigación.

Implementación del lenguaje en *Ciao Prolog* Implementaremos el lenguaje sobre *Ciao Prolog* [3], una plataforma de desarrollo y ejecución multiparadigma desarrollada y mantenida por el grupo *CLIP*. Esta herramienta brinda facilidades para incorporar nuevas características mediante las técnicas de *metaprogramación* [4] y *compilación de control* [2]. Además, en versiones recientes de *Ciao Prolog* se incorporaron construcciones para la comunicación entre agentes que nos resultarán útiles para construir sobre ellas nuestro lenguaje permitiendo abstraernos de esas cuestiones.

4. Resultados esperados

A través de la presente línea de investigación pretendemos alcanzar dos resultados fundamentales. Por un lado, establecer las bases sobre las cuáles poder llegar a una definición definitiva y consensuada entre las diferentes áreas de la comunidad de ciencias de la computación de la orientación a agentes. Por otro lado, implementar un lenguaje que respete estrictamente la definición antes mencionada, demostrando su utilidad práctica a través del desarrollo de diversas aplicaciones en diferentes dominios de la vida real.

5. Formación de recursos humanos

La presente línea de investigación constituirá el tema para la tesis de fin de carrera de Maximiliano Klemen, para alcanzar el título de grado de *Licenciatura en Ciencias de la Computación*, proyectándose a realizar estudios de postgrado en colaboración con algunos de los grupos de investigación mencionados anteriormente.

Referencias

- [1] Gregory R. Andrews and Fred B. Schneider. Concepts and notations for concurrent programming. *ACM Computing Surveys*, 15:3–43, 1983.

- [2] Maurice Bruynooghe, Danny de Schreye, and Bruno Krekels. Compiling control. *J. Log. Program.*, 6(1-2):135–162, 1989.
- [3] F. Bueno, D. Cabeza, M. Carro, M. Hermenegildo, P. López, and G. Puebla. *The Ciao Prolog System - A Next Generation Multi-Paradigm Programming Environment*. Grupo Clip, <http://www.ciaohome.org/>, the ciao system documentation series edition, August 2004.
- [4] D. Cabeza and M. Hermenegildo. A New Module System for Prolog. In *International Conference on Computational Logic, CL2000*, LNCS. Springer-Verlag, July 2000.
- [5] S Costantini and A Tocchio. The dali logic programming agent-oriented language. In *Logics in Artificial Intelligence, Proc. of the 9th European Conference, Jelia 2004*, pages 685–688. Springer-Verlag, 2004.
- [6] C++’s web site. <http://www.research.att.com/~bs/C++.html>.
- [7] Winton H E Davies and Peter Edwards. Agent-k: an integration of aop and kqml. In *in Proceedings of the CIKM’94 Intelligent Information Agents Workshop, Y. Labrou & T. Finin (Eds, 1994*.
- [8] Marc Denecker. Ailp: abductive inductive logic programming. In *In Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1201–1207. Morgan Kaufmann, 1995.
- [9] Thom Frühwirth. *J. logic programming 1994:19, 20:1–679 1 theory and practice of constraint handling rules, 1994*.
- [10] Martin L Griss, Steven Fonseca, Dick Cowan, and Robert Kessler. Smartagent: Extending the jade agent behavior model. In *In Proceedings of the Agent Oriented Software Engineering Workshop, Conference in Systemics, Cybernetics and Informatics*. ACM Press, 2002.
- [11] V. Wiktor Marek. Reflexive autoepistemic logic and logic. In *Programming, in L M Pereira & A Nerode (eds), Logic Programming and Non-Monotonic Reasoning*. MIT Press, 1993.
- [12] Stuart Russell and Peter Norvig. *Artificial Intelligence: A modern Approach*. PEARSON Prentice Hall, 2004.
- [13] Chiaki Sakama. Nonmonotonic inductive logic programming. In *In Proceedings of the 6th International Conference on Logic Programming and Nonmonotonic Reasoning*, page 80.
- [14] John F. Shoch. An overview of the programming language smalltalk-72. *ACM SIGPLAN Notices*, 14:64–73, 1979.
- [15] Sun. Java’s official web site. <http://java.sun.com>.
- [16] Michael Thielscher. Programming of reasoning and planning agents with flux. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 435–446. Morgan Kaufmann, 2002.
- [17] Michael Wooldridge and Nicholas R. Jennings. Intelligent agents: Theory and practice. *Knowledge Engineering Review*, 10:115–152, 1995.

Metaheurísticas poblacionales aplicadas a la resolución de problemas complejos

Laura Lanzarini, Javier López
{javierhernanlopez@speedy.com.ar ; laural@lidi.info.unlp.edu.ar}

Instituto de Investigación en Informática LIDI
Facultad de Informática. UNLP.

CONTEXTO

Las investigaciones descritas en esta presentación han sido desarrolladas en el marco del Subproyecto “Sistemas Inteligentes” perteneciente al Proyecto “Algoritmos Distribuidos y Paralelos. Aplicación a Sistemas Inteligentes y Tratamiento Masivo de Datos” del Instituto de Investigación en Informática LIDI.

RESUMEN

Esta línea de trabajo está basada en la investigación y aplicación de metaheurísticas poblacionales para la resolución de problemas multi-objetivo. En particular, se selecciona la metaheurística conocida como Optimización por Cúmulo de Partículas (de aquí en más PSO por sus siglas en inglés de “*Particle Swarm Optimization*”), dado los buenos resultados alcanzados por la misma.

En una primera instancia, la investigación se centra en la optimización de problemas con un solo objetivo, y se propone una nueva versión del algoritmo PSO. Se utiliza un conjunto estandarizado de funciones de prueba con el objeto de realizar una evaluación comparativa entre la versión propuesta y la versión original de la metaheurística.

En una segunda instancia se expande el alcance de la investigación, abordando problemas con dos o más objetivos, sujeto a restricciones, y se espera proponer una variante inédita del algoritmo PSO. Se empleará un conjunto estandarizado de funciones de prueba multi-objetivo con el propósito de comparar el rendimiento de la versión propuesta con otros populares métodos de resolución representativos del estado del arte.

Por último, se considera la utilización de los algoritmos desarrollados para resolver problemas del mundo real del tipo de planificación de actividades y asignación de recursos.

Palabras clave: *Optimización Multi-objetivo, Optimización por cúmulo de partículas, Scheduling*

1. INTRODUCCIÓN

La optimización de problemas es un proceso muy frecuente en la vida real. Permanentemente, nos vemos ante la exigencia de encontrar soluciones que maximicen o minimicen algún recurso escaso. Cuando se trata de situaciones sencillas, el cerebro humano es capaz de resolverlas directamente pero a medida que se incrementa la complejidad del problema, es preciso contar con herramientas adicionales. En esta dirección, existe una amplia variedad de técnicas que pueden clasificarse en exactas y aproximadas.

Las técnicas exactas, si bien garantizan la obtención de la solución óptima para cualquier tipo de problema, requieren un costo computacional elevado. Esto ha volcado el interés hacia las técnicas aproximadas que permiten obtener una solución “casi” óptima en un tiempo “razonable”; en particular al uso de las metaheurísticas a las que se puede definir como “*métodos que integran de diversas maneras, procedimientos de mejora local y estrategias de alto nivel para crear un proceso capaz de escapar de óptimos locales y realizar una búsqueda robusta en el espacio de búsqueda.*”

En su evolución, estos métodos han incorporado diferentes estrategias para evitar la convergencia a óptimos locales, especialmente en espacios de búsqueda complejos” [4].

Una primera clasificación para las metaheurísticas está dada por el número de soluciones que procesan en un determinado momento durante el proceso de búsqueda, es decir, aquellas “basadas en trayectoria” y las “basadas en población” o poblacionales. En el primer caso, se caracterizan por la mejora continua de la solución actual mediante la inspección de un vecindario, mientras que las metaheurísticas basadas en población trabajan con un conjunto de soluciones que iterativamente son mejoradas a través de un proceso inteligente de exploración del espacio de búsqueda.

Dentro del conjunto de metaheurísticas poblacionales, la más difundida corresponde a las técnicas basadas en la Computación Evolutiva: Algoritmos Genéticos (AGs), Estrategias Evolutivas (EEs), Programación Evolutiva (PE) y Programación Genética (PG). Todos estos, generalmente referidos como Algoritmos Evolutivos (AEs) [8]. Sin embargo, en los últimos años han surgido diferentes enfoques alternativos, que si bien presentan algunas de las características de los AEs, tienen rasgos particulares que los diferencian de estos últimos. Dentro de estos podemos mencionar a Optimización mediante Cúmulo de Partículas (PSO) [1] [2], Evolución Diferencial (DE) [10], Optimización por Colonia de Hormigas (ACO) [3] o Algoritmos de Estimación de Distribución (EDA) [9]. Estas metaheurísticas han mostrado ser sumamente eficientes cuando se aplican a problemas complejos de laboratorio o del mundo real [7]. Si bien en ningún caso aseguran la obtención del mínimo o máximo global, han mostrado que son capaces de alcanzar soluciones de calidad en periodos de tiempo aceptables.

PSO es una metaheurística relativamente nueva que ha sido utilizada exitosamente en una amplia variedad de problemas de optimización. El diseño de PSO se basa en el comportamiento de grupos (generalmente en movimiento) o *swarm* de organismos biológicos que se comunican entre sí (en forma directa o indirecta) dentro de su hábitat (e.g., cardúmenes, bandadas, enjambres, colonias de hormigas, etc.)

El comportamiento destinado a la resolución de problemas y que emerge de la interacción de los agentes, conforman lo que se denomina “Swarm Intelligence” o Inteligencia Colectiva, mientras que los algoritmos diseñados siguiendo estos conceptos, son agrupados bajo el término “Inteligencia Computacional Colectiva”. Aplicaciones de PSO incluyen una amplia variedad de problemas continuos y discretos, sin embargo, en su concepción original, su diseño está orientado a resolver problemas mono-objetivos.

Por otro lado, es claro que muchos problemas del mundo real son multi-objetivo por naturaleza; habitualmente las decisiones a tomar no dependen de un único criterio y es por este motivo que resulta de interés el estudio de metaheurísticas que estén diseñadas para manejar esta característica.

Dado que la versión original de PSO no puede utilizarse en problemas multi-objetivo, se han desarrollado distintas modificaciones al respecto. En este punto es importante destacar que a diferencia de los problemas mono-objetivo (y dentro de éstos los uni-modales), en un problema con dos a más objetivos no existe una única solución mejor que optimice los respectivos objetivos involucrados. Por el contrario, una buena técnica para resolverlos es aquella que sea capaz de encontrar la mayor cantidad de “mejores” soluciones que cumplan satisfactoriamente con los objetivos implicados. En este sentido, para evitar una posible pérdida de diversidad en el proceso de búsqueda por parte de una metaheurística poblacional, es necesario proveer de algún mecanismo adecuado a fin de obtener un conjunto mayor de soluciones de buena calidad.

PSO como muchas otras metaheurísticas poblacionales, necesita de mecanismos eficientes para promover una adecuada exploración del espacio de búsqueda que en el caso de los problemas multi-objetivo, se vería reflejada por la diversidad del conjunto de soluciones óptimas. En [5] [6] [7] puede consultarse sobre distintos mecanismos para controlar la diversidad de la población en el contexto de Algoritmos Evolutivos que podrían ser aplicados o que pueden servir como base para ser usados en el contexto de PSO.

Si nos referimos a optimización mono-objetivo, existen diferentes versiones desarrolladas a partir de la idea original de PSO, la mayoría de ellas relacionadas con variación en los diferentes parámetros del algoritmo o combinación y variación de diferentes topologías, tamaños y cantidad de poblaciones. [13][16]. Incluso existen versiones con agregados de nuevos operadores inspirados en otras metaheurísticas evolutivas [14], y se han propuesto además variantes del algoritmo con cambios en la fórmula de actualización de velocidad, con el objetivo de mantener la diversidad y evitar el estancamiento en el proceso de búsqueda [12] [17] [18].

En esta línea de investigación se define y evalúa una nueva versión que utiliza el esquema general de funcionamiento propuesto por la metaheurística original, con el agregado de un método de búsqueda local que se activa de forma automática dependiendo de las características de movimiento de las partículas.

Se plantea como objetivo hacer más eficiente el proceso de búsqueda en la fase de aproximación final de las partículas a óptimos locales (globales). De acuerdo a trabajos anteriores realizados por otros autores [11], se encontró que este proceso de acercamiento final es ineficiente en la versión original del algoritmo PSO, utilizando esta versión muchas iteraciones para lograr aproximar el óptimo local (global).

Con esta consideración en mente, surgen dos cuestiones principales a resolver

1. Descubrir cuando el algoritmo está oscilando alrededor de algún óptimo local (global)
2. Desarrollar un procedimiento que haga más eficiente la aproximación final de la partícula al óptimo.

Para resolver el punto 1), se incorporó un procedimiento para detectar las oscilaciones, que analiza como varía el signo del componente velocidad de cada partícula en cada iteración. Para resolver el punto 2), se incorporó un procedimiento de búsqueda local determinista. El mismo intenta encontrar en la menor cantidad de iteraciones posibles donde se halla el óptimo.

De acuerdo a las pruebas realizadas, se puede afirmar que se alcanzó el objetivo previsto, dado que la variante propuesta igualó o mejoró el rendimiento mostrado por la versión canónica del algoritmo en las 10 funciones de prueba utilizadas [21].

El paso siguiente de esta línea de investigación será el análisis y la posterior propuesta de una nueva versión de un algoritmo del tipo MOPSO (Multi Objective Particle Swarm Optimization) [19], teniendo en consideración como premisas centrales de trabajo la manutención de la diversidad en la población, así como la eficiencia en la aproximación al conjunto de soluciones óptimas. Estas dos características permitirán que el algoritmo sea capaz de encontrar soluciones no dominadas preferentemente con distribución uniforme a lo largo del frente de Pareto [20]. Se efectuarán comparaciones de rendimiento de este algoritmo con aquellos métodos de optimización y búsqueda representativos del estado del arte [19] en optimización multi-objetivo, utilizando un conjunto de pruebas estándar desarrollado para tal propósito,

Como aportación final de esta línea de investigación, se avanzará en la implantación de los algoritmos estudiados en problemas del mundo real, aplicado a los procesos de producción de la industria de televisión. Existen antecedentes aplicados a la problemática de “*scheduling*” [15], pero no en la industria referida en particular, de acuerdo a lo mejor de nuestros conocimientos. Esta problemática de planificación y asignación de recursos se caracteriza por ser dinámica, de corto plazo, requiere múltiples recursos calificados escasos, estacional y administra recursos de terceros como propios.

2. LINEAS DE INVESTIGACION Y DESARROLLO

- PSO aplicado problemas mono-objetivo.
 - Estudio del estado del arte.

- Propuesta de nueva variante.
- Evaluación de rendimiento en funciones de laboratorio.
- PSO aplicado a problemas multi-objetivo con restricciones (MOPSO).
 - Estudio del estado del arte.
 - Propuesta de nueva variante.
 - Evaluación de rendimiento en funciones de laboratorio.
- Aplicaciones de la metaheurística PSO a problemas del mundo real, dentro del área de planificación y asignación de recursos.

3. RESULTADOS OBTENIDOS/ESPERADOS

La versión propuesta del algoritmo PSO dirigido a la resolución de problemas mono-objetivos mostró resultados prometedores. El desempeño de esta versión se reveló superior a la versión PSO original. En 5 de los 10 problemas de prueba, se alcanzó un resultado superior. En los 5 problemas restantes, no se pudo determinar cual fue el algoritmo con mejor desempeño. Los resultados obtenidos demuestran que se alcanzaron los objetivos planteados, y que esta nueva versión del algoritmo PSO exhibe un comportamiento mejor que la versión original del mismo, por lo menos en el conjunto de pruebas utilizado.

Respecto de la versión propuesta para la optimización multi-objetivo, la misma se encuentra en la fase de análisis y definición. Se prevé realizar una evaluación de rendimiento utilizando las funciones de prueba desarrolladas para este tópico. Se espera, al menos, alcanzar el rendimiento observado en los métodos representativos del estado del arte en esta rama de investigación.

Por último, con la aplicación de los métodos anteriores a la problemática del mundo real, se espera hacer mas eficiente los procesos de producción de la industria de televisión, utilizando una menor cantidad de recursos (mano de obra, equipos y locaciones) manteniendo los niveles de calidad requeridos.

4. FORMACIÓN DE RECURSOS HUMANOS

Dentro de los temas involucrados en esta línea de investigación se está desarrollando actualmente una tesis de Doctorado en Ciencias de la Computación dentro del ámbito de la Universidad Nacional de La Plata, Facultad de Informática.

5. BIBLIOGRAFIA

- [1] Engelbrecht A. **Fundamentals of Computational Swarm Intelligence** , John Wiley & Sons, Ltd, 2005.
- [2] Kennedy J., Eberhart R. **Particle Swarm Optimization**, in Proceedings of the IEEE International Joint Conference on Neural Networks, pp 1942-1948, IEEE Press, 1995.
- [3] Dorigo M., Stützle T. **Ant Colony Optimization** , MIT Press, 2004.
- [4] Glover F., Kochenberger G. **Handbook of Metaheuristics** , Kluwer Academic Publishers, 2003.
- [5] Deb G. **Multi-objective Optimization using Evolutionary Algorithms**, John Wiley & Sons, Ltd. 2001.
- [6] Coello Coello C. **EMOO Web Page**, <http://delta.cs.cinvestav.mx/~ccoello/EMOO/>
- [7] Jain A., Goldberg R. **Evolutionary Multiobjective Optimization. Theoretical Advances and Applications**, Springer, USA.
- [8] Bäck T. , Fogel T., Michalewicz Z. **Evolutionary Computation I** , Institute of Physics Publishing, Bristol and Philadelphia, 2000.
- [9] Larraña P., Lozano J. **Estimation of Distribution Algorithms - A New Tool for Evolutionary Computation**, Kluwer Academic Publishers, 2002.

- [10] Price K., Storn R, Lampinen J. **Differential Evolution**, Springer-Verlag, Berlin Heidelberg, 2005.
- [11] Monson C. **No Free Lunch, Bayesian Inference, and Utility: A Decision-Theoretic Approach to Optimization**, *Ph.D. dissertation*. Department of Computer Science. Brigham Young University. August 2006
- [12] Van den Bergh F. **An Analysis of Particle Swarm Optimizers**, *Ph.D. dissertation*. Department Computer Science. University Pretoria. South Africa. 2002
- [13] Esquivel S., Cagnina L. **Global Numerical Optimization with a bi-population Particle Swarm Optimizer**, LIDIC (Research Group) - Universidad Nacional de San Luis
- [14] Zhang W., Xiao-Feng Xie **DEPSO: Hybrid Particle Swarm with Differential Evolution Operator**, Institute of Microelectronics, Tsinghua University - IEEE International Conference on Systems, Man & Cybernetics (SMCC), Washington D C, USA, 2003: 3816-3821
- [15] Grobler J, Engelbrecht A, Joubert J., Kok S. **A starting-time-based approach to production scheduling with Particle Swarm Optimization**, Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Scheduling (CI-Sched 2007)
- [16] Lanzarini L., Leza V., De Giusti A. **Particle Swarm Optimization with Variable Population Size**, *Proceedings of the 9th international conference on Artificial Intelligence and Soft Computing Zakopane, Poland*. Section: Evolutionary Algorithms and Their Applications pp: 438 - 449 Year of Publication: 2006
- [17] Riget J., Vesterstrøm J. **A Diversity-Guided Particle Swarm Optimizer – the ARPSO**. *Technical Report 2002-02*, Department of Computer Science, University of Aarhus, 2002.
- [18] Peer E.S., Van Den Bergh F., Engelbrecht A. Using **Neighbourhoods with the Guaranteed Convergence PSO**, *Swarm Intelligence Symposium, 2003*. SIS '03. Proceedings of the 2003 IEEE pp: 235-242 ISBN: 0-7803-7914-4
- [19] Reyes-Sierra M, Coello Coello C. **Multi-Objective Particle Swarm Optimizers: A Survey of the State-of-the-Art**, International Journal of Computational Intelligence Research. ISSN 0973-1873 Vol.2, No.3 (2006), pp. 287–308 Research India Publications <http://www.ijcir.info>
- [20] Dupont G., Adam S., Lecourtier Y., Grilheres B. **Multi objective particle swarm optimization using enhanced dominance and guide selection**, International Journal of Computational Intelligence Research. ISSN 0973-1873 Vol.4, No.2 (2008), pp.145–158 © Research India Publications <http://www.ijcir.info>
- [21] Suganthan P., Hansen N., Liang J., Deb K., Chen Y., Auger A. , Tiwari S. **Problem Definitions and Evaluation Criteria for de CEC 2005 Special Session on Real-Parameter Optimization**. *Special Session on Real-Parameter Optimization at CEC-05*, Edinburgh, UK, 2-5 Sept. 2005

Sistemas Inteligentes. Aplicaciones en Minería de Datos, Procesamiento de Señales y Redes de Computadoras

Laura Lanzarini¹, Waldo Hasperue², Leonardo Corbalán³, César Estrebou⁴, Juan La Battaglia⁵

Instituto de Investigación en Informática LIDI (III-LIDI)⁶
Facultad de Informática. UNLP

CONTEXTO

Esta presentación corresponde al Subproyecto “Sistemas Inteligentes” perteneciente al Proyecto “Algoritmos Distribuidos y Paralelos. Aplicación a Sistemas Inteligentes y Tratamiento Masivo de Datos” del Instituto de Investigación en Informática LIDI.

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes basados en mecanismos de adaptación. Actualmente el énfasis está puesto en la transferencia de tecnología a las áreas de minería de datos, procesamiento de señales y redes de computadoras.

En el área de la Minería de Datos, los temas centrales se encuentran relacionados con la investigación de nuevas estrategias adaptativas que generen modelos a partir de grandes volúmenes de datos y además sean capaces de acomodar su estructura ante algún cambio en los datos y/o nueva información que se obtenga, reflejando estos cambios en el conocimiento actual adquirido.

En el área del procesamiento de señales, el énfasis está puesto en la combinación de las técnicas habitualmente utilizadas con estrategias adaptativas inteligentes que permitan desarrollar aplicaciones con capacidades para reconocer e identificar objetos o personas en tiempo real. Se trabaja en el desarrollo de métodos adaptativos factibles de ser aplicables en distintos contextos.

La aplicación de distintas metaheurísticas sobre redes Peer-to-Peer (P2P), es otra línea de investigación que se está llevando a cabo en el III-LIDI. El objetivo central es mejorar la capacidad de adecuación del sistema a los cambios rápidos del entorno de información.

Palabras claves : Redes Neuronales, Algoritmos Evolutivos, Minería de Datos, Técnicas de Optimización, Procesamiento de Señales.

1. INTRODUCCIÓN

Los Sistemas Inteligentes han demostrado ser herramientas sumamente útiles en la resolución de problemas complejos. Su capacidad de adaptación al entorno de información les permite brindar buenos resultados en distintas áreas.

En el Instituto de Investigación en Informática LIDI se está trabajando en este tema desde hace varios años. Inicialmente se desarrollaron estrategias basadas en Redes Neuronales y Algoritmos Evolutivos aplicables al Reconocimiento de Patrones [Lan00, Lan04] así como al control de agentes autónomos [Cor03a, Cor03b, Oli05]. Luego se profundizó en el estudio de la capacidad de caracterización de este tipo de estrategias a fin de establecer un modelo de la

¹ Profesor Titular. Facultad de Informática. UNLP

² Becario de Perfeccionamiento de la CIC – Ayudante Diplomado - Facultad de Informática. UNLP

³ Becario de Formación Superior de la UNLP. Jefe de Trabajos Prácticos - Facultad de Informática. UNLP

⁴ Jefe de Trabajos Prácticos DS - Facultad de Informática. UNLP

⁵ Becario de Iniciación de la UNLP - Facultad de Informática. UNLP

⁶ Calle 50 y 115 1er Piso, (1900) La Plata, Argentina, TE/Fax +(54) (221) 422-7707. <http://weblidi.info.unlp.edu.ar>

información disponible. Se definieron y desarrollaron varias estrategias basadas en redes neuronales competitivas con aplicaciones concretas en distintas áreas [Has07a, Cor06a, Cor06b]. Actualmente, con el objetivo de mejorar la transferencia tecnológica de los resultados obtenidos, el énfasis está puesto por un lado, en la definición de estrategias que faciliten la interpretación del modelo y por otro en la incorporación de mecanismos de adaptación a entornos dinámicos.

A continuación se detallan brevemente los avances realizados últimamente.

1.1. Minería de Datos

Las investigaciones realizadas en el III-LIDI relacionadas con diferentes mecanismos de aprendizaje y adaptación de redes neuronales competitivas aplicables a la Minería de Datos han facilitado la construcción de modelos de la información disponible principalmente a partir de reglas de asociación y clasificación [Has05, Has06, Has07]. Sin embargo, existen situaciones en las cuales el modelo por si solo es incapaz de transmitir al usuario el conocimiento adquirido así como su importancia y relación con los datos subyacentes. En esta dirección, las técnicas visuales resultan una manera intuitiva de presentación, dándole una perspectiva exploratoria y contextual.

El gran obstáculo que actualmente presenta la transferencia tecnológica en temas referidos a estrategias de minería de datos, es el desconocimiento por parte del usuario acerca de la forma de adaptar el modelo a sus necesidades. Como forma de ayudar a revertir esta situación, se han desarrollado estrategias que definen, a partir del modelo basado en reglas de clasificación, las acciones a seguir para lograr el beneficio esperado [Has07b].

Otro aspecto importante a considerar es el dinamismo de la información. Hoy en día, la tecnología posibilita el almacenamiento de enormes volúmenes de datos. Gran parte de las estrategias inteligentes existentes permiten modelizarlos para un instante de tiempo dado. Sin embargo, existe una amplia gama de problemas que requiere disponer de mecanismos capaces de adaptar el modelo existente a los cambios del entorno. En [Has08] se ha generado un método de obtención de reglas difusas las cuales facilitan su adaptación ante nueva información.

Siguiendo con esta línea de trabajo se espera conseguir nuevos métodos y técnicas de minería de datos que sean capaces de generar conocimiento útil, produciendo resultados que sean de provecho al usuario final.

1.2. Procesamiento de Señales

En el III-LIDI interesa especialmente la combinación de técnicas de procesamiento de señales con estrategias adaptativas inteligentes que permitan desarrollar aplicaciones con capacidades para reconocer e identificar objetos o personas en tiempo real. Este tipo de aplicaciones resulta de sumo interés en área tales como seguridad y telefonía celular, entre otras.

En esta línea, se han realizado algunas investigaciones sobre señales de voz a fin de verificar la identidad de la persona que habla. Este es un problema difícil de resolver debido a la variabilidad de la información de entrada; no sólo porque dos personas distintas pueden tener un registro parecido sino porque una misma persona puede generar dos registros diferentes en distintos instantes de tiempo. Esto motiva la necesidad de utilizar herramientas adaptativas capaces de procesar información dinámica.

En el III-LIDI se han utilizado redes neuronales competitivas dinámicas para modelizar las voces de un conjunto de personas a reconocer. El entrenamiento se realizó a partir de los coeficientes ceptrales de las señales de voz logrando un buen reconocimiento al aplicarse tanto a voces sintéticas como reales.

También se está trabajando en el desarrollo de un protocolo de comunicación para teléfonos celulares con capacidad de transmisión de voz y audio en un formato propietario reducido y eficiente.

1.3. Redes de Computadoras

La aplicación de distintas metaheurísticas sobre redes Peer-to-Peer (P2P), es otra línea de investigación que se está llevando a cabo en el III-LIDI. Las redes P2P constituyen un sistema distribuido, descentralizado y sumamente dinámico, donde la información disponible cambia con mucha rapidez, al igual que la topología misma de la red. El objetivo central de esta investigación es mejorar la capacidad de adecuación del sistema a los cambios rápidos del entorno de información utilizando metaheurísticas poblacionales.

Si bien existen soluciones previas [Shi07] basadas en cúmulos de partículas para determinar la interconexión de los nodos en redes P2P con algunas restricciones, investigaciones llevadas a cabo en el III-LIDI han permitido definir y desarrollar una extensión original de PSO que, a diferencia de [Ken95][Shi99], permite trabajar con una población de tamaño variable. De esta forma, no es necesario definir a priori la cantidad de soluciones a utilizar, evitando así condicionar la calidad de la solución a obtener [Lan08].

Para la búsqueda de recursos en sistemas P2P se ha propuesto una estrategia que utiliza redes neuronales locales para asistir a los nodos en la propagación selectiva de solicitudes dirigidas sólo a los vecinos más adecuados. El algoritmo de aprendizaje tradicional de las redes neuronales que se utilizan ha sido reemplazado por un apropiado intercambio de información entre vecinos que permite acelerar los tiempos de entrenamiento y mantener actualizado el conocimiento adquirido por cada nodo [Cor09].

Actualmente se está trabajando en la reducción del costo computacional de la solución propuesta. Cuando la información disponible se encuentra distribuida, los modelos globales resultan ineficientes. Por tal motivo se propone dividir el proceso en dos partes, una primera parte que utiliza un algoritmo local muy eficiente para modelizar la información cercana y luego una segunda etapa que utiliza la información generada localmente para monitorear el sistema completo. Con esto se espera lograr una reducción considerable en el tiempo de adaptación y una mayor flexibilidad a los cambios de la información en cada nodo.

2. TEMAS DE INVESTIGACIÓN Y DESARROLLO

- Estudio y análisis de diferentes estructuras de modelización dinámicas. Interesa especialmente la obtención de árboles de decisión incluyendo los árboles difusos e incrementales.
- Desarrollo y aplicación de diferentes métricas que permitan analizar el conjunto de reglas a utilizar para representar el modelo. Esto incluye considerar distintas estrategias de poda que permitan maximizar la representación obtenida.
- Desarrollo e implementación, a partir de los métodos existentes, de estrategias adaptativas capaces de construir y mantener modelos adecuados en entornos de información dinámicos.
- Análisis de los distintos tipos de Redes Neuronales competitivas dinámicas. Estudio de las estrategias existentes que permiten determinar, durante la adaptación, el tamaño de la arquitectura y forma de conexión de los elementos que componen la red neuronal.

- Estudio y aplicación de diferentes de métricas que permitan analizar la preservación de la topología de los datos tanto en el espacio de los patrones de entrada como en el espacio de salida de la red.
- Estudio de distintas metaheurísticas de optimización basadas en trayectoria y en población aplicables al problema de ruteo en redes P2P.
- Análisis de la importancia de los parámetros de las distintas metaheurísticas en el funcionamiento y eficiencia de la estrategia seleccionada. Análisis de la función de aptitud a utilizar en el caso de redes P2P.
- Estudios de performance de los algoritmos desarrollados. Análisis de eficiencia en la resolución de problemas concretos.

3. RESULTADOS OBTENIDOS/ ESPERADOS.

- Desarrollo e implementación de una estrategia para la obtención de reglas de clasificación a partir de una matriz de co-asociación.
- Desarrollo e implementación de mecanismos que permitan la especificación de un conjunto de acciones a seguir a fin de objetivar la interpretación de la información modelizada.
- Desarrollo e implementación de una red neuronal entrenada a partir del método AVGSOM como herramienta para reconocer patrones de señales de voz.
- Desarrollo e implementación de un sistema de votación a partir de varias redes neuronales competitivas dinámicas que permite reducir el error en la etapa de reconocimiento.
- Desarrollo e implementación de una estrategia basada en cúmulos de partículas (PSO) con tamaño de población variable basado en los conceptos de edad y vecindario. Se ha comprobado que el mecanismo utilizado para incorporar nuevos individuos así como la forma de calcular el tiempo de vida preserva la diversidad de la población.
- Resolución de problemas concretos, tanto en ambientes simulados como en el mundo real. En este último caso, resulta de fundamental importancia la optimización del algoritmo propuesto.

4. FORMACIÓN DE RECURSOS HUMANOS

Dentro de los temas involucrados en esta línea de investigación se están desarrollando actualmente 2 tesis de doctorado, 1 una de maestría y al menos 2 tesinas de grado de Licenciatura. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

5. REFERENCIAS

- [Cor03a] Corbalán, Lanzarini. GNE. Grupos Neuronales Evolutivos. *XXIX Conferencia Latinoamericana de Informática*. CLEI 2003. Bolivia. 2003
- [Cor03b] Corbalán, Lanzarini. Evolving Neural Arrays A new mechanism for learning complex action sequences. *Special Issue of Best Papers presented at CLEI'2002*. Volumen 6. Number 1, December 2003. Montevideo, Uruguay.
- [Cor05] Corbalán, Lanzarini, De Giusti A. ALENA. Adaptive-Length Evolving Neural Arrays". *Journal of Computer Science and Technology*. Vol 5, nro. 4. 2005. Pags. 59-65.

- [Cor06a] Corbalán, Osella Massa, Russo, Lanzarini. Image recovery using a new nonlinear Adaptive Filter based on Neural Networks. *CIT - Journal of Computing and Information Technology*. Vol. 14, No.4. December 2006 - Pág. 315-320.
- [Cor06b] Corbalán, Hasperue, Osella Massa, Lanzarini. BPNn-CPN. Nuevo método para segmentación de Imágenes basado en Redes Neuronales Artificiales. *IV Workshop de Computación Gráfica, Imágenes y Visualización (WCGIV)*. CACIC 2006. San Luis. Argentina. Octubre de 2006.
- [Cor09] Corbalán, Lanzarini, De Giusti. Resources NeuroSearch in Peer-to-Peer Networks. *31th International Conference of Information Technology Interfaces (ITI)*. Cavtat, Croacia. June 2009
- [Has05] Hasperué, Lanzarini. Dynamic Self-Organizing Maps. A new strategy to enhance topology preservation. *XXXI Conferencia Latinoamericana de Informática*. CLEI 2005.
- [Has06] Hasperué, Lanzarini. Classification Rules obtained from Dynamic Self-organizing Maps. *VII Workshop de Agentes y Sistemas Inteligentes*. CACIC 2006. San Luis. Argentina. Octubre de 2006.
- [Has07a] Hasperue, Corbalán, Lanzarini, Bría. Skeletonization of Sparse Shapes using Dynamic Competitive Neural Networks. *Ibero-American Journal of Artificial Intelligence*. Vol.11. Nro.35. pp.33-42 2007.
- [Has07b] Hasperué, Lanzarini. Extracting Actions from Classification Rules. *Workshop de Inteligencia Artificial. Jornadas Chilenas de Computación 2007*. Iquique, Chile. Noviembre de 2007.
- [Has08] Hasperué, W., G. Osella Massa & L. Lanzarini. Obtaining a Fuzzy Classification Rule System from a non-supervised Clustering. *30th International Conference of Information Technology Interfaces (ITI)*. Cavtat, Croacia. June 2008.
- [Ken95] Kenedy, Eberhart. Particle Swarm Optimization. *IEEE International Conference on Neural Networks*. Vol IV, pp.1942-1948. Australia 1995.
- [Lan00] Lanzarini. Reconocimiento de Patrones en Imágenes Médicas utilizando Redes Neuronales. *Journal of Computer Science and Technology*. Vol.4 . Dic 2000.
- [Lan04] Lanzarini, Yanivello. Reconocimiento de Comandos Gestuales utilizando GesRN. *V Workshop de Agentes y Sistemas Inteligentes*. CACIC 2004. Bs.As. Argentina. 2004. ISBN 987-9495-58-6
- [Lan08] Lanzarini, Leza, De Giusti. Particle Swarm Optimization with Variable Population Size. *Lecture Notes in Computer Science*. Vol 5097/2008. Artificial Intelligence and Soft Computing. Pags. 438-449. Junio de 2008. ISBN 987-3-540-69572-1.
- [Oli05] Olivera y Lanzarini. Cyclic Evolution. A new strategy for improving controllers obtained by layered evolution. *Journal of Computer Science and Technology*. Vol 4, nro. 1. 2005. Pags 211-217.
- [Shi07] Shichang, Ajito, Guiyong, Hongbo. A Particle Swarm Optimization Algorithm for Neighbor Selection in Peer-to-Peer Networks. *6th International Conference on Computer Information. Systems and Industrial Management Applications (CISIM'07)*. Pp. 166-172. ISBN 0-7695-2894-5. June 2007.
- [Shi99] Shi Y., Eberhart R. An empirical study of particle swarm optimization. *IEEE Congress Evolutionary Computation*. pp.1945-1949. Washington DC, 1999.

GENERADOR DE MOTORES DE SCHEDULING EN DOMINIOS INDUSTRIALES

Daniel Díaz, Francisco Ibañez, Raymundo Forradellas

LISI- Instituto de Informática – Dpto. de Informática

FCEfN - Universidad Nacional de San Juan

CUIM – Av. Ignacio de la Roza 590 (O), Rivadavia – J5402DCS San Juan

{ddiaz, fibanez, kike}@iinfo.unsj.edu.ar

CONTEXTO

Este trabajo está encuadrado en una línea de I+D+i que este grupo de trabajo desarrolla en el LISI (Laboratorio Integrado de Sistemas Inteligentes) desde 1996, orientado a resolver problemas complejos de Planning, Scheduling y Asignación de Recursos, en el ámbito de los sistemas industriales. El trabajo actual guarda relación con esta línea, incorporando nuevas tecnologías y funcionalidades provenientes de la ingeniería de software.

RESUMEN

En este trabajo se describe una propuesta para desarrollar sistemas de scheduling en el dominio de sistemas industriales, utilizando como base un generador de motores de scheduling, para una instancia específica de un dominio determinado. El constructor de motores de scheduling utiliza una combinación de métodos provenientes de la ingeniería del software y de la inteligencia artificial. Más específicamente, se presenta una integración entre SPL - *Software Product Line*, y la programación con restricciones.

Palabras clave: Scheduling, Inteligencia Artificial, Reutilización de software, Software Product Line.

1 INTRODUCCION

Existen varias definiciones para scheduling: “Scheduling es el problema de asignar recursos limitados a tareas en el tiempo con el objeto de optimizar uno o más objetivos” (Barker, 1974). “Scheduling trata con la asignación temporal de trabajos a recursos limitados donde se debe considerar un conjunto de restricciones” (Sauer, 2000b).

Los problemas de scheduling aparecen en una gran cantidad de dominios, en este artículo tratamos con los problemas de scheduling en la industria, los cuales se conocen como *production scheduling* o programación de la producción.

Considerando la definición del problema de scheduling, la solución constituye la asignación temporal de máquinas a trabajos (referidos también como actividades o tareas). En la industria, un sistema de scheduling es el corazón del sistema de planificación y control de la producción (Vollmann et al., 1997). Normalmente el sistema de scheduling interactúa con otros sistemas de fabricación. Por ejemplo, un sistema de scheduling necesita conocer la capacidad de los recursos y disponibilidad de la materia prima. Este tipo de información la brinda el MRP - *Material requirement planning* (Orlicky, 1975). El sistema de scheduling genera como salida un plan de producción, el cual es puesto en práctica por la planta.

El sistema encargado de gestionar la operación de planta es el MES - *Manufacturing Execution System* (Boucher and Yalcin, 2006). Este sistema interactúa con el sistema de scheduling a fin de adecuar el scheduling predictivo al scheduling real (re-scheduling). Los sistemas contenidos en el ERP - *Enterprise Resource Planning* (Shanks and Seddon, 2000) también pueden interactuar con sistema de scheduling. A modo de ejemplos, ventas puede observar a futuro la carga de máquinas y así ajustar su política de ventas, los mantenimientos preventivos se pueden planificar en función de las simulaciones de carga de máquinas obtenidas desde el sistema de scheduling.

Según (Yen and Pinedo, 1994), un sistema de scheduling se compone diferentes módulos: (1) módulos de base datos y base de conocimiento, (2) módulos del motor de scheduling y (3) módulo de interface de usuario.

1.1 EVOLUCIÓN DE LOS SISTEMAS DE SCHEDULING

La investigación en el diseño y desarrollo de sistemas de scheduling se puede clasificar en 4

generaciones. Inicialmente, esta clase de problemas fue abordada por el área de la ingeniería industrial, y posteriormente por el dominio de la investigación operativa. En los comienzos del siglo pasado, Henry L. Gantt (Gantt, 1903), fue uno de los primeros en proponer un sistema de scheduling: *daily balance* “el equilibrio diario” que consiste en un método de programar y de registrar el trabajo. En (MacNiece, 1951), se discuten dos técnicas de scheduling *Loading*, y *Planning Board*. En 1960 se desarrolló *Planalog Control Board*, una herramienta que constaba de una pizarra y entre otras cosas podía forzar restricciones de precedencia. Los primeros sistemas de scheduling computarizados nacieron a la par del MRP - *Material Requirement Planning*, (Wight, 1984). Estos fueron desarrollados con las primeras tecnologías de hw y sw disponibles. En aquel entonces el desarrollo requería un gran esfuerzo y sus costos y tiempos eran elevados. En la segunda generación surgen los ISS - *Intelligent Scheduling Systems*, tales como ISIS (Fox and Smith, 1984) y OPIS de (Smith, 1994). Esta fue una etapa donde el problema de scheduling se comenzó a abordar con técnicas de inteligencia artificial tales como algoritmos genéticos, *simulated annealing*, redes neuronales y *constraint satisfaction techniques*. En esta generación aparecen los sistemas de scheduling que se construyen utilizando técnicas de KBS - *Knowledge Based Systems* (Sauer, 2000a). La tercera generación se inició con el surgimiento del paradigma orientado, dando origen a sistemas de scheduling desarrollados con esta tecnología (Yen and Pinedo, 1994). Esto permitió el desarrollo de *frameworks* especializados en scheduling, los cuales encapsulan técnicas de inteligencia artificial (Le Pape, 1994).

Una cuarta generación de sistemas de scheduling está surgiendo actualmente. En esta etapa, el constante avance tecnológico hace que la arquitectura de sistemas de la empresa crezca día a día y esto requiere que los sistemas sean integrables, flexibles, de rápida implementación y bajo mantenimiento. La flexibilidad tiene un alto impacto en los costos de mantenimiento, dado que modificar un

sistema inflexible suele ser costoso, tanto en tiempo como en dinero. Últimamente la reutilización sistemática está dando lugar a esta cuarta generación de sistemas de scheduling, así lo demuestran los últimos trabajos de investigación sobre sistemas de scheduling en el área de KBS (Rajpathak et al., 2003). Esta propuesta adhiere a esta tendencia, usando técnicas de inteligencia artificial embebidas en *frameworks* orientados a objetos para resolver problemas de scheduling, y por otro lado, utilizando los últimos avances en la ingeniería de software, como son el desarrollo de familias de productos de software y SPL, para modelar, diseñar, implementar sistemas de scheduling.

1.2 PROBLEMA DE INVESTIGACIÓN

El objetivo principal del tema de investigación es demostrar que dado un dominio industrial, es posible desarrollar un proceso constructor de motores de scheduling, considerando una instancia X del dominio industrial.

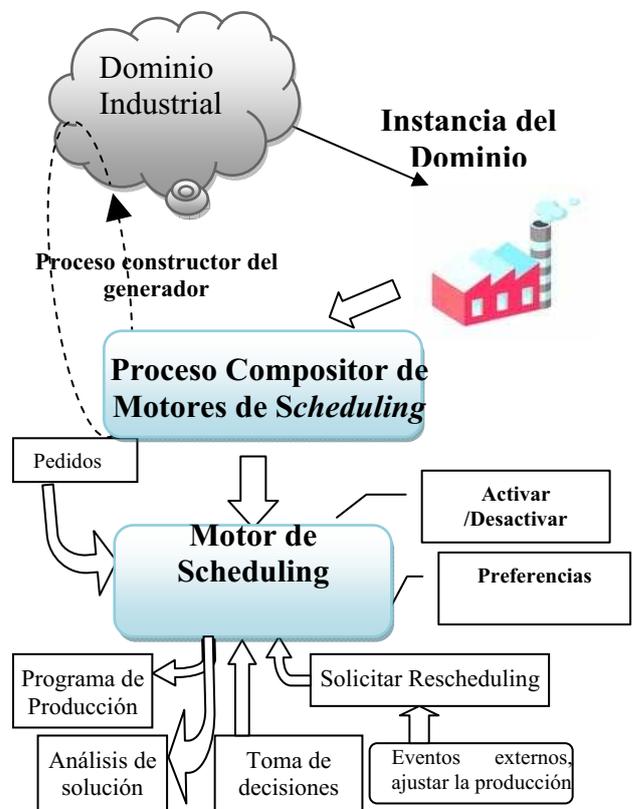


Figura 1 Problema de Investigación

La instancia queda descrita por un modelo conceptual, que describe a las máquinas, procesos, hojas de rutas, y demás objetos de la

instancia. Este modelo conceptual constituye la entrada al proceso compositor de motores de solución, el cual produce como salida, un motor de solución que debe responder a las distintas instancias del problema que tiene dicha compañía.

Una instancia del problema queda determinada por las distintas necesidades que ocurren en el día a día de la empresa. Por ejemplo, máquinas no disponibles, nuevos pedidos, urgencias, etc. Cada instancia es en realidad un problema de scheduling.

Los componentes básicos del problema son “el proceso compositor de motores de solución” y el “motor de solución”, los cuales se muestran en la Figura 1.

1.3 PROPUESTA DE SOLUCIÓN

En esta propuesta, se muestra que es posible construir un generador de motores de scheduling para dar solución al problema planteado, mediante programación por restricciones y SPL.

La programación por restricciones es una de las técnicas que permite solucionar problemas de scheduling (Barták., 2003) (Baptiste et al., 2001).

La ingeniería de SPL es un paradigma para desarrollar aplicaciones de software (sistemas y productos de software intensivos) usando una plataforma de componentes reusables y de customización en masa (Pohl et al., 2005).

Nuestra propuesta tiene como eje principal la reutilización, tanto de la construcción con reuso y para reuso. La ingeniería SPL utiliza el reuso como camino para construir sistemas, utilizando los conceptos de la ingeniería de dominios, pero acotando el dominio bajo estudio a las características del conjunto concreto de productos (existentes, planeados, o futuros) a desarrollar.

La ingeniería de línea de productos de software tiene dos procesos (Weiss and Lai, 1999, Boeckle et al., 2004, van der Linden, 2002).

La Ingeniería de Dominio: Es el proceso responsable de establecer la plataforma reusable y por lo tanto definir las similitudes y las variabilidades de la línea de producto.

La Ingeniería de Aplicación: Es el proceso responsable de derivar aplicaciones de la línea

de producto, a partir de la plataforma establecida en la ingeniería de dominio.

Regresando a nuestro problema, el Proceso Compositor de Motores de Scheduling es una línea de productos. Podemos ver entonces que este proceso se compone de dos subprocesos, ingeniería de dominio e ingeniería de aplicación. En el primero se establece una plataforma reusable que contiene todos los artefactos de software para construir un motor de scheduling a medida para una instalación industrial del dominio bajo estudio. Mientras que el segundo proceso, la ingeniería de aplicación, construye el motor de scheduling.

Nuestro proceso de construcción de motores de scheduling también adhiere al enfoque generativo de desarrollo de software (Czarnecki, 2005), permitiendo la generación automáticamente de una aplicación de la línea de producto, mediante la definición y uso de un DSL – *Domain Specific Language* o Lenguaje de Propósito Específico. (Greenfield and Short, 2004, Mernik et al., 2003, Fowler, 2005). En nuestro caso el modelo conceptual de la instancia del dominio se formaliza usando un DSL. Así, el proceso generador toma esta especificación y genera un motor de scheduling para la instancia del dominio. Una de las ventajas de usar DSL es que el motor puede ser construido por usuarios expertos del dominio, sin que sean expertos en tecnologías informáticas. Cuando el motor no se adecua a la realidad de la planta, se puede modificar usando el conocimiento del dominio, sin necesidad de programación. Sin embargo, para cada instancia (empresa del dominio) se tienen diferentes necesidades que no pueden ser capturadas al momento de definir la plataforma, y por ese motivo, la ingeniería de dominio resulta un proceso que está permanentemente modificando o agregando conocimiento a la plataforma reusable. Por último, otra particularidad de este enfoque es que el conocimiento obtenido en cada implementación se puede proveer a las demás instancias del dominio.

2 LINEAS DE I + D

Las líneas de I+D son las siguientes:

- Algoritmos de scheduling para soportar preferencias parametrizables
- Calidad en Scheduling
- Software product lines y software product families
- Enterprise reference models y Scheduling References models
- Variabilidad y similitudes en constraint programming
- Implementaciones en la industria del envase flexible

3 RESULTADOS OBTENIDOS/ESPERADOS

Como resultados obtenidos se pueden mencionar los siguientes: (1) Formulación del problema, (2) Estudio de estado del arte sobre problemas de scheduling, técnicas de solución de problemas y sistemas de scheduling, integración del sistema de scheduling, *software product lines* y *software product families*. (3) Vinculación con empresas que tienen la problemática abordada, tareas de campo. (4) Obtención de una beca para hacer una pasantía de 6 meses en el Institut National Polytechnique de Lorraine (INPL) ubicado en Nancy Francia con fecha de comienzo en septiembre de 2009.

Como tareas actuales, se está terminando el reporte de estado del arte en la temática abordada (examen de calificación programa de doctorado, capítulo de tesis).

Diseñando e Implementando un prototipo de generador de motores de scheduling para el dominio de la industria del envase flexible.

Como tareas Futuras. Generalizar el método para generador de motores de scheduling.

Publicar resultados, redactar y defender la tesis.

4 FORMACION DE RECURSOS HUMANOS

La Formación de Recursos Humanos está sustentada en el trabajo de tesis de doctorado de Daniel Diaz Araya becario de CONICET con beca Tipo II.

5 BIBLIOGRAFIA

BAPTISTE, P., PAPE, C. L. & NUIJTEN, W. (2001) *Constraint-Based Scheduling - Applying Constraint Programming to Scheduling Problems*, Springer.

BARKER, K. R. (1974) *Elements of sequencing and scheduling*, New York, John Wiley and Sons.

BARTÁK., R. (2003) Constraint-based Scheduling: An Introduction for Newcomers. *Intelligent Manufacturing Systems* 69-74.

BOECKLE, G., KNAUBER, P., POHL, K. & SCHMID, K. (2004) *Software-Produktlinien: Methoden, Einführung und Praxis (in German)*, Heidelberg, dpunkt.

BOUCHER, T. & YALCIN, A. (2006) *Design of Industrial Information Systems*, Academic Press.

CZARNECKI, K. (2005) Overview of Generative Software Development. *Unconventional Programming Paradigms*.

FOWLER, M. (2005) Language Workbenches: The Killer-App for Domain Specific Languages?

FOX, M. S. & SMITH, S. F. (1984) ISIS - a knowledge-based system for factory scheduling. *Expert Systems*, 1, 25-49.

GANTT, H. L. (1903) A graphical daily balance in manufacture. *Transactions of the American Society of Mechanical Engineers*, 24, 1322-1336.

GREENFIELD, J. & SHORT, K. (2004) *Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools*, Indianapolis, Wiley.

LE PAPE, C. (1994) Implementation of Resource Constraints in ILOG SCHEDULE: A Library for the Development of Constraint-Based Scheduling Systems. *Intelligent Systems Engineering* 3, 3, 55-66.

MACNIECE, E. H. (1951) *Production Forecasting, Planning, and Control*, New York, John Wiley & Sons.

MERNIK, M., HEERING, J. & SLOANE, A. M. (2003) When and how to develop domain-specific language. *Technical Report SEN-E0309* Amsterdam, CWI Available from <http://www.cwi.nl/ftp/CWIreports/SEN/SEN-E0309.pdf>.

ORLICKY, J. (1975) *MRP, The New Way of Life in Production and Inventory Management*, McGraw-Hill Book Company.

POHL, K., BÖCKLE, G. & LINDEN, F. J. V. D. (2005) *Software Product Line Engineering: Foundations, Principles and Techniques*, Springer.

RAJPATHAK, D., MOTTA, E., ZDRAHAL, Z. & ROY, R. I. (2003) A Generic Library of Problem Solving Methods for Scheduling Applications. *Second International Conference on Knowledge Capture (KCAP'2003)*. Florida, USA,.

SAUER, J. (2000a) Knowledge-based Systems in Scheduling. IN LEONDES, T. L. (Ed.) *Knowledge-Based Systems Techniques and Applications*. San Diego, USA, Academic Press.

SAUER, J. (2000b) Knowledge-Based Systems Techniques and Applications in Scheduling. IN

- LEONDES, C. T. (Ed.) *Knowledge-Based Systems Techniques and Applications* Academic Press.
- SHANKS, G. & SEDDON, P. (2000) Enterprise resource planning (ERP) systems. *Journal of Information Technology*, 15, 243-244.
- SMITH, F. S. (1994) *OPIS: A Methodology and architecture for reactive scheduling*, Morgan Kaufmann.
- VAN DER LINDEN, F. (2002) Software Product Families in Europe: The ESAPS and CAFÉ Projects. *IEEE Software*, 19, 41-49.
- VOLLMANN, T. E., BERRY, W. L. & WHYBARK, D. C. (1997) *Manufacturing Planning and Control Systems*, New York, Irwin/McGraw-Hill.
- WEISS, D. M. & LAI, C. T. R. (1999) *Software Product-Line Engineering - A Family-Based Software Development Process*, Massachusetts, Addison-Wesley.
- WIGHT, O. W. (1984) *Production and Inventory Management in the Computer Age*, Wight (Oliver) Publications Inc., U.S.
- YEN, B. P.-C. & PINEDO, M. (1994) On the design and development of scheduling systems. *Fourth International Conference on Computer Integrated Manufacturing and Automation Technology*.

Extensión de algoritmos ACO usando conceptos de *Tabu Search*

Franco Arito Guillermo Leguizamón

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional
Departamento de Informática
Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis
e-mail: {farito, legui}@unsl.edu.ar

Resumen

En los algoritmos ACO (del inglés Ant Colony Optimization) las hormigas artificiales iterativamente construyen soluciones a un problema de optimización combinatorio. La construcción de dichas soluciones está guiada por rastros de feromona los cuales actúan como un mecanismo de adaptación que permite sesgar la muestra de nuevas soluciones hacia regiones prometedoras del espacio de búsqueda. Adicionalmente, el sesgo de la búsqueda está influenciado por información heurística dependiente del problema.

En este trabajo, que describe parte de la línea de investigación relacionada al uso de memoria externa en algoritmos ACO, se propone la incorporación de conceptos de la metaheurística *Tabu Search* (TS) en el proceso de construcción de soluciones. Estos conceptos se refieren específicamente a la forma en la que TS utiliza la historia del proceso de búsqueda para evitar visitar soluciones ya analizadas.

Palabras Clave: Ant Colony Optimization, Tabu Search, Quadratic Assignment Problem.

1. Introducción

Los algoritmos ACO generan soluciones candidatas para un problema de optimización con un mecanismo de construcción donde la elección de la componente de solución a ser agregada en cada paso de la construcción está influenciada probabilísticamente por rastros de feromona e

información heurística [3]. En este trabajo, se examina la posibilidad de alternar la forma en la que se eligen las componentes de solución, introduciendo una memoria externa como mecanismo auxiliar para tomar las decisiones en cada paso de la construcción de una solución. Esta modificación está inspirada en la metaheurística *Tabu Search*¹, la cual usa explícitamente la historia de la búsqueda, para escapar de óptimos locales e implementar una estrategia explorativa. Este tipo de enfoque pertenece a una de las tendencias actuales en algoritmos ACO, en los cuales se incorpora una memoria externa como alternativa al mecanismo de elección de componentes de soluciones [1, 16, 17].

El algoritmo *Tabu Search* simple aplica búsqueda local con el criterio de “*best-improvement*” como componente básico y usa una memoria de corto plazo, para poder escapar de óptimos locales y evitar ciclos en la búsqueda. La memoria de corto plazo está implementada como una *lista tabú* que mantiene registro de las soluciones visitadas más recientemente y prohíbe movimientos hacia ellas. De esta forma, se restringe la vecindad de la solución actual a soluciones que no pertenecen a la

¹El significado de la palabra tabú designa a una conducta, actividad o costumbre prohibida por una sociedad, grupo humano o religión, es decir, es la prohibición de algo natural, de contenido religioso, económico, político, social o cultural por una razón de utilidad social. La búsqueda tabú no se refiere obviamente a ninguna de estas temáticas, sino al hecho de imponer restricciones para guiar un proceso de búsqueda con el objetivo de superar regiones del espacio de búsqueda. Estas restricciones operan de varias formas, como por ejemplo mediante la exclusión directa de alternativas de búsqueda clasificadas como “prohibidas”.

lista tabú, previniendo de ciclos infinitos y forzando a la búsqueda a aceptar movimientos que incluso pueden generar soluciones peores. Sin embargo, la implementación de la memoria a corto plazo como una lista que contiene soluciones completas puede no ser efectiva, por lo que en lugar de almacenar soluciones, se almacenan *atributos de soluciones* (e.g., componentes de soluciones, movimientos o diferencias entre dos soluciones).

La lista tabú, identificada usualmente con el uso de memoria a corto plazo, es solo una de las maneras posibles de aprovechar la historia de la búsqueda. La información recolectada durante todo el proceso de búsqueda también puede ser muy útil, especialmente para una guía estratégica del algoritmo. Este tipo de memoria a largo plazo generalmente se agrega a TS de acuerdo a cuatro principios: lo reciente, frecuencia, calidad e influencia. La memoria basada en lo reciente registra para cada solución (o atributo) la iteración más reciente en la que estuvo involucrada. En contraposición, la memoria basada en frecuencia mantiene registro de cuántas veces cada solución (o atributo) ha sido visitada (usado). Esta información identifica las regiones del espacio de soluciones donde la búsqueda estuvo concentrada o donde permaneció por mayor número de iteraciones. Es decir, es información que representa la memoria del proceso de búsqueda y que puede ser explotado para diversificar la búsqueda. Por su parte, el atributo calidad hace referencia a la habilidad para diferenciar la bondad de las soluciones visitadas a lo largo del proceso de búsqueda. De esta forma, la memoria puede ser utilizada para la identificación de elementos comunes a soluciones buenas o a ciertos caminos que conducen a ellas. La calidad constituye un fundamento para el aprendizaje basado en refuerzos, donde se premian las acciones que conducen a buenas soluciones y se penalizan aquellas que, por el contrario, conducen a soluciones pobres. Por último, la cuarta dimensión de memoria, referida a la influencia, considera el impacto de las decisiones tomadas durante la búsqueda, no solo en lo referente a la calidad de las soluciones, sino también en lo referente a la estructura de las mismas.

2. Uso de memoria externa en algoritmos ACO

El enfoque abordado agrega un mecanismo con características determinísticas en la construcción de soluciones, en contraposición con la filosofía presente en los algoritmos ACO, los cuales construyen soluciones en forma totalmente probabilística (basada en los rastros de feromona y en la información heurística). Las estructuras de memoria utilizada permiten que las hormigas en ciertas ocasiones² no tomen decisiones en forma aleatoria sino que elijan componentes de soluciones, de manera determinística, influenciadas por los valores registrados en dicha memoria. Como esta memoria almacena información específica de la historia de la búsqueda desde el comienzo del algoritmo, permite efectivamente enfocarse en regiones del espacio de búsqueda no visitadas (no registradas en la memoria y con valores que reflejen ésto) o por el contrario concentrarse en regiones ya visitadas y que sean prometedoras (también registradas en la memoria y con valores que lo reflejen). Estos usos de la memoria reflejan los mecanismos de intensificación y diversificación que, aunque ya estén presentes en los algoritmos ACO, son de utilidad para evitar convergencia prematura y lograr un mayor rendimiento frente a problemas de optimización con características particulares.

Se debe recalcar que aunque el mecanismo en estudio está inspirado en Tabu Search, existen ciertas diferencias en cuanto al uso de la memoria y también tiene sus diferencias respecto de otros algoritmos ACO que utilizan el enfoque de memoria explícita[1, 16, 17]. En este sentido, cabe destacar por un lado, que en TS las memorias basadas en lo reciente y en frecuencia se complementan la una a la otra para lograr el balance entre intensificación y diversificación. Así, la memoria basada en lo reciente es utilizada como memoria a corto plazo y aquella basada en frecuencia, como memoria a largo plazo (en el enfoque propuesto no existe esta distinción). Por otro lado, en los algoritmos ACO citados previamente, se propone almacenar en la memo-

²Para esto se utilizan parámetros que indican el nivel de exploración que tiene el algoritmo en determinados instantes de tiempo

ria partes de soluciones que pueden estar constituidas por las mejores soluciones encontradas durante la búsqueda (en el enfoque propuesto la memoria almacena atributos de soluciones recolectados durante el proceso de la búsqueda, los cuales representan indirectamente patrones interesantes de las soluciones). En lo que sigue se describen los principios considerados en nuestra propuesta.

2.1. Memoria basada en lo reciente

Este tipo de memoria almacena componentes de las soluciones que han cambiado en el pasado reciente. La forma habitual de explotar este tipo de memoria es etiquetando las componentes seleccionadas de soluciones visitadas recientemente. Lo que se intenta lograr es “prohibir” o “incentivar” ciertas elecciones, que impidan o favorezcan a que las hormigas construyan soluciones ya evaluadas en el pasado reciente y con lo cual se logra abarcar una mayor región del espacio de búsqueda o concentrarse en una región particular. Es importante tener en cuenta que en algunas instancias, un buen proceso de búsqueda resultará en volver a visitar una solución encontrada anteriormente. El objetivo más general es continuar estimulando el descubrimiento de nuevas soluciones de alta calidad.

2.2. Memoria basada en frecuencia

Este tipo de memoria almacena componentes de las soluciones que forman parte de una solución con mayor frecuencia, i.e., la mayor cantidad de veces presentes en una solución y en una posición particular de la solución. La forma habitual de explotar este tipo de memoria es etiquetando las componentes seleccionadas de soluciones más frecuentemente escogidas. Esta memoria, permite “prohibir” que una hormiga escoja una componente de solución porque es muy frecuentemente escogida en las soluciones. Se busca a través de la “prohibición”, generar soluciones que efectivamente se diferencien de las ya generadas, extendiendo de esta manera, la exploración del espacio de búsqueda. Inversamente, se puede usar esta información para “promover” su elección por considerarla atractiva ya que una gran mayoría de las hormigas la han elegido como parte de sus soluciones y por lo tanto, se considera deseable de que forme parte de una nueva solución.

3. Trabajos realizados

Como parte de la evaluación empírica de la propuesta, se ha implementado el enfoque de memoria explícita en el algoritmo $MAX - MIN$ Ant System [12] hibridado con aplicación de búsqueda local. Para la realización de los experimentos se ha considerado como caso de estudio, el problema de asignación cuadrática [5] (QAP, del inglés, Quadratic Assignment Problem). La formulación general de QAP describe las distancias entre n ubicaciones y los flujos entre n objetos. El objetivo es encontrar una asignación de objetos a ubicaciones, de manera tal de minimizar el producto de la distancia entre cada par de ubicaciones por el flujo entre el par de objetos asignados a ese par de ubicaciones. Distintos algoritmos ACO han sido aplicados a QAP [7, 9, 15, 4, 6]. Uno de los motivos de la elección de QAP como caso de estudio es que muchos tipos de instancias del mismo tienen características interesantes para estudiar el beneficio del enfoque de memoria explícita. En cuanto a la elección del algoritmo $MAX - MIN$ Ant System, se debe a que actualmente es uno de los mejores algoritmos ACO [3] para resolver QAP. Cabe destacar además, que dicho enfoque de memoria explícita puede ser fácilmente adaptado para ser aplicado a otros problema de optimización.

Las instancias utilizadas son las disponibles en QAPLIB³, una librería de instancias de benchmark para QAP. Los algoritmos ACO están actualmente entre los mejores algoritmos aproximados para resolver QAP, en particular para instancias estructuradas presentes en la vida real (ver [14] para una descripción detallada de los tipos de instancias). Sin embargo, en instancias no estructuradas (generadas aleatoriamente), su performance es inferior a métodos de mejora iterativa tales como búsqueda local iterada [10, 11], búsqueda tabú [13, 8], entre otros.

Las instancias de QAPLIB de tipo no estructuradas, generadas aleatoriamente tienen características que hacen que los algoritmos ACO inicialmente propuestos para este problema no sean competitivos. Una de estas características es que la *distancia* entre las soluciones que son óptimas

³<http://www.seas.upenn.edu/qaplib>

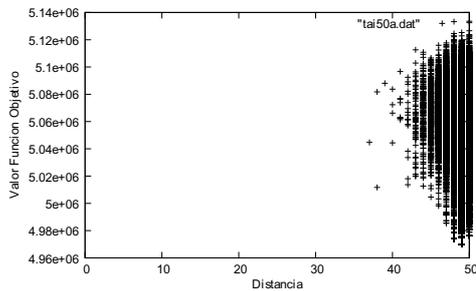


Figura 1: Instancia `tai50a` tomada de QAPLIB

locales es muy grande. Aquí asumimos que la distancia $d(\phi, \phi')$ entre dos soluciones ϕ y ϕ' para el QAP es:

$$d(\phi, \phi') = |\{i \in \{1, \dots, n\} \mid \phi_i \neq \phi'_i\}|$$

Si la distancia entre dos soluciones que son óptimos locales es grande, significa que esas soluciones tienen pocas componentes en común. La dificultad que este tipo de característica plantea a los algoritmos ACO, es que necesitan de un mecanismo de diversificación que le permita a las hormigas elegir (en ocasiones) componentes de soluciones que difieran de manera significativa de las componentes de otras soluciones ya construidas; y de esta manera adaptarse a las grandes distancias entre las soluciones que implica esta característica.

La Figura 1 muestra 10000 soluciones generadas por el algoritmo $\mathcal{MAX} - \mathcal{MLN}$ Ant System para la instancia `tai50a` de QAPLIB. Las soluciones con distancias grandes respecto del óptimo tienen, en general, valores objetivo bajos; y las soluciones con distancias cercanas al óptimo tienen, en general, valores objetivo más altos. Por lo tanto para hacer frente a este tipo de instancias, se incorpora el uso de la memoria como mecanismo de diversificación adicional.

Los resultados obtenidos hasta el momento han sido alentadores, tanto si se compara el algoritmo $\mathcal{MAX} - \mathcal{MLN}$ Ant System con y sin la memoria externa. Igualmente, cuando se compara nuestra propuesta con otros algoritmos ACO que utilizan otro tipo de memoria externa.

4. Conclusión y Trabajo futuro

Los trabajos futuros a corto plazo involucran un estudio detallado de la mejor configuración de los parámetros que el enfoque de memoria requiere; así como la implementación paralela del algoritmo $\mathcal{MAX} - \mathcal{MLN}$ Ant System para resolver el QAP, con instancias de mayor tamaño.

Una de las líneas a seguir es la aplicación de algoritmos ACO con el enfoque de memoria, aplicados a otros problemas de optimización combinatoria. Asimismo, se planea la aplicación del enfoque de memoria a otros algoritmos ACO eficientes, como el algoritmo Ant Colony System [2].

5. Recursos Humanos

La propuesta particular planteada aquí junto con los estudios derivados de la misma han dado lugar a la formación de becarios de grado y posgrado. En particular se pretende continuar en este y otros temas relacionados a través del contacto realizado por los autores con grupos de la Universidad de Málaga (España) y del INRIA, Lille (Francia).

Referencias

- [1] A. Acan. An external memory implementation in ant colony optimization. In M. Dorigo, M. Birattari, C. Blum, L. M. Gambardella, F. Mondada, and T. Stützle, editors, *ANTS 2004*, volume 3172 of *LNCS*, pages 73–84. Springer, Heidelberg, 2004.
- [2] M. Dorigo and L. M. Gambardella. Ant Colony System: A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1(1):53–66, 1997.
- [3] M. Dorigo and T. Stützle. *Ant Colony Optimization*. MIT Press, Cambridge, MA, 2004.
- [4] L. M. Gambardella, É. D. Taillard, and M. Dorigo. Ant colonies for the quadratic assignment problem. Technical report, IDSIA-4-97, IDSIA, Lugano, Switzerland, 1997.

- [5] T. C. Koopmans and M. J. Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25(1):53–76, 1957.
- [6] V. Maniezzo and A. Coloni. The ant system applied to the quadratic assignment problem. *IEEE Transactions on Knowledge and Data Engineering*, 1999.
- [7] V. Maniezzo, A. Coloni, and M. Dorigo. The ant system applied to the quadratic assignment problem. Technical report, IRIDIA/94-28, IRIDIA, Université Libre de Bruxelles, Belgium, 1994.
- [8] A. Misevicius. A tabu search algorithm for the quadratic assignment problem. *Computational Optimization and Applications*, 30(1):95–111, 2005.
- [9] T. Stützle. *MAX – MIN* for the quadratic assignment problem. Technical report, AIDA-97-4, FG Intellektik, FB Informatik, TU Darmstadt, Germany, 1997.
- [10] T. Stützle. Iterated local search for the quadratic assignment problem. Technical report, AIDA-99-03, FG Intellektik, FB Informatik, TU Darmstadt, 1999.
- [11] T. Stützle. Iterated local search for the quadratic assignment problem. *European Journal of Operational Research*, 174:1519–1539, 2006.
- [12] T. Stützle and H. Hoos. *MAX – MIN* ant system. *Future Generation Computer Systems*, 16(8):889–914, 2000.
- [13] É. D. Taillard. Robust taboo search for the quadratic assignment problem. *Parallel Computing*, 17:443–455, 1991.
- [14] É. D. Taillard. Comparison of iterative searches for the quadratic assignment problem. *Location Science*, 3(2):87–105, 1995.
- [15] É. D. Taillard and L. M. Gambardella. Adaptive memories for the quadratic assignment problem. Technical report, IDSIA-87-97, IDSIA, Lugano, Switzerland, 1997.
- [16] S. Tsutsui. *cAS*: Ant colony optimization with cunning ants. In T. P. Runarsson *et al.*, editor, *Proc. of the 9th Int. Conf. on Parallel Problem Solving from Nature (PPSN IX)*, volume 4193 of *LNCS*, pages 162–171. Springer, Heidelberg, 2006.
- [17] W. Wiesemann and T. Stützle. Iterated ants: An experimental study for the quadratic assignment problem. In M. Dorigo, L. M. Gambardella, M. Birattari, A. Martinoli, R. Poli, and T. Stützle, editors, *ANTS 2006*, volume 4150 of *LNCS*, pages 179–190. Springer, Heidelberg, 2006.

Modelos Formales de Diálogo para Sistemas Multi-Agente

Julieta Marcos **Marcelo A. Falappa** **Guillermo R. Simari**
mjm@cs.uns.edu.ar mfalappa@cs.uns.edu.ar grs@cs.uns.edu.ar

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA)
Departamento de Ciencias e Ingeniería de la Computación (DCIC)
Universidad Nacional del Sur (UNS)

Resumen

En un sistema multi-agente los agentes necesitan comunicarse, por diferentes motivos, dando lugar a diferentes tipos de *diálogos*. Existe interés por automatizar estas interacciones, sin embargo la mayoría de las propuestas existentes en la literatura son *ad hoc* y carecen de un fundamento teórico sólido. Nuestro trabajo de investigación está orientado principalmente al estudio abstracto de estas interacciones con el fin de brindar metodologías que faciliten el diseño de sistemas de diálogo.

Contexto

Esta línea de investigación tiene lugar dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial, y está asociada a los siguientes proyectos de investigación:

- “Programación en lógica distribuida con primitivas de sincronización, comunicación y ejecución en paralelo de agentes, para desarrollo de sistemas multi-agente en ambientes dinámicos” Código: PIP 5050. Director: Alejandro Javier García. Financiamiento: CONICET (Proyecto de Investigación Plurianual).
- “Formalismos y Tecnologías de Dinámica de Conocimiento aplicadas a Robótica Cognitiva, Sistemas Multi-agente y Web Semántica”. Código: PGI 24/ZN11. Director: Dr. Marcelo A. Falappa. Co-director: Dr. Alejandro J. García. Financiamiento: Universidad Nacional del sur.

1. Introducción

En un sistema multi-agente los agentes necesitan comunicarse, por diferentes motivos: resolver diferencias de opinión o intereses en conflicto, cooperar para resolver dilemas o encontrar pruebas, o simplemente informarse uno a otro sobre hechos pertinentes. En muchos casos no alcanza con intercambiar mensajes aislados, sino que los agentes necesitan entablar *diálogos* (secuencias de mensajes sobre el mismo tema) [10]. Existe una gran variedad de interacciones con características diferentes que podrían quererse modelar. Una posible tipología, teniendo

en cuenta el objetivo común del diálogo y las metas particulares de cada participante, es la siguiente [13]:

Diálogo de Búsqueda de Información. Un agente busca la respuesta a una pregunta en el conocimiento de otro agente. Se supone que este último conoce la respuesta.

Diálogo de Investigación. Todos los agentes colaboran para encontrar la respuesta a una pregunta. Se supone que ninguno de ellos conoce la respuesta.

Diálogo Persuasivo. Un agente trata de convencer a otro para que se adhiera a cierta creencia o punto de vista.

Negociación. Los agentes tratan de llegar a un acuerdo aceptable sobre la división de recursos escasos. Cada uno trata de maximizar su ganancia.

Diálogo Deliberativo. Los agentes colaboran para decidir que acción realizar en cierta situación.

Se puede hacer una distinción entre diálogos *colaborativos* y diálogos *no-colaborativos*. En un diálogo colaborativo los agentes no tienen metas individuales más allá de la meta común del diálogo; por lo tanto, todos colaboran en aras del mismo fin. En un diálogo no-colaborativo, en cambio, los agentes tienen metas individuales que podrían estar en conflicto. Por ejemplo, un diálogo de investigación es colaborativo (el único objetivo compartido por todos los agentes investigadores es descubrir la verdad) pero una negociación no lo es (cada agente negociador tiene como meta maximizar su propia ganancia). Un diálogo deliberativo podría ser colaborativo o no (dependiendo de si los agentes tienen algún interés particular por tomar cierto curso de acción). Un diálogo persuasivo puede ser visto como un diálogo *semi-colaborativo*, donde el agente que persuade tiene una meta particular, pero el persuadido podría no tenerla. Los *agentes colaborativos* expondrán toda la información que consideren relevante, mientras que los *agentes no-colaborativos* podrían ocultar información, sabiendo que es relevante, porque no favorece el cumplimiento de sus metas individuales.

Otra posible diferenciación es entre aquellos diálogos que son *sobre creencias*, y aquellos que no lo son. En un diálogo sobre creencias los participantes hablan sobre la verdad de cierta proposición. A esta categoría corresponden los primeros tres tipos de diálogo: búsqueda de información, investigación y diálogo persuasivo. Sin embargo, un diálogo persuasivo no es *necesariamente* sobre creencias (podría ser, por ejemplo, sobre acciones).

El objetivo final de este área de investigación es la automatización de los diferentes tipos de diálogo en un sistema multi-agente. A grandes rasgos, podemos identificar algunos elementos que están presentes en un sistema de diálogo:

1. Un **conjunto de agentes con estado mental**. Los elementos involucrados en el estado mental de los agentes dependerán de los distintos tipos de diálogo.
2. Un **tópico** en torno al cual se desarrolla el diálogo.
3. Una **secuencia de movimientos** realizados por los agentes, que conforman al diálogo propiamente dicho. En qué consisten exactamente los movimientos, y si éstos pueden causar modificaciones en el estado mental de los agentes o no, dependerá de los distintos tipos de diálogo.
4. Una **conclusión** sobre el tópico en cuestión, abordada al finalizar el diálogo.

Por ejemplo, el estado mental de un agente podría incluir simplemente creencias (en el caso de diálogos sobre creencias), o podría incluir también deseos, intenciones, preferencias, etc., en el caso de diálogos deliberativos o negociaciones. La forma de los movimientos también es variante según los distintos tipos de diálogo. Por ejemplo, en los diálogos de investigación los movimientos

en general consisten en hacer afirmaciones o aceptar afirmaciones hechas por otros, mientras que en una negociación tienen sentido también otros tipos de movimientos como por ejemplo: realizar/aceptar/rechazar propuestas y contra-propuestas, hacer promesas, etc.. En general, los sistemas de diálogo definen un conjunto de *locuciones* (por ejemplo: *afirmar*, *proponer*, *rechazar*, *aceptar*) y luego los movimientos especifican el tipo de locución junto con cierto contenido (el contenido corresponde a lo que se afirma, o lo que se promete, etc.). También suele especificarse un *protocolo de interacción*, que es un conjunto de reglas que gobiernan el diálogo, indicando principalmente qué tipos de movimientos son válidos en determinado momento (por ejemplo: después de que un agente hace una propuesta, otro agente puede aceptarla, rechazarla, o hacer una contra-propuesta). Si bien el protocolo de interacción establece en líneas generales la forma de los diálogos válidos, esto no es suficiente para lograr su automatización. Para tal fin, hace falta una *estrategia* que permita a cada agente determinar de manera autónoma el movimiento (locución + contenido) a realizar en cada momento.

Se han realizado varios trabajos con el objetivo de estudiar y modelar formalmente estas interacciones. Existe un grupo creciente de investigadores que consideran que se puede mejorar la calidad de los diálogos si los agentes exponen los *argumentos* que justifican lo que dicen [12], es decir, entablan *diálogos basados en argumentación*. Este tipo particular de diálogo respeta la formal general mencionada más arriba, pero se asumen en particular *agentes argumentativos* (es decir, agentes capaces de generar y evaluar argumentos) de manera que los movimientos que éstos realizan incluyen (o pueden incluir), como parte del contenido, argumentos y contra-argumentos. Habitualmente, los sistemas de diálogo basados en argumentación se diseñan sobre la base de un formalismo argumentativo ¹ existente. El trabajo realizado en [12] brinda un amplio panorama en lo que respecta a la *negociación basada en argumentación*, identificando y describiendo los principales elementos (tanto internos como externos a los agentes) necesarios para su modelamiento, y señalando también las principales motivaciones y ambiciones en la investigación.

La mayoría de las propuestas existentes en la literatura definen formalmente un sistema para un tipo particular de diálogo, e investigan propiedades de los diálogos generados por el sistema. Sin embargo, las soluciones propuestas en general son *ad hoc* y carecen de un fundamento teórico sólido. Algunos trabajos sólo especifican la representación del estado mental de los agentes, el conjunto de locuciones, y el protocolo de interacción, dejando sin especificar la estrategia de selección de movimientos. Otros trabajos incluyen también una estrategia, pero en general ésta se diseña de manera *ad hoc* y muchas veces no alcanza todas las propiedades deseables. A continuación mencionaremos brevemente algunas de las propuestas existentes. Todos los trabajos que mencionaremos estudian diálogos basados en argumentación. En [4] proponen un framework para diálogos basados en argumentación, con el objetivo de llegar a un acuerdo en la toma una decisión colectiva entre un conjunto de agentes autónomos, y estudian propiedades como *terminación* y *optimalidad del resultado*. En [2], [9] y [3] los autores se concentran en diálogos sobre creencias (búsqueda de información, investigación, y persuasión). Definen un conjunto de *locuciones* para que los agentes puedan intercambiar argumentos, un conjunto de *actitudes* que marcan una relación entre los argumentos que puede construir un agente y las locuciones que puede realizar (intuitivamente, los agentes “menos atrevidos” sólo afirman proposiciones soportadas por “buenos argumentos”), y definen también un conjunto de *protocolos* para llevar a cabo los diálogos. Luego consideran algunas propiedades de los diálogos generados bajo estos protocolos: en principio *terminación* y *complejidad*, y luego extienden el trabajo analizando propiedades del resultado e investigando en qué medida el resultado es dependiente de las tácticas utilizadas por los agentes.

¹Mayor información en el área de sistemas argumentativos puede encontrarse, por ejemplo, en [5] y [11].

2. Líneas de Investigación y Desarrollo

Dado que, desde nuestro punto de vista, hace falta una base teórica sólida que facilite el diseño y desarrollo de estos sistemas de diálogo, nuestro trabajo de investigación actual tiene por objetivo estudiar de manera abstracta y formal los requerimientos esenciales que deberían alcanzar estos sistemas, así como también identificar los principales desafíos en la persecución de los mismos, y esbozar estrategias generales para su cumplimiento. Actualmente estamos considerando una noción de diálogo restringida en ciertos aspectos, en particular: (1) el conjunto de posibles locuciones está reducido a *afirmar* conjuntos de proposiciones, (2) el lenguaje de representación de conocimiento y el mecanismo para obtener conclusiones son compartidos por todos los agentes participantes, y (3) el estado mental de los agentes no se modifica durante el diálogo. Se pretende brindar un enfoque abstracto, con el objetivo de abarcar un rango amplio de diálogos, ya sea basados en argumentación o no, sin especificar la representación del estado mental de los agentes ni el mecanismo para obtener conclusiones. De todas maneras, se pretende llevar, luego, los resultados generales a casos particulares para analizarlos en profundidad. En este sentido, estamos instanciando los conceptos y resultados obtenidos con formalismos lógicos específicos, en particular *Lógica Clásica Proposicional*, *Programación en Lógica* y *Programación en Lógica Rebatible* (DeLP) [6].

3. Resultados Obtenidos/Esperados

En [7] propusimos un algoritmo abstracto que simula una interacción entre dos o más participantes. Básicamente, vemos al diálogo como un proceso mediante el cual los agentes provocan sucesivos cambios sobre una base de conocimiento pública (que representa el *estado* del diálogo). Los agentes tienen *metas* que dictan qué conocimiento exponer en determinado momento. El uso de *Operadores de Cambio*² nos permitió abstraernos de la teoría lógica subyacente (lenguaje y mecanismo de inferencia). Por otro lado, el uso de una noción abstracta de *meta* nos permitió parametrizar las actitudes de los agentes en el diálogo, abstrayéndonos, en cierta medida, del *tipo* de diálogo. Mostramos cómo pueden diseñarse las metas de los agentes para modelar diferentes tipos de diálogos sobre creencias (*investigación*, *búsqueda de información* y *persuasión*) y brindamos ejemplos en Lógica Proposicional. En un trabajo posterior, también exploramos su aplicabilidad en diálogos de *negociación*.

En [8] nos concentramos en diálogos colaborativos específicamente, proponiendo un conjunto básico de *propiedades deseables* que éstos deberían satisfacer. Las propiedades propuestas persiguen la caracterización de diálogos finitos, con conclusiones razonables (considerando todo lo que ha sido dicho durante el diálogo), en los cuales toda la información relevante es expuesta y toda la información expuesta es realmente relevante. Para tal fin, definimos dos niveles de *relevancia* en el diálogo, *relevancia directa* y *relevancia potencial*, basados en las nociones de *inferencia* y *abducción* respectivamente. Mostramos ejemplos concretos utilizando Lógica Proposicional y Programación en Lógica Rebatible.

Como trabajo futuro, pretendemos generalizar los resultados a otros tipos de diálogos y relajar algunas de las restricciones impuestas en trabajos anteriores, en particular ampliar el conjunto de locuciones, y considerar conjuntos de agentes no homogéneos en cuanto al lenguaje de representación de conocimiento y al mecanismo de razonamiento.

²La *Teoría de Cambio de Creencias* estudia la *dinámica del conocimiento* en agentes o mundos, es decir, los cambios provocados en una base de conocimiento por el arribo de nueva información.

Referencias

- [1] Leila Amgoud, Yannis Dimopoulos, and Pavlos Moraitis. A unified and general framework for argumentation based negotiation. *In 6th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), Honolulu, Hawaii*, May 2007.
- [2] Leila Amgoud, Nicolas Maudet, and Simon Parsons. Modelling dialogues using argumentation. *In 4th International Conference on MultiAgent Systems, ICMAS'2000, Boston, USA. IEEE Press*, 2000. July.
- [3] Leila Amgoud, Simon Parsons, and Michael Wooldridge. On the formal outcomes of formal inter-agent dialogues. *In 2nd International joint conference on Autonomous Agents and Multi-Agent systems, AAMAS'03, Melbourne, Australia. ACM Press*, pages 616–623, July 2003.
- [4] Leila Amgoud, Henri Prade, and Sihem Belabbes. Towards a formal framework for the search of a consensus between autonomous agents. *In the proceedings of the 4th International joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'2005, Utrecht. Franck Dignum, Michael Wooldridge, Sven Koenig, Sarit Kraus (Eds.), ACM Press*, pages 537–543, July 2005.
- [5] Carlos Iván Chesñevar, Ana Gabriela Maguitman, and Ronald Prescott Loui. Logical models of argument. *ACM Comput. Surv.*, 32(4):337–383, 2000.
- [6] Alejandro J. García and Guillermo R. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4(1):95–138, 2004.
- [7] M. Julieta Marcos, Marcelo A. Falappa, and Guillermo R. Simari. Un modelo abstracto de diálogo sobre creencias para sistemas multi-agente. *Actas del XIII Congreso Argentino de Ciencias de la Computación (CACIC)*, pages 1371–1382, 2007.
- [8] M. Julieta Marcos, Marcelo A. Falappa, and Guillermo R. Simari. Properties for a formal model of collaborative dialogue. *Actas del XIV Congreso Argentino de Ciencias de la Computación (CACIC)*, 2008.
- [9] Simon Parsons, Leila Amgoud, and Michael Wooldridge. An analysis of formal inter-agent dialogues. *In 1st International Joint Conference on Autonomous Agents and Multi-Agent systems, AAMAS'2002, Bologna, Italy. ACM Press*, pages 394–401, July 2002.
- [10] Simon Parsons, Michael Wooldridge, and Leila Amgoud. Properties and complexity of some formal inter-agent dialogues. *Journal of Logic and Computation*, 13:347–376, 2003.
- [11] Henry Prakken and Gerard Vreeswijk. Logics for defeasible argumentation. In Dov M. Gabbay, editor, *Handbook of Philosophical Logic*. Kluwer Academic, 1999.
- [12] Iyad Rahwan, Sarvapali D. Ramchurn, Nicholas R. Jennings, Peter McBurney, Simon Parsons, and Liz Sonenberg. Argumentation-based negotiation. *The Knowledge Engineering Review*, 18:343–375, 2003.
- [13] Douglas Walton and Erik C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, NY, 1995.

Dinámica de bases de conocimiento compartidas y distribuidas

Luciano H. Tamargo †
Alejandro J. García †

Marcelo A. Falappa †
Guillermo R. Simari

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur,
† Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Av. Alem 1253, (B8000CPB) Bahía Blanca, Argentina
Tel: (0291) 459-5135 / Fax: (0291) 459-5136
e-mail: {lt, maf, ajg, grs}@cs.uns.edu.ar

Contexto

Esta línea de investigación se realizará dentro del ámbito del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial, y está asociada a los siguientes proyectos de investigación:

“Programación en lógica distribuida con primitivas de sincronización, comunicación y ejecución en paralelo de agentes, para desarrollo de sistemas multi-agente en ambientes dinámicos” Código: PIP 5050. Director: Alejandro Javier García. Financiamiento: CONICET (Proyecto de Investigación Plurianual).

“Representación de conocimiento, argumentación y apoyo a la toma de decisiones: Herramientas inteligentes para la web semántica” (código 24/N023). Inicio 01/01/2007 Finalización: 31/12/2009. Acreditado en el Programa de Incentivos. Director: Simari, Guillermo Ricardo.

Resumen

El objetivo principal de esta línea de investigación es el estudio y desarrollo de técnicas y formalismos para actualización de bases de conocimiento de agentes inteligentes en sistemas multi-agente. Dichas bases de conocimiento podrán ser compartidas por varios agentes y estar distribuidas dentro del sistema, pretendiendo además que el conocimiento representado en los agentes pueda tener referencias temporales.

En particular, se estudiará combinar formalismos de revisión de creencias y actualización de conocimiento, con técnicas de mantenimiento de bases de datos distribuidas en sistemas multi-agente. De esta manera, se permitirá actualizar en forma dinámica bases de conocimiento de diferentes agentes que participan de un sistema multi-agente dentro de un ambiente distribuido. Además, se estudiarán diferentes modelos donde los agentes del sistema estén ordenados por fiabilidad y se definirá un operador de revisión que cambie dinámicamente el orden mencionado. De esta manera, el formalismo propuesto en esta línea de investigación obtendrá más flexibilidad y dinamismo.

Palabras clave: Revisión de Creencias, Sistemas Multi-Agente, Bases Distribuidas

1. Introducción

Los agentes son entidades computacionales autónomas, ya sea programas o robots, con la capacidad de percibir el entorno en que se desenvuelven y actuar para llevar a cabo alguna tarea. La percepción del entorno por parte de un agente es usualmente limitada, y diferentes agentes dentro de un sistema multi-agente pueden tener una percepción diferente del entorno. Por lo tanto, es importante que los agentes puedan cooperar compartiendo su conocimiento acerca del entorno. Además, si cada agente almacena conocimiento producto de su experiencia, es importante que pueda compartir este conocimiento con otros agentes con los cuales coopera. De esta manera, agentes especialistas en cierto aspecto podrán intercambiar conocimiento con otros que tienen experiencia en otras áreas.

Cada agente inteligente necesita de una base de conocimiento donde almacenar la información referida al entorno en el cual se desenvuelve. El conocimiento aquí almacenado debe ser actualizado de manera periódica y respetar ciertas restricciones de integridad para garantizar el correcto razonamiento del agente. En un entorno multi-agente, el conocimiento puede estar distribuido, fragmentado, y/o replicado.

En esta línea de investigación se pretende combinar formalismos de revisión de creencias y actualización de conocimiento con técnicas de mantenimiento de bases de datos distribuidas. De esta manera, se permitirá actualizar en forma dinámica bases de conocimiento de diferentes agentes en un ambiente distribuido. El estudio y desarrollo de agentes inteligentes, y la investigación en robótica cognitiva han demandado en los últimos años la investigación y desarrollo de formalismos de representación y mantenimiento de conocimiento. En particular esto ha dado un fuerte impulso al área de revisión de creencias (belief revision) [1, 10, 9, 16, 17, 22, 11, 12, 13, 14, 21, 15, 7] donde se estudia formalmente parte de estos temas. Además, en los últimos años se han producido importantes avances en el área de revisión de creencias en sistemas multi-agente, entre cuyos artículos cabe destacar [6, 5, 4, 19, 20, 18].

2. Líneas de Investigación y Desarrollo

Esta línea de investigación busca combinar resultados teóricos obtenidos en las áreas de sistemas multi-agentes, mecanismos de representación de conocimiento y revisión de creencias. El objetivo es actualizar en forma dinámica bases de conocimiento de diferentes agentes en un ambiente distribuido. Disponer de un formalismo como el propuesto permitirá lograr un avance en las áreas antes mencionadas, así como también el desarrollo de nuevas aplicaciones que utilicen la tecnología de sistemas multi-agente. Esto facilitará el desarrollo de aplicaciones basadas en agentes para resolver problemas complejos. Los resultados obtenidos serán utilizados para su aplicación inicialmente en el campo de agentes inteligentes y sistemas multi-agente y posteriormente en robótica cognitiva, sistemas de información así como también sistemas de recomendación.

2.1. Revisión de creencias

La meta principal de la teoría del cambio es intentar modelar la dinámica del conocimiento. No hay una única manera de realizar un “cambio” en el conocimiento de un agente, existen numerosas propuestas para ello. El modelo AGM [1] es uno de los principales referentes (expansión, contracción y revisión). La revisión es una de las operaciones más comunes que desarrolla un agente que se desenvuelve en un entorno dinámico. Si un agente cree en un conjunto K , y se produce cierta entrada epistémica X , el mismo deberá hacer lo siguiente: revisar el contenido de K para determinar cuáles de sus creencias están en desacuerdo con X , eliminar algunas de ellas, luego incorporar X y producir un nuevo conjunto K' consistente. En este contexto pueden surgir diferentes situaciones:

- a) La entrada epistémica tiene primacía sobre el estado epistémico del agente. En tal caso deben eliminarse del estado epistémico todas las creencias que sean inconsistentes con la entrada. Algunos ejemplos de este tipo de operadores son Partial meet revision [1] y Kernel revision [12].
- b) La entrada no tiene primacía y el agente, según corresponda, pueda preservar su estado epistémico (idealmente en equilibrio), pues puede ser el caso en que no le de “mayor credibilidad” a la entrada epistémica. Por ejemplo, Semi-Revision [13], Revision by a Set of Sentences [7] y Selective Revision [8].

2.2. Revisión de creencias en sistemas multi-agentes

Una variedad de nociones han sido adoptadas por las personas que realizan investigaciones en el área de Revisión de Creencias en Sistemas Multi-Agentes (MABR). Un buen entendimiento de las relaciones entre estas aproximaciones es esencial antes de llevar a cabo cualquier investigación futura. Es por esto, que como trabajo a futuro en esta línea, en principio se planea profundizar sobre el estudio de formalismos específicos de MABR. Por ejemplo, en el trabajo de Wei Liu [19] se presenta un análisis exhaustivo y una jerarquía sobre los distintos sistemas existentes (ver Figura 1) entre los que podemos destacar: Multi-Agent Belief Revision (MABR) y Belief Revision using information from Multiple Sources (MSBR).

La capacidad de realizar revisión de creencias podría ser considerado como parte de las habilidades de los agentes para mantener consistencia de su propio estado epistémico. En este caso, un proceso de revisión de creencias individual es llevado a cabo en una ambiente multi-agente, donde la nueva información puede venir desde múltiples fuentes y podría generar conflictos. Revisiones de ese estilo se denominan MSBR. Sobre esta alternativa, algunos autores tratan de resolver el conflicto de información ordenando las fuentes sobre las bases de su fiabilidad [5, 4]. Otros autores, usan técnicas de revisión de creencias para lograr metas de grupos o sociedades de agentes [18, 20] (por ejemplo, alcanzar consenso antes de llevar a cabo los planes). En este escenario más de un agente toma parte en el proceso y, para este propósito, los agentes involucrados necesitan comunicarse, cooperar, coordinar y negociar con los otros. Un sistema MABR es un sistema multi-agente (MAS) cuya meta global involucra utilizar técnicas de revisión de creencias.

MSBR estudia el comportamiento de las revisiones en agentes individuales, es decir, cuando un agente recibe información desde múltiples agentes desde quienes tiene opiniones sociales. MABR investiga el comportamiento de la revisión de creencias global de grupos o sociedades de agentes. MSBR es uno de los componentes esenciales de MABR.

El paradigma AGM [1] ha sido ampliamente aceptado como un framework estándar para revisión de creencias. Pero sólo es capaz de describir comportamientos de revisión de un único agente. El proceso de revisión de creencias es más complejo en el caso de múltiples agentes. Más allá del principio de mínimo cambio, existen otros requisitos debido a la sofisticada interacción entre los agentes.

Un agente es capaz de llevar a cabo Individual Belief Revision (IBR), mientras que una sociedad o equipo de agentes es capaz de MABR. IBR en un ambiente de un único agente (Single Belief Revision, SBR) podría ser logrado usando belief revision tradicional satisfaciendo o adaptando los postulados AGM. IBR en un ambiente de múltiples agentes es MSBR, es decir, un único agente tendrá que procesar información que viene desde más de una fuente.

3. Resultados Obtenidos/Esperados

Anteriormente en esta línea se ha desarrollado un formalismo de revisión de creencias en sistemas multi-agente el cual está orientado a Multi-Source Belief Revision [23]. Este formalismo incluye un nuevo operador de revisión basado en un ordenamiento de fuentes (como es sugerido en [6, 5, 4, 3])

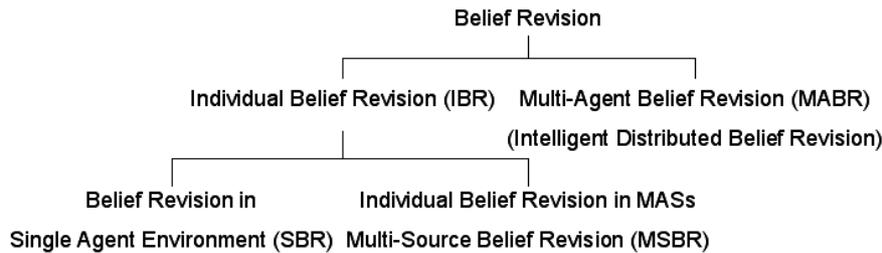


Figura 1: Jerarquía de Revisión de Creencias

y en la técnica de contracción propuesta por Sven Ove Hanson en [12]. Esta aproximación tiene similitudes con las propuestas en lógica posibilística [2, 3]. Actualmente, esta línea de investigación está orientada en el análisis de los operadores de revisión basados en lógica posibilística.

En primera instancia, el objetivo de esta línea de investigación seguirá orientado a Multi-Source Belief Revision con la idea de continuar con el desarrollo del formalismo propuesto. En base a este nuevo formalismo de revisión de creencias en sistemas multi-agente, se definirá una caracterización axiomática del operador proveyendo de esta manera teoremas de representación que le dará a la propuesta un sólido marco teórico. Además, se estudiarán diferentes modelos donde los agentes de un sistema multi-agente estén ordenados por fiabilidad y se definirá un operador de revisión que cambie dinámicamente el orden mencionado. De esta manera, el formalismo propuesto obtendrá más flexibilidad y dinamismo.

La importancia de esta línea de investigación radica en el estudio y desarrollo de nuevos formalismos para su aplicación en el área de sistemas multi-agente. Estos sistemas se utilizan en la actualidad para la resolución de problemas complejos en ciencias de la computación y en otras disciplinas. Su campo de aplicación ha ido en aumento en los últimos años, fundamentalmente por los desarrollos teóricos y prácticos producidos en las áreas de inteligencia artificial distribuida y robótica cognitiva. Se prevee que el campo de aplicación de estos sistemas seguirá en aumento, lo cual requerirá el desarrollo de nuevos formalismos teóricos para su futura aplicación.

El desarrollo de técnicas de actualización de bases de conocimiento distribuidas, se utilizará para mejorar el mantenimiento y la consulta de las bases de conocimiento de los agentes. Esto permitirá contar con nueva tecnología para el desarrollo de aplicaciones con sistemas multi-agente, que podrán ser utilizados en diferentes áreas, tanto en las ciencias de la computación como en otras ciencias.

Referencias

- [1] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *J. of Symbolic Logic*, 50(2):510–530, 1985.
- [2] S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *In Proc. of Uncertainty in Artificial Intelligence*, pages 411–419. Morgan Kaufmann, 1993.
- [3] S. Benferhat, D. Dubois, H. Prade, and M. A. Williams. A practical approach to revising prioritized knowledge bases. *Studia Logica*, 70(1):105–130, 2002.
- [4] J. Cantwell. Resolving conflicting information. *Journal of Logic, Language and Information*, 7(2):191–220, 1998.

- [5] A. Dragoni, P. Giorgini, and M. Baffetti. Distributed belief revision vs. belief revision in a multi-agent environment: First results of a simulation experiment. In *MAAMAW*, pages 45–62, 1997.
- [6] A. Dragoni, P. Giorgini, and P. Puliti. Distributed belief revision versus distributed truth maintenance, 1994.
- [7] M. A. Falappa, G. Kern-Isberner, and G. R. Simari. Explanations, belief revision and defeasible reasoning. *Artificial Intelligence*, 141(1):1–28, 2002.
- [8] E. L. Fermé and S. O. Hansson. Selective revision. *Studia Logica*, 63(3):331–342, 1999.
- [9] P. Gärdenfors. *Knowledge in Flux - Modeling the Dynamic of Epistemic States*. The MIT Press, Cambridge, Mass, 1988.
- [10] P. Gärdenfors and D. Makinson. Revisions of knowledge systems using epistemic entrenchment. In *Second Conference on Theoretical Aspects of Reasoning about Knowledge Conference*, pages 83–95. Morgan Kaufmann, 1988.
- [11] S. O. Hansson. In defense of base contraction. *Synthese*, 91(3):239–245, june 1992.
- [12] S. O. Hansson. Kernel contraction. *Journal of Symbolic Logic*, 59(3):845–859, 1994.
- [13] S. O. Hansson. Semi-revision. *J. of Applied Non-Classical Logic*, pages 151–175, 1997.
- [14] S. O. Hansson. *A Textbook of Belief Dynamics: Theory Change and Database Updating*. Kluwer Academic Publishers, march 1999.
- [15] Sven Ove Hansson, Eduardo L. Fermé, John Cantwell, and Marcelo A. Falappa. Credibility limited revision. *J. Symb. Log.*, 66(4):1581–1596, 2001.
- [16] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In *KR*, pages 387–394, 1991.
- [17] Hirofumi Katsuno and Alberto O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligent*, 52(3):263–294, 1992.
- [18] N. E. Kfir-Dahav and M. Tennenholz. Multi-agent belief revision. In Yoav Shoham, editor, *Theoretical Aspects of Rationality and Knowledge: Pro. of the Sixth Conf. (TARK 1996)*, pages 175–196. Morgan Kaufmann, San Francisco, 1996.
- [19] W. Liu and M. A. Williams. A framework for multi-agent belief revision, part i: The role of ontology. In *Australian Joint Conference on A. I.*, pages 168–179, 1999.
- [20] W. Liu and M. A. Williams. A framework for multi-agent belief revision. *Studia Logica*, 67(2):291–312, 2001.
- [21] David Makinson. Screened revision. *Theoria: Special Issue on Non-Prioritized Belief Revision*, 1997.
- [22] H. Rott. Preferential belief change using generalized epistemic entrenchment. *Journal of Logic, Language and Information*, 1(1):45–78, 1992.
- [23] L. H. Tamargo, A. J. García, M. A. Falappa, and G. R. Simari. Consistency maintenance of plausible belief bases based on agents credibility. *12th International Workshop on Non-Monotonic Reasoning (NMR)*, pages 50–58, 2008.

PROPUESTA PARA OPTIMIZACIÓN DE SISTEMAS PRODUCTIVOS MODELADOS CON SIMULACIÓN POR EVENTOS DISCRETOS

Ricardo R. Palma, Raymundo Forradellas

CEAL - Facultad de Ingeniería - Universidad Nacional de Cuyo
Centro Universitario – M5502JMA Mendoza
{rpalma, kike}@uncu.edu.ar

CONTEXTO

El presente trabajo está encuadrado en una línea de investigación que el CEAL (Centro de Estudio y Aplicaciones Logísticas) está desarrollando desde 2002 entorno a la aplicación de las TIC's a los dominios de la manufactura y su cadena de abastecimiento. Bajo de denominación de Supply Chain Management e Internal Logistics se han desarrollado en los últimos años una serie de trabajos científicos y aplicaciones de los mismos a los sistemas de manufacturas a efecto de optimizar su funcionamiento y dotarlos de flexibilidad para afrontar los frecuentes e intensos cambios que el entorno les impone.

RESUMEN

Este trabajo propone una instancia de optimización que ayuda al gerente de operaciones a configurar una nueva red de su sistema productivo agregando o quitando eslabones dentro de su planta industrial. Se complementan con las herramientas de optimización existentes para sistemas modelados y simulados por eventos discretos. De la revisión realizada hemos encontrado la necesidad de incorporar en la herramienta de optimización la posibilidad de modificar la red de producción.

Palabras clave: *Simulation, Supply Chain Management, Discret event simulation, Internal logistics.*

1. INTRODUCCION

1.1 Génesis de los SED

El sendero de la simulación ha estado siempre vinculado a la complejidad de los sistemas de manufactura. Desde un capítulo de la investigación operativa (Operational Research) surgió lo que se denomina Simulación por Eventos Discretos (SED).

En los últimos años han aparecido una serie de herramientas vinculadas a la simulación que han tenido un rápido crecimiento en su adopción en los medios industriales para mejorar la calidad del proceso de toma de decisión. Si bien algunos los SED son

conocidos desde los años 90, la gran mayoría de esta primera generación de herramientas estaba basada en el uso de leguajes específicos para su implementación. Ejemplos clásicos de estos leguajes son SIMAN y sus derivados. Esta primera generación de útiles para modelado de sistemas de manufactura incluía en su semántica elementos como las colas (filas) centros de manufactura y recursos necesarios para realizar los procesos. Todos estos lenguajes apelaban a 2 instancias denominadas *entry point* y *exit point* que permitían modelar una generalidad de procesos de fabricación utilizando descripción estadística de la tasa de arribos de objetos al sistema y un arsenal de distribuciones de probabilidades para describir cuanto tiempo demandaba el paso de estos objetos por los centro de manufactura.

1.2 Segunda generación SED

Los trabajos de W.D. Kelton [1] de 1999 dieron nacimiento a la segunda generación de herramientas para simulación que ampliaron el horizonte de utilización a un gran número de usuarios finales utilizando un lenguaje gráfico para el modelado. Si bien el trabajo de base se seguía siendo delegado al lenguaje como SIMAN, la estrategia de “dibujar” permitió ganar mucho en el terreno de la simplicidad del desarrollo de modelo. Por otro lado la posibilidad de documentar gráficamente lo que realmente ocurría en el piso de manufactura fue un elemento que aportó claridad a la comunicación dentro del entorno vinculado a la toma de decisiones industrial. Como resultado de esta segunda ola en la evolución de las herramientas aparecieron productos comerciales como Simul8 [1] PROMODEL [2] y ARENA [3] que tienen en nuestros días una gran difusión y uso a nivel del ámbito de programación de operaciones en empresas de manufactura.

1.3 Estado actual en la optimización

El crecimiento en las tecnologías de comunicaciones ha permitido que las herramientas de SED tomen contacto con le mundo real. Hoy casi todos lo productos comerciales que se señalaran en el párrafo

anterior son capaces de tomar contacto con las bases de datos o con los denominados sistemas ERP (*Enterprise Resource Planning*) y basados en el modelo creado de antemano más la información que de ellos se dispone puedan predecir inconvenientes como cuellos de botella transitorios o retrasos que se producirán al alterar el plan de producción. Las herramientas para optimización que han desarrollado en las soluciones comerciales en parte imitan el comportamiento del gerente de producción. Algunas de ellas como *OptQuest Solver* y *OptQuest Callable Library* [4] interrogan al programador de la producción sobre cuales son las variables sobre las cuales él tiene control y cuales son fijas o externas utilizando así métodos como el de simulación comparables a “Montecarlo” para darle una idea general de en que rango debería moverse todas las variables que se pueden controlar a fin de obtener un resultado tan próximo al óptimo como sea posible. Si bien los resultados de estos métodos de optimización han sido apreciados por quienes enfrentan a diario los complejos escenarios del mundo de la manufactura, es preciso reconocer que estas instancias de optimización están restringidas a iniciativa humana. Es decir que parten del hecho impuesto por el gerente de operaciones que designa las variables controlables y su rango de operación o variación.

En el presente trabajo se están explorando herramientas que aporten soluciones a este problema mediante cambios en la estructura misma de los proceso de manufactura. La técnica esta inspirada en los trabajos de líneas de investigación como es la Automatización de Reconocimiento de planes de fabricación [1].

El objetivo final es operar con un agente inteligente que sea capaz de interpretar que está tratando de hacer el gerente de operaciones al correr las simulaciones en diferentes oportunidades y que pueda sugerirle, respetando las restricciones tecnológicas el agregado o la eliminación de centros de manufactura, el incremento o disminución de recurso de producción y el incremento o descenso de el *tack time* o ritmo de producción y tiempo de ciclo.

1.4 Condiciones iniciales para el diseño del agente inteligente

Uno de los requisitos esenciales de para el que agente inteligente pueda inferir que pretende cambiar el gerentes de operaciones, es contar con una biblioteca de modelos de

sistemas de manufactura que le permitan hallar variantes que aporten resultados de performance superior al sistema de manufactura en operación.

Para ello es necesario desarrollar una nueva forma para describir el proceso de manufactura. En parte esta misión está resuelta por la metodología de la representación gráfica que las herramientas de simulación comerciales mencionadas han aportado.

A fin de poder establecer un marco que posibilite el intercambio de datos y modelos H.K. Concannon y K.I Hunter [6] propusieron en 2003 el desarrollo de una técnica que descripción de modelos de manufactura basada en XML. Esta propuesta ha sido adoptada por Simul8 y como parte de las actividades del CEAL se han desarrollado herramientas que permitan generar los archivos en formato XML a partir de ficheros de los modelos generados por las herramientas mencionadas como, ARENA y PROMODEL. Además se están ensayando métodos de generación del archivo XML a partir de modelos generados en formato VSD (Microsoft Visio) y de otras herramientas GNU como DIA.

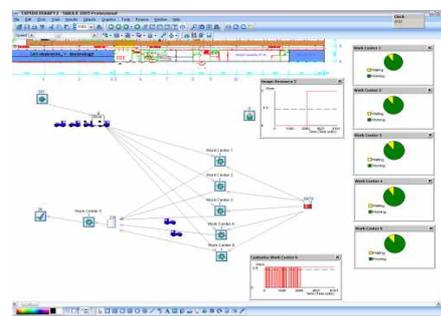


Fig.1: Ejemplo de Modelo de Manufactura S8

1.5 Estructura general en XML para el modelado de sistema de manufactura

A los efectos de comprender la estructura básica con la que se logra describir un sistema real de manufactura mediante marcadores de texto se expondrá un ejemplo de sistema extremadamente simple.

El sistema de manufactura en cuestión se representa en la Figura 2. En el mismo podemos identificar a derecha un punto de entrada (*entry point*) en el que las órdenes de proceso arriban cada cierto intervalo de tiempo establecido por una ley de distribución de probabilidades. Conforme estas órdenes entran en la simulación se encolan frente al proceso de manufactura en una fila (*queue 1*) que tiene cierta capacidad de almacenamiento antes de bloquear su

entrada y en la que las órdenes de proceso pueden estar solo un tiempo antes de considerarse retrasada o vencida.

De allí los productos son admitidos en el centro de manufactura que es capaz de procesar órdenes a razón de cierto tiempo distribuido según otra ley de probabilidades.

Para poder operar el centro de manufactura emplea un recurso (operador 10) de modo que la manufactura no comienza hasta que la orden de producción está disponible en la cola y el recurso operador es librado por el otro proceso o la orden de proceso anterior.

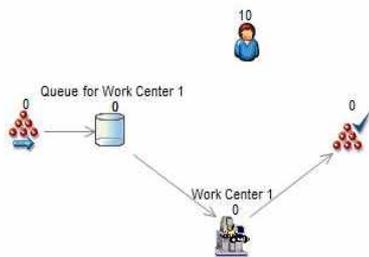


Fig 2 Ejemplo simple de Sistema de Manufactura

Este ejemplo se puede describir mediante un archivo en formato XML como:

```
<SED_SIM.XML>
<Simulation parameters>
Parametros de simulación , semillas random .etc
</Simulation parameters>
<Work entry point>
  <interarrival time>
  < bernoulli 20,12>
  </interarrival time>
  < fixed cost>
  <unit cost>
  <stetup cost>
</Work entry point>
  <Queue for Work Center 1>
  <capacity> 50 units</capacity>
  <self_time> 12 min </self_time>
  <arrivals> from entry point>
</Queue for Work Center 1>
<Work Center 1>
<setup time> 7 min </setup time>
<work time per item>
  <batch tipe>
<nromal , 3 min , sigma 0,5 min>
  </barch>
<Resource 10>
  <capacity 2 units>
  <schedule : odbc//200.55.21.122>
  <initial cost>
  <per hour cost>
  <waiting cost>
</Resource 10>
```

</Work Center 1>

</SED_SIM.XML>

Además es posible ver aspectos vinculados a costos operativos, parámetros estadísticos de los tiempos de espera y operación e incluso mecanismos de acceso a los turnos y horarios de disponibilidad del recurso humano accedido desde terceras partes, mediante una consulta QSL [5].

2. LINEAS DE I+D

2.1 Optimización por cambio de configuración

Basados en la propuesta de descriptor XML de sistema de manufactura se han montado 3 laboratorios gemelos en las carreras de Ingeniería Industrial de las Uiversidades:

UNLitoral, UNSan Luis y UNCuyo que han relevado con esta técnica uno 120 sistemas de manufactura existentes en el interior del país. De esta forma, es posible extrapolar instancias de optimización tomadas en terrenos de una línea de fraccionamiento de vinos y llevarlas a un sistema de ambulancias [7y5]. La modificación que un operador humano encontró como mejor alternativa para la línea de fraccionamiento fue aumentar el tiempo de ciclo en el llenado de botellas e incorporar una segunda línea de llenado (repetir líneas amarillas del XML) con el mismo número de personas que la operaban. Este mecanismo de optimización puede imitarse en un sistema de manufactura a condición de que su estructura tenga un grado de semejanza importante con el que ya ha obtenido la mejora. Es finalmente un parámetro de performance como la relación costo beneficio o el número probable de unidades procesadas al año lo que lleva al operador del simulador a indicar que la solución obtenida es mejor que la inicial. El agente inteligente detecta este hecho y lo guarda en su base de datos para utilizarlo como sugerencia a otro operador humano que está consultando por alternativas para optimizar su sistema[7].

2.2 Caracterización de sistemas de manufactura y factores de semejanza

Dado que la colección de modelos de sistemas de manufactura está validada por intervención humana, es esperable entonces que la biblioteca generada pueda servir para comprender muchos aspectos vinculados a la dinámica de evolución y tipificación de los sistemas de manufactura de la región. Pensando en esto se ha tratado de encontrar patrones comunes y características que permitan a los docentes de administración de

operaciones encontrar “denominadores comunes”[8]. Basados en esta estrategia se han podido establecer por el método de mínimos componentes un conjunto de 4 parámetros que logra explicar el 85% de las diferencias entre los archivos XML que representan al total de los sistemas relevados a la fecha. En rigor estos 4 parámetros constituyen una base ortogonal que representa un espacio multidimensional en el que es posible caracterizar a los sistemas de manufactura. Los 4 parámetros son: a) El número de proveedores (*enrty points*), b) La altura del árbol a la que se encuentra un nodo concentrador de procesos, c) La cantidad de instancias de almacenamiento interno (colas de productos semielaborado), d) La forma en que arriban los productos o materia prima el sistema. Es decir si es en forma estacional siguiendo ciclos anuales, o mensualmente, o por días u horas semejante a lo expuesto en la referencia [6].

Basándose en este hecho se han construido 4 modelos de simulación o configuraciones típicas con las que es posible explicar un gran número de casos que representan a empresas de la región.

3. RESULTADOS OBTENIDOS/ESPERADOS

Según lo presentado en el punto 2.2 se muestran las 4 configuraciones que aparecen

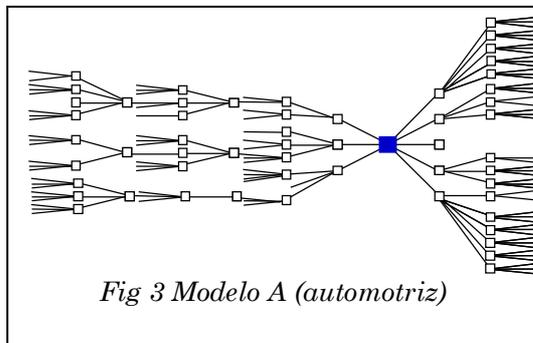


Fig 3 Modelo A (automotriz)

como casos extremos o más representativas de los 4 parámetros hallados para explicar las

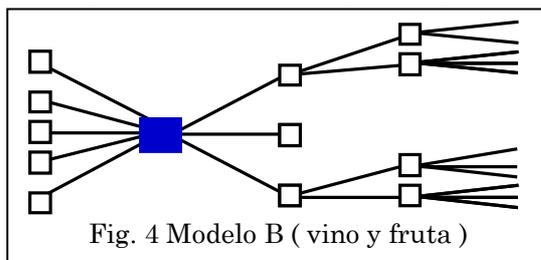


Fig. 4 Modelo B (vino y fruta)

diferencias de los sistemas de manufactura. El primer modelo identificado está relacionado por su proximidad a los que se encuentran de la industria del montaje de

automotores. El segundo tiene 99% de proximidad con los que se encuentran en la

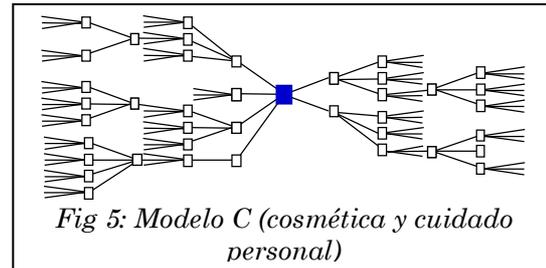


Fig 5: Modelo C (cosmética y cuidado personal)

industria del mueble y otros semejantes. El tercero tiene gran parecido con el que se encuentra en la elaboración del vino y frutas

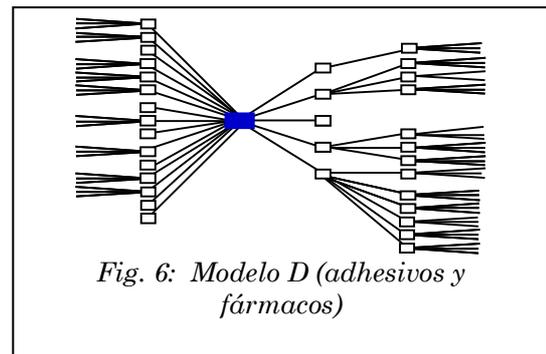


Fig. 6: Modelo D (adhesivos y fármacos)

frecas.

Y finalmente un modelo de manufactura que tiene como instancia más próxima a la de las actividades de producción de cosméticos, productos de perfumería artesanal y productos de tocador.

4. FORMACION DE RECURSOS HUMANOS

A la fecha la técnica de modelado de sistemas de manufactura ha sido impartida a recursos humanos que se encuentran en su etapa de redacción de tesis de maestría y doctorado. Se han desarrollado 2 talleres. Tesistas han adoptado el método como marco teórico o método de comprobación de hipótesis. Este trabajo está enmarcado en el trabajo de tesis doctoral del autor bajo dirección de coautor

5. BIBLIOGRAFIA

- 1- Designing simulation experiments Kelton, W.D. Dept. of Quantitative Anal. & Oper. Manage., Cincinnati Univ., OH; ISBN: 0-7803-5780-9
- 2- Decision Making Tool Techniques <http://www.promodel.com/>
- 3- Simul 8 software www.simul8.com
- 4- The OptQuest Callable Library M. Laguna, R. Martí
University of Colorado at Boulder and Universitat de Valencia

5- Generalized plan recognition - H Kautz, JF Allen - Proceedings of the Fifth National Conference on Artificial, 1986 - aaai.org

6- SIMUL8-Planner simulation-based planning and scheduling Concannon, K.H. Hunter, K.I. Tremble, J.M. Visual8 Corp., Mississauga, Ont., Canada; Volume: 2, On page(s): 1488- 1493 vol.2 ISBN: 0-7803-8131-9 INSPEC Accession Number: 8045456 DOI: 10.1109/WSC.2003.1261593

7- Simulation data exchange (SDX) implementation and use Source -Winter Simulation Conference archive Proceedings of the 33rd conference on Winter simulation table of contents Arlington, Virginia SESSION: Future of simulation table of contents Pages: 1473 - 1477 Year of Publication: 2001

ISBN:0-7803-7309-X Authors Dave Sly UGS e-Factory, Ames, IA Shreekanth Moorthy UGS e-Factory, Southfield, MI

8- Panel on future challenges in modeling methodology Taylor, S.J.E. Lendermann, P. Paul, R.J. Reichenthal, S.W. Strassburger, S. Turner, S.J. Dept. of Inf. Syst. & Comput., Brunel Univ., Uxbridge, UK;

Publication Date: 5-8 Dec. 2004 Volume: 1, On page(s): - 335 ISSN: ISBN: 0-7803-8786-4

Digital Object Identifier: 10.1109/WSC.2004.1371333

Current Version Published: 2005-01-03

6. AGRADECIMIENTOS

El autor desea manifestar el agradecimiento a las autoridades del Instituto Politécnico de Lorenia (INPL Francia), las autoridades de la Facultad de Ingeniería de la Universidad Nacional de Cuyo y al programa ARFITEC por la posibilidad de avance en el desarrollo de su tesis doctoral.

Una plataforma de Servicios Web

Ana Alonso de Armiño. aalonso@uncoma.edu.ar
Universidad Nacional del Comahue. Tel 0299-4490300

Pablo Rubén Fillotrani. prf@uns.edu.ar
Universidad Nacional del Sur. Tel 0291-4595135. Fax 0291 -4595135

Este trabajo de investigación se realiza en el marco del proyecto de investigación Software para Aprendizaje y Trabajo Colaborativos (UNComa) y del proyecto Un framework para gobernabilidad electrónica mediante servicios de la web semántica (UNS).

Resumen: La combinación de dos tecnologías, los Servicios Web y la Web Semántica, para posibilitar la integración de las aplicaciones, tanto dentro como entre las empresas, han influenciado en el diseño y desarrollo de estándares Web que afrontan diversos aspectos, así como el de plataformas de agentes que implementan servicios web. La composición de servicios es un tema vigente actualmente en el ámbito de investigación y es a lo que nos abocamos en este trabajo. Presentamos una propuesta de implementación de servicios usando una plataforma de agentes que describiremos a continuación para experimentar la composición de servicios.

Palabras clave. Servicios Web, Web Semántica, Composición de Servicios, Plataformas de Agentes, JADE

1. Introducción

La evolución de la Web viene acompañada con el desarrollo de nuevos estándares y tecnologías, dentro de los cuales podemos destacar a la **Web Semántica** y a los **Servicios Web**.

La Web Semántica pretende conseguir una comunicación efectiva entre computadoras, para lo cual centra sus esfuerzos en la búsqueda de descripciones enriquecidas semánticamente para los datos en la Web. Para lograrlo se requieren descripciones que incluyan no sólo las estructuras de datos, sino también las relaciones existentes con otros conceptos, las restricciones, reglas que permitan realizar inferencias, etc. En este contexto también se promueve la definición y reutilización de vocabularios u ontologías de conceptos que faciliten el procesamiento por parte de las máquinas.

Los Servicios Web promueven la interacción entre aplicaciones. El servicio es un componente software que puede procesar documentos XML recibidos a través de protocolos de transporte y de aplicación.

Ambas tecnologías constituyen lo que se conoce como la Web Semántica de Servicios, la cual proveerá otras funciones y requerirá otras técnicas y modelos que afectarán todos los ámbitos: educativos, e-commerce, colaboración y comunicación, etc.

El paradigma de los servicios web surge como un intento de mejorar las prácticas de desarrollo. En este enfoque el software deja de ser considerado como un producto manufacturado y pasa a ser visto como un servicio. Los servicios Web pretenden fortalecer las características de interoperabilidad, confiabilidad, escalabilidad y disponibilidad. Para alcanzar estos objetivos debe ser posible el descubrimiento, invocación y composición de servicios, y aquí es donde la semántica entra en juego.

En la sección 2 y 3 de este trabajo presentamos las nuevas tecnologías que surgen como consecuencia de la evolución de la Web. En la sección 4 presentaremos una propuesta de implementación de servicios y los resultados esperados.

2. Tecnologías de la Web Semántica

La Web ha tenido un crecimiento exponencial que nos coloca ante una enorme cantidad de datos. Allí anidan problemas bien conocidos como la dificultad para realizar consultas semánticas, elevados costos de mantenimiento de los sitios Web, contenidos dependientes del lenguaje, alta dependencia del factor humano para depurar la información buscada.

Es así que surge la necesidad de construir una Web Semántica (WS) [β] concebida como “una extensión de la actual (Web), en la cual la información está dada por significados bien definidos, mejorando la relación entre las computadoras y los humanos para su trabajo cooperativo” [1], que debe permitir que sea usada para un “descubrimiento más efectivo, automatización, integración y reutilización entre varias aplicaciones” [2].

Los mecanismos de representación del conocimiento, las ontologías y los agentes inteligentes, son las tecnologías requeridas en la construcción de esta nueva Web. El consorcio de la W3C, realiza la tarea de estandarización de

estas tecnologías e impulsa proyectos que tienden a masificar su uso.

La visión propuesta, una Web con semántica, plantea el reto de garantizar que las máquinas “comprendan”, interactúen en un medio certificado y bajo estrictos niveles de confianza.

2.1. Lenguajes de ontologías basados en Web.

El boom de Internet ha generado la proliferación de lenguajes para explotar las características de la Web. La figura 1, presenta la relación que se establece, entre los lenguajes de marcado, orientados a mejorar la semántica de la Web.

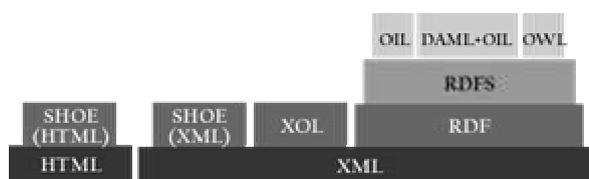


Fig. 1. Lenguajes de marcado de ontologías

Aquí se puede observar que todos ellos están basados en la capacidad sintáctica de XML.

SHOE (Simple HTML Ontology Extensión), es un lenguaje de representación del conocimiento diseñado para la WEB, basado en asociar contenido de páginas Web ligándolas con una o varias ontologías. Se trata de un anotador de contenido asociado a páginas Web, su modelo de conocimiento son los marcos.

Basado en la sintaxis de XML se encuentran XOL, lenguaje para intercambio de Ontologías basado en XML, fue desarrollado tomando como base OML y Ontolingua, por parte de investigadores del área bioinformática en los Estados Unidos.

2.2. Tecnologías de los Servicios Web

Los Servicios Web constituyen un sistema de software diseñado para soportar la interacción de las máquinas sobre una red. Que tiene una interfase descrita en un formato procesable por las máquinas (WSDL). Otros sistemas interactúan con el servicio Web en una forma prescrita por esa descripción usando mensajes SOAP, típicamente usando HTTP y XML entre otros estándares.

El modelo de procesos de negocio (business process) prescribe el orden en el cual deben ejecutarse las actividades que lo constituyen y las condiciones dentro de las cuales estas pueden ser realizadas. Esto se conoce como flujo de control. Modelar un proceso de negocio requiere un sistema que verifique las propiedades y condiciones requeridas por el mismo para que el proceso sea ejecutado exactamente como lo

requiere el proceso. Un application service provider (ASP), permite correr partes del proceso del negocio y ejecutar el software de monitoreo y de backups periódicamente.

La tecnología de Servicios Web facilita y extiende este modelo.

Para posibilitar el descubrimiento dinámico y la comunicación con los servicios, se requiere un bajo nivel de acoplamiento entre quien requiere y el servicio: no hace falta conocer la plataforma sobre la que cada uno se ejecuta, no hace falta conocer su implementación y patrones usados por cada uno, no hace falta conocer los formatos y protocolos usados para interoperar.

La tecnología de Servicios Web [4] es construida sobre el concepto de mensajería. Como consecuencia el solicitante y el servicio pueden correr sobre diferentes plataformas con canales conectándolos y haciendo las adaptaciones y transformaciones necesarias.

2.3. Lenguajes para la especificación de la composición de los servicios Web

• BPEL4WS

Éste es un estándar para especificar procesos del negocio y los protocolos de interacción del negocio. BPEL define un estado y la lógica de coordinación entre esas interacciones y formas semánticas de tratar las condiciones excepcionales. El modelo de procesos definido por BPEL se basa en el modelo de descripción de servicios de WSDL [5].

• WSFL (Web Services Flow Language)

WSFL es un lenguaje XML para la descripción de la composición de los servicios Web. WSFL considera dos tipos de composiciones de servicios Web: el primer tipo (modelo de flujos) especifica un patrón de uso de una colección de servicios Web, de forma tal que la composición resultante describa cómo se alcanza una meta en particular. El segundo tipo (modelo global) especifica un patrón de interacción de una colección de servicios Web, en este caso el resultado es una descripción del patrón de interacción total [7].

• WSCL (Web Services Conversation Language)

Provee una forma para modelar los procesos públicos de un servicio, habilitando a los servicios para participar en ricas interacciones. WSCL ha sido desarrollado como un complemento de WSDL. Este último especifica como enviarle mensajes a un servicio sin establecer el orden en el cual se pueden enviar esos mensajes, WSCL define la secuencia de documentos intercambiados entre los servicios Web. [6]

El objetivo principal de WSCL es definir una conversación.

- **BPML**

Su objetivo es expresar procesos de negocios abstractos y ejecutables. Permite definir procesos entre empresas, servicios Web complejos y colaboración entre múltiples partes. Un proceso en BPML es una composición de actividades que realizan funciones específicas. Los procesos dirigen la ejecución de estas actividades. A su vez puede ser parte de otra composición definiéndose como parte de un proceso padre o invocándolo desde otro proceso.

Las especificaciones en BPML soportan la importación de definiciones en WSDL [6].

3. Propuesta y Resultados Esperados

La plataforma que se pretende implementar consiste en un servicio de reservación de pasajes para trasladarse dentro de las provincias de Río Negro y Neuquén, llamamos al servicio RePaCo el cual es implementado combinando los servicios correspondientes a las empresas que realizan viajes dentro de las provincias antes mencionadas y un servicio de búsqueda de una ruta (cubierta por una o mas empresas) que permita ir desde un lugar a otro.

El servicio RePaCo debe consultar el servicio de las empresas de colectivo, el cual puede brindar información y efectuar reservas, y al servicio de búsqueda para combinar los viajes que sean necesarios para satisfacer el requerimiento del usuario.

A continuación describimos brevemente los actores que intervienen:

El *servicio de Transporte* acepta los requerimientos para proveer información sobre un viaje (origen-destino) dado y, si el trayecto es cubierto por una empresa provee información sobre el costo, horario de salida desde origen y llegada al destino, el tipo de coche y el tipo de servicio que brinda. Este servicio también acepta requerimientos para reservar un pasaje, en este caso retorna la lista de asientos disponibles en el colectivo elegido. La oferta puede ser aceptada (con la elección de un asiento) o rechazada por el servicio externo que ha invocado al servicio de Transporte.

El *servicio de Búsqueda* será invocado cuando el servicio de Transporte de una respuesta negativa, lo que significa que ninguna empresa tiene viajes directos desde origen a destino. Este servicio se encarga de verificar si existe alguna combinación de colectivos (que pueden pertenecer a diferentes

empresas) para ir de una ciudad a otra. Este servicio se debe comunicar con el servicio de Transporte para confeccionar un grafo de los viajes provistos por las empresas y buscar en él un camino (el más corto).

El *Usuario* envía un requerimiento para obtener un pasaje para ir desde una ciudad a otra y espera una respuesta negativa si esto no es posible, o una oferta indicando el precio y demás datos del servicio. El usuario puede aceptar o rechazar la oferta, en tal caso se confirma la reserva.

Con esta implementación se pretende experimentar la composición automática de los servicios desarrollados para brindar el servicio de reservación.

La implementación será realizada usando la plataforma de agentes JADE [8]. Cada uno de los servicios descritos será implementado por agentes que estarán disponibles a través de esta plataforma. La interfaz con el usuario será implementada como servlet accesibles desde una página web y es desde donde se hará la invocación al servicio RePaCo que usando composición automática podrá implementar su funcionalidad.

4. Bibliografía

[1] Tim Berners-Lee, James Hendler and Ora Lassila, "The Semantic Web", Scientific American, May 2001

[2] Ian Horrocks And Peter F. Patel-Schneider. "Three Theses of Representation in the Semantic Web". Conference WWW 2003, mayo 20 -24 2003, Budapest.Hungary.p39-47

[3] Ivan Herman, W3C. "Introduction to the Semantic Web".

[4] Benjamín González C. "Introducción a los servicios Web". <http://www.desarrolloweb.com/>

[5] Specifying and constraining W.S. behaviour through policies. <http://www.w3.org/2004/08/ws-cc/hp-20040908>

[6] Web Services Conversation Language (WSCL) 1.0. <http://www.w3.org/TR/wscl10/>

[7] Web Services Flow Language (WSFL). <http://xml.coverpages.org/wsfl.html>

[8] JADE (Java Agent DEvelopment Framework) <http://jade.cselt.it/>

TÉCNICAS DE MINERÍA DE DATOS Y METAHEURÍSTICAS APLICADAS A LA EXPLOTACIÓN EFICIENTE DE ENERGÍA EÓLICA

de San Pedro M.^{*}, Pandolfi D.^{*}, Lasso M.^{*}, Villagra A.^{*}, Lorenzetti D.[#], Fernandez C.[#], Valdez J.^{*}, Varas V.^{*}, Vidal P.^{*}, Bilbao M.^{*}

(^{*}) Laboratorio de Tecnologías Emergentes (LabTEem)

([#]) Laboratorio de Energías Renovables

Unidad Académica Caleta Olivia - Universidad Nacional de La Patagonia Austral
e-mail: {edesanpedro, dpandolfi, mlasso, avillagra, dloren, cfernandez, jcvvaldez, vvaras, pjvidal}@uaco.unpa.edu.ar; martb82@hotmail.com

Leguizamón, G.

Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)

Departamento de Informática - Universidad Nacional de San Luis

e-mail: legui@unsl.edu.ar

CONTEXTO

Esta línea de Investigación forma parte del proyecto “*Minería de Datos y Técnicas Metaheurísticas: su aplicabilidad al problema de explotación eficiente de energía eólica a través de la obtención de modelos aproximados y distribución óptima de máquinas eólicas*” del Laboratorio de Tecnologías Emergente (LabTEem) en el marco del programa de Investigación en Ciencia y Tecnología de la Universidad Nacional de la Patagonia Austral.

RESUMEN

La presente propuesta involucra tres temáticas principales: Metaheurísticas, Minería de Datos y Energía eólica (problemas asociados a la explotación eficiente de un parque eólico). En el caso de las metaheurísticas, el grupo tiene una amplia experiencia demostrada a través de la sostenida producción científica desde sus inicios en el año 1998, en publicaciones nacionales ([Villagra et al, 2005a] [Villagra et al, 2005b] [Villagra et al, 2004b] [Pandolfi et al, 2003a] [de San Pedro et al, 2003a] [Lasso et al, 2003a] [Lasso et al, 2003b] [de San Pedro et al, 2002a] [Lasso et al, 2002a] [Pandolfi et al, 2002c]) e internacionales ([de San Pedro et al, 2005] [Villagra et al, 2004a] [Lasso et al, 2004] [de San Pedro et al, 2004] [Lasso et al, 2003c] [de San Pedro et al, 2003c] [Pandolfi et al, 2003b] [Pandolfi et al, 2002a] [Pandolfi et al, 2002d]) en eventos de reconocida jerarquía, como también varios artículos en revistas [Pandolfi et al, 2004] [Pandolfi et al, 2002b].

Asimismo, el campo de Minería de Datos fue incorporado al grupo como objeto de estudio y sus posibles aplicaciones en el último proyecto presentado y actualmente en ejecución cuyo principal objetivo es el uso de metaheurísticas en el campo de minería de datos para mejorar el desempeño de las técnicas de minería de datos o como técnicas de minería de datos en sí.

Finalmente, sobre la problemática de la instalación de parques eólicos, el grupo ha incorporado dos nuevos integrantes quienes a su vez son integrantes de un grupo de investigación especializado en energías renovables. A partir de una fluida interacción con los mismos, los objetivos propuestos podrán ser alcanzados.

Palabras clave: *Minería de datos, Energía Eólica, Modelos descriptivos y predictivos, Metaheurísticas, Problemas de Asignación*

1. INTRODUCCION

En la actualidad, la producción de energía basada en recursos no-renovables, como el petróleo, ha alcanzado importantes niveles de desarrollo, sin embargo, su continuidad está muy limitada por diversos factores que tienen que ver, por ejemplo, con el incremento de los costos para su explotación, efectos nocivos en el medio-ambiente y principalmente, por su agotamiento en un corto plazo.

En consecuencia, diversas fuentes alternativas para la producción de energía, vienen siendo consideradas. Una de ellas, la producción de energía eólica. En este caso particular, es importante poder realizar una explotación eficiente de su principal recurso (el viento) a los efectos de lograr niveles adecuados en cuanto a la relación del costo de la inversión para realizar la explotación y el beneficio obtenido en cuanto a la capacidad de energía producida. En este sentido, un gran número de factores deberán ser considerados al momento de diseñar una estrategia que genere importantes beneficios a un costo razonable. Por ejemplo, factores como la intensidad y regularidad del viento, las características topográficas de la región en donde se instalará el parque, tipo y cantidad de máquinas eólicas (aerogeneradores), y su respectiva distribución, entre muchos otros factores.

Desde un punto de vista global y vinculado con la presente propuesta, podemos diferenciar varias etapas en el desarrollo de un plan de explotación eólica:

a) Recolección de datos sobre las características del viento (mediciones a través de veletas y anemómetros). Esta etapa también implica la definición de una estrategia adecuada para la selección de la ubicación y tipo de los equipos de medición a utilizar en función de al menos, los siguientes aspectos: velocidad media y máxima del viento, distribución de frecuencias en las diferentes direcciones, variación del viento con respecto a la altura y posición, estadística de ráfagas, etc. [Escudero 2003].

b) Inferencia de posibles modelos del comportamiento del viento según los datos recolectados. En este caso, el estudio se puede complementar con aproximaciones basadas en la distribución de Weibull [Weibull 1951] la cual puede ser usada para caracterizar los datos asociados a la velocidad del viento.

c) Determinación del número y distribución óptima de las máquinas eólicas en lo que se denomina conformación del parque eólico o “*wind farm*”. Más precisamente, un parque eólico es una agrupación de máquinas eólicas o aerogeneradores, usadas para producir energía eléctrica y las cuales se distribuyen sobre una región generalmente extensa.

d) Uso de los modelos encontrados para realizar predicciones sobre el comportamiento del viento y por lo tanto, lograr un uso más eficientemente del equipamiento instalado.

Los puntos a), b) y d) se enmarcan, dentro de nuestra propuesta, en el contexto del proceso conocido como KDD (*Knowledge Discovering in Databases*). Si bien el término KDD fue utilizado originalmente sobre datos almacenados en bases de datos, hoy en día, se lo considera un proceso que permite la extracción automática de conocimiento a partir de volúmenes de datos estadísticamente significativos (no necesariamente almacenados en bases de datos tradicionales o estructurados).

La minería de datos constituye el núcleo del análisis inteligente de los datos y ha recibido un gran impulso en los últimos tiempos motivado por distintas causas: i) el desarrollo de algoritmos eficientes y robustos para el procesamiento de grandes volúmenes de datos, ii) un poder computacional más barato que permite utilizar métodos computacionalmente intensivos, y iii) las ventajas comerciales y científicas que han brindado este tipo de técnicas en las más diversas áreas. Entre las áreas donde han sido utilizadas exitosamente las técnicas de minería de datos podemos mencionar distintas aplicaciones financieras y bancarias, análisis de mercado, seguros y salud privada,

educación, procesos industriales, medicina, biología, bioingeniería, telecomunicaciones, Internet, turismo, deportes, etc.

Es importante diferenciar en la minería de datos, el tipo de tareas que se suelen abordar y las técnicas utilizadas en cada caso. Como ejemplos de tareas generales, se pueden mencionar el aprendizaje de conceptos, clasificación, categorización, regresión, agrupamiento (o *clustering*), correlaciones y análisis de asociación. Estas tareas pueden ser abordadas mediante distintos métodos o técnicas que suelen adaptarse mejor de acuerdo a la tarea sobre la cual se trabajará. Entre las técnicas más conocidas se puede mencionar el aprendizaje de reglas de clasificación, reglas de asociación, reglas relacionales, reglas difusas, árboles de decisión (y regresión), ecuaciones de regresión, redes neuronales, metaheurísticas, etc. [Witten 1999], [Witten 2005], [Orallo et al. 2003]. En términos generales, el análisis inteligente de datos surge como una necesidad para dar respuesta a las limitaciones que exhiben hoy en día las técnicas tradicionales de análisis e interpretación manual de los datos.

Actualmente, llevar a cabo un análisis manual de los datos se torna impráctico si se consideran los grandes volúmenes de información originados desde distintas fuentes: registro automático de transacciones comerciales, actividades desarrolladas en Internet, disponibilidad *on-line* de resultados experimentales y mediciones científicas (velocidad y dirección del viento en distintas horas del día), etc.

En el contexto de construcción y mantenimiento de plantas productoras de energía, es indudable que a medida que los costos de los equipos para captura de diversos datos disminuyen y por ende se expande su utilización, la cantidad de datos recolectados se incrementa proporcionalmente. En la actualidad existe un término (*Energy Data Mining*), que identifica a todos aquellos procesos de búsqueda automática en grandes volúmenes de datos, derivados de observaciones vinculadas a plantas productoras de energía, que intentan descubrir patrones ocultos, correlaciones entre diferentes atributos y modos operacionales sostenidos. Por lo tanto, tareas típicas de minería de datos, tal como *clustering*, regresión, clasificación, reglas de asociación, etc., pueden ser aplicadas sobre esos datos que ayuden a la toma de decisiones para un adecuado diseño y/o funcionamiento de una planta productora de energía. En el caso de la energía eólica y desde la perspectiva del proyecto propuesto, estamos ante al siguiente situación: por un lado, es importante tratar de ajustar los modelos del comportamiento del viento en ciertas regiones de la Patagonia, a través del análisis de datos que permitan, describir o explicar el comportamiento de los datos; y por otro lado, poder realizar ciertas predicciones para mejorar el aprovechamiento de los recursos disponibles cuando la planta se encuentre

en proceso de producción. (<http://www.osdpd.noaa.gov/ml/index.html>).

Sin embargo, existen otras decisiones importantes antes de poner en funcionamiento un parque eólico de manera tal que pueda aprovechar al máximo las condiciones climáticas (comportamiento del viento en la zona elegida). En este caso, es de vital importancia la distribución de las máquinas eólicas, ya que este tipo de decisiones no son tan simples (es decir, no se trata solamente de distribuir aerogeneradores en una región ventosa). Cabe aclarar, que hay otro aspecto de fundamental importancia (fuera del alcance de nuestra propuesta) vinculado al tipo y diseño de los aerogeneradores a ser usados en la instalación.

Un proyecto bien encarado para la explotación eólica debe cuidadosamente definir la locación en donde se ubicarán las máquinas eólicas considerando, entre otras cosas, dirección y velocidad del viento en conjunción con otros aspectos específicos como por ejemplo, tener en cuenta que la velocidad del viento se incrementa con la altitud en regiones abiertas y que los sitios más adecuados son aquellos en la cima de cerros lisos y redondeados, en praderas y líneas costeras abiertas o entre montañas cercanas. En este último caso, por la posibilidad de la ocurrencia del efecto "embudo".

Claramente, el problema de la distribución de las máquinas eólicas puede ser representado como un problema de optimización en donde la función objetivo podría representar la energía eléctrica producida (un problema de maximización de beneficios). Siendo éste además, un problema de alta complejidad, puede existir más de un objetivo a optimizar e inclusive varias restricciones (lineales y no-lineales) asociadas al problema. En este sentido, diversas propuestas han sido consideradas para resolver el problema de distribución de las máquinas eólicas según las características de la función objetivo y la existencia de restricciones [Donovan 2005], [Donovan 2007], [Donovan 2008], [Eliknton 2008].

Sin embargo, hay situaciones en que los métodos tradicionales están limitados en su aplicación, y por ende los enfoques metaheurísticos se presentan como una alternativa altamente viable, no sólo por su robustez y flexibilidad, sino por la calidad de los resultados encontrados en tiempos razonables [Mora et al. 2006], [Mora et al. 2007a], [Mora et al. 2007b], [Grady et al. 2005]. Más precisamente, las metaheurísticas proveen de un marco general que permite crear nuevos híbridos a través de la combinación de conceptos derivados de: heurísticas clásicas, inteligencia artificial, evolución biológica, sistemas naturales, mecánica estadística, etc. Esta familia de enfoques incluyen, pero no están limitados, a algoritmos evolutivos (AEs), optimización basada en colonia de hormigas (ACO),

búsqueda local, búsqueda local guiada, búsqueda local iterada, *simulated annealing* (SA), *tabu search* (TS), *scatter search* (SS), *greedy randomized adaptive search process* (GRASP), redes neuronales, etc. [Glover et al. 2003]. En su evolución, estos métodos han incorporado diferentes estrategias para evitar la convergencia a óptimos locales, especialmente en espacios de búsqueda complejos. En la actualidad es indiscutible el gran impacto que el desarrollo de las metaheurísticas ha tenido en muchos campos de aplicación, tanto en el ámbito académico como en problemas del mundo real. Además, varias de las metaheurísticas mencionadas anteriormente se caracterizan por ser poblacionales y por llevar a cabo un proceso de búsqueda altamente distribuido. Esto, en términos de mejoras en la eficiencia computacional, ha derivado en desarrollos de modelos y algoritmos paralelos, que no sólo se traduce en el aprovechamiento de las capacidades de cómputo existentes en la actualidad, sino que además, un incremento en la calidad de los resultados esperados [JH 2002], [Alba 2005].

En síntesis, el presente proyecto tiene dos objetivos principales, uno de ellos, destinado a la aplicación de técnicas de minería de datos para el análisis inteligente de datos obtenidos a través de diversas mediciones *in situ* (provisas por diferentes fuentes), que permitan inferir modelos descriptivos y predictivos del comportamiento del viento. La otra parte, orientada a la investigación y desarrollo de metaheurísticas (versiones seriales y paralelas), para resolver uno de los problemas claves (distribución) asociados al proceso de instalación de aerogeneradores en un parque eólico.

2. LINEAS DE INVESTIGACION y DESARROLLO

De lo anterior, se desprenden en detalle las siguientes líneas temáticas específicas:

A.1) Organización de los datos provenientes de mediciones *in situ* (de la región patagónica o de otras fuentes disponibles para uso de la comunidad científica).

A.2) Aplicación de técnicas de Minería de Datos (tareas descriptivas y predictivas).

B.1) Investigación, desarrollo y aplicación de Metaheurísticas para la distribución eficiente de los aerogeneradores o máquinas eólicas.

B.2) Paralelización de las Metaheurísticas desarrolladas (aprovechamiento de sus características inherentes para acelerar el proceso de búsqueda y mejorar la calidad de las soluciones).

3. RESULTADOS OBTENIDOS/ESPERADOS

Avance del conocimiento científico en el área.

A partir de los objetivos planteados, se espera lograr un importante avance en conocimiento en el área de aplicación de técnicas de minería de datos en el ámbito de explotación de energía eólica. Además, se espera lograr un desarrollo de algoritmos avanzados basados en inteligencia computacional para distribución de aerogeneradores considerando situaciones próximas a la realidad dado el contexto geográfico y de trabajo del grupo de investigación.

Es importante destacar además, la consolidación del grupo en este campo de investigación a nivel regional, nacional e internacional a través de sus propios desarrollos y de la interacción en progreso con grupos de otras Universidades, en particular con el grupo del LIDIC de la Universidad Nacional de San Luis con el cual este grupo mantiene una estrecha relación de colaboración con perspectivas de consolidación hacia el futuro. Específicamente el Dr. Guillermo Leguizamón que oficiará como asesor científico externo del Proyecto con la posibilidad de realizar visitas periódicas a la Unidad Académica Caleta Olivia.

Contribución al desarrollo socioeconómico del país, región, provincia, etc.

Existen posibilidades concretas de transferencia directa o indirecta de los resultados, dada la aplicabilidad de los métodos a estudiar, a instituciones y/o empresas del medio que necesiten estudiar y resolver determinados problemas vinculados al aprovechamiento de la energía eólica.

Más importante aún, el presente proyecto se desarrollará en un contexto geográfico altamente beneficioso dada la existencia de planes concretos de explotación de energía eólica en la región, como también la existencia de grupos de investigación vinculados a la temática, con los cuales se pueden realizar intercambios de gran utilidad e impacto entre los grupos involucrados.

4. FORMACION RECURSOS HUMANOS

Dos de los integrantes, están actualmente realizando sus estudios de postgrado (Doctorado), que iniciaron en Noviembre de 2008, en la Universidad de Málaga bajo la dirección del Dr. Anrique Alba, para realizar estudios sobre Modelos y Algoritmos Paralelos en el ámbito de Metaheurísticas.

Un integrante está desarrollando su Tesis de Maestría estimando su defensa para mediados de 2009, en la Universidad Nacional de San Luis bajo la dirección del Dr. Guillermo Leguizamón.

Dos integrantes, han comenzado a definir su plan de tesis en temas afines a los propuestos en el proyecto, y han comenzado a realizar cursos de postgrado como parte inicial de sus tareas para definir su plan.

5. BIBLIOGRAFIA

- [Alba 2005] E. Alba - Parallel Metaheuristics: A New Class of Algorithms, Wiley-Interscience, 2005.
- [de San Pedro et al, 2002a] de San Pedro M., Pandolfi D., Villagra A., Lasso M., Vilanova G, Gallard R. - Adding problem-specific knowledge in evolutionary algorithms to solve W-T scheduling problems - CACIC 2002, Buenos Aires, 2002.
- [de San Pedro et al, 2003a] de San Pedro M., Lasso M., Villagra A., Pandolfi D., Gallard R. - Influence of Crossover Operators in Evolutionary Scheduling Under Multirecombined Schemes - CACIC 2003, La Plata, 2003.
- [de San Pedro et al, 2003c] de San Pedro M., Villagra A., Lasso M., Pandolfi D., Díaz Vivar M., Gallard R. - Solutions for the Weighted Number of Tardy Jobs in Single Machine Environments via Evolutionary Algorithms - CSITeA03, Rio de Janeiro, 2003.
- [de San Pedro et al, 2004] de San Pedro M., Pandolfi D., Lasso M., Villagra A., Gallard R. - Effect of Crossover Operators under Multirecombination: Weighted Tardiness, a Test Case - CEC '04 - Portland, U.S.A., 2004.
- [de San Pedro et al, 2005] de San Pedro M., Pandolfi D., Lasso M., Villagra A. - Dynamic Scheduling Approaches to solve Single Machine Problems, ASC 2005, Benidorm, España, 2005.
- [Donovan 2005] S. Donovan - Wind Farm Optimization, Department of Engineering Science, University of Auckland, New Zealand. (<http://www.orsnz.org.nz/Prizes/Papers/2005WindFarmOptimization.pdf>)
- [Donovan 2007] S. Donovan, H. Waterer, R. Archer - Mixed Integer Programming Models for Wind Farm Design. Department of Engineering Science, The University of Auckland, Electric Power Optimization Centre, Winter Workshop 2007.
- [Donovan 2008] S. Donovan - An improved mixed integer programming model for wind farm layout optimization. YPP entry, 2008.
- [Elkinton 2008] C.N. Elkinton J.F. Manwell & J.G. McGowan - Algorithms for Offshore Wind Farm Layout Optimization. Wind Engineering, Volume 32, Number 1, January 2008 , pp. 67-84(18). Multi-Science Publishing Co Ltd.
- [Escudero 2003] J.M. Escudero López - Manual de Energía Eólica. Ediciones Mundi Prensa, 2003.
- [Glover et al. 2003] F. Glover, G.H. Kochenberger (editors)- Handbook of Metaheuristics, Kluwer Academic Publishers, 2003.
- [Grady et al. 2005] S. A. Grady, M. Y. Hussaini, M. M. Abdullah - Placement of wind turbines using genetic algorithms Renewable Energy, Volume 30, Issue 2, February 2005, pp. 259-270
- [JH 2002] Journal of Heuristics - Special Issue: Parallel Meta-Heuristics. Volume 8, Number 3, Springer Netherlands, 2002.

- [Lasso et al, 2002a] Lasso M., Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Algorithms to solve the dynamic weighted tardiness problem - CACIC 2002, Buenos Aires, 2002.
- [Lasso et al, 2003a] Lasso M., de San Pedro M., Villagra A., Pandolfi D., Gallard R. - Solutions to the Dynamic Average Tardiness Problem in Single Machine Environments - CACIC 2003, La Plata, 2003.
- [Lasso et al, 2003b] Lasso M., Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Heuristics for partial and total dynamic w-t problems In single machine environments - WICC'03, Tandil, 2003.
- [Lasso et al, 2003c] Lasso M., Pandolfi D., de San Pedro M., Villagra A., Gallard R. - Heuristics to Solve Dynamic W-T problems in Single Machine Environments - CSITeA03, Rio de Janeiro, 2003.
- [Lasso et al, 2004] Lasso M., Pandolfi D., de San Pedro M., Villagra A., Gallard R. - Solving Dynamic Tardiness Problems in Single Machine Environments" – CEC '04 –Portland, U.S.A., 2004.
- [Mora et al. 2006] J.C. Mora, J.M. Calero Barón, J.M. Riquelme Santos, M. Burgos Payán - A Hybrid Evolutive Algorithm for Wind Farm Optimum Network Design. Aiesp 2006. Artificial Intelligence in Energy Systems and Power Aiesp 2006. Madeira, Portugal. pp. 1-5, 2006
- [Mora et al. 2007a] J.C. Mora, J.M. Calero Barón, J.M. Riquelme Santos, M. Burgos Payán - A Mixed Hybrid Algorithm for Integral Wind Farm Optimum Design. Proceeding of Icrepq-07. Sevilla, España. European Association for the Development of Renewable Energy, pp. 227-228, 2007.
- [Mora et al. 2007b] J.C. Mora, J.M. Calero Barón, J.M. Riquelme Santos, M. Burgos Payán - An evolutive algorithm for wind farm optimal design. Neurocomputing, Volume 70 , Issue 16-18, pp. 2651-2658, October 2007.
- [Orallo et al. 2003] C. Ferri Ramírez; J. Hernández Orallo; M.J. Ramírez Quintana Introducción A La Minería De Datos (Pearson Educación), 2003.
- [Pandolfi et al, 2002a] Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Multirecombining Random and Seeds with Studs in evolutionary algorithm to solve W-T Scheduling problems - CSITeA-02, Foz Iguazú, Brasil, 2002.
- [Pandolfi et al, 2002b] Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Stud Mating Immigrants in Evolutionary Algorithms to solve the Earliness-Tardiness Scheduling Problem - Cybernetics and Systems: An International Journal, Ed Taylor & Francis, pp 391-400, 2002 Vol 33 Number 4 June 2002 ISSN 0196-9722
- [Pandolfi et al, 2002c] Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Inserting Problem Specific-Knowledge in Multirecombined Evolutionary Algorithms - WICC'02, 2002.
- [Pandolfi et al, 2002d] Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - A survey an multirecombined evolutionary approach for single machine scheduling – AEB'02, Merida, España, 2002.
- [Pandolfi et al, 2003a] Pandolfi D., Lasso M., de San Pedro M., Villagra A., Gallard R. - Knowledge Insertion: an Efficient Approach to Reduce Search Effort in Evolutionary Scheduling - CACIC 2003, La Plata, 2003.
- [Pandolfi et al, 2003b] Pandolfi D., Lasso M., de San Pedro M., Villagra A., Gallard R. - Evolutionary algorithms to solve average tardiness problems in single machine environments - CSITeA03, Rio de Janeiro, 2003.
- [Pandolfi et al, 2004] Pandolfi D., Lasso M., de San Pedro M., Villagra A., Gallard R. - Knowledge Insertion: an Efficient Approach to Reduce Search Effort in Evolutionary Scheduling - Journal of Computer Science & Technology, pp 109-114, Vol 4 Number 2 Agosto 2004 ISSN 1666-6038
- [Villagra et al, 2004a] Villagra A., de San Pedro M., Lasso M., Pandolfi D. - Multirecombined Evolutionary Algorithm inspired in the Selfish Gene Theory to face the Weighted Tardiness Scheduling Problem - Iberamia 2004 - Puebla, Mexico, 2004.
- [Villagra et al, 2004b] Villagra A., de San Pedro M., Lasso M., Pandolfi D. - Algoritmos Evolutivos inspirados en la Teoría del Gen Egoísta - WICC'04 – Neuquén, 2004.
- [Villagra et al, 2005a] Villagra A., de San Pedro M., Lasso M., Pandolfi D. – Algoritmo Evolutivo basado en el mecanismo de haplodiploidia para resolver el problema de planificación de weighted tardiness. CACIC 2005.
- [Villagra et al, 2005b] Villagra A., de San Pedro M., Lasso M., Pandolfi D. - Optimización guiada por Algoritmos Evolutivos Multirecombinados inspirados en la Teoría del Gen Egoísta para resolver Problemas de Weighted Tardiness – XI Reunión de Trabajo en Procesamiento de la Información y Control, 2005.
- [Weibull 1951] W. Weibull - A statistical distribution function of wide applicability. Journal of Applied Mechanics, Trans. ASME 18(3), 293-297, 1951.
- [Witten 1999] I.H. Witten - Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, The Morgan Kaufmann Series in Data Management Systems, 1999.
- [Witten 2005] I.H. Witten - Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, Morgan Kaufmann Series in Data Management Systems, 2005.

Propuesta para resolver el problema de predicción de complejos de proteínas usando agentes inteligentes que aprenden usando técnicas de sistemas inmunológicos artificiales

Cristian S. Rocha

Laboratorio de Investigación y Desarrollo de Nuevas Tecnologías (LIDeNTec),
Gerencia de Sistemas y Telecomunicaciones (GSyT),
Administración Nacional de la Seguridad Social (ANSES)

5 de abril de 2009

Contexto

La ANSES, coordinada por la GSyT, busca explotar sus recursos computacionales no aprovechados fuera del horario laboral en beneficio de otras instituciones que los requieran. En este contexto inició un convenio marco con el Hospital Italiano (HI) para desarrollar herramientas de bioinformática que puedan ejecutarse bajo éstas restricciones.

El LIDeNTec, dentro de este convenio, decidió aplicar técnicas de Inteligencia Artificial (IA) de fácil paralelización, como son los Agentes Inteligentes (AI), para resolver el problema de inferir estructuras de complejos de proteínas.

Resumen

En este trabajo se propone una técnica original del área de la inmunobiología para predecir las estructuras tridimensionales de complejos de proteínas - macromoléculas conformadas por dos o más proteínas encastradas para cumplir una función metabólica o estructural de la célula. La inmunobiología describe las técnicas usadas por el sistema inmune de los vertebrados para identificar patógenos. Esas mismas técnicas fueron recogidas por la IA

para resolver problemas de clasificación. El trabajo recoge la capacidad innata del Sistema Inmunológico (SI) natural para aprender la geometría de las moléculas, pero no para identificar cuerpos extraños, sino con el objetivo de reconstruir complejos de proteínas comparando las geometrías aprendidas.

Palabras claves: reconstrucción de complejos, docking de proteínas, interacción proteína-proteína, agentes inteligentes, sistema inmunológico artificial.

1. Introducción

Muchos de los procesos biológicos requieren conocer la estructura tridimensional de complejos macromoleculares. Conocer la información estructural de dichos complejos puede ayudar a desarrollar nuevos compuestos farmacéuticos. Sin embargo los estudios cristalográficos y de Resonancia Magnética Nuclear (RMN) siguen una visión reduccionista donde las unidades de los complejos son más estudiadas que los mismos complejos. Esto se refleja en la base de datos de proteínas Protein Data Bank (PDB) donde existen una discrepancia en el número de unidades de proteínas (c. 3000) con respecto al número de complejos (c. 300).[7]

Es por ello que hay una fuerte necesidad de algoritmos de predicción de estructuras de complejos usando únicamente las coordenadas espaciales de los átomos de las proteínas que los componen. [7]

Existen diversas estrategias donde los algoritmos actuales se especializan: reconstruir complejos conocidos (*bounded docking*), y reconstruir complejos desconocidos (*unbounded docking*). En el primer caso existen técnicas que pueden recorrer todo el espacio de soluciones en un tiempo aceptable, mientras que la segunda, al requerir flexibilidad de los cuerpos involucrados, la cantidad de dimensiones y el tamaño del espacio de solución hacen al problema de muy difícil solución.[6]

En la biología de los vertebrados existe un sistema capaz de recorrer el espacio de soluciones en forma heurística y eficientemente: el Sistema Inmunológico (SI). Éste utiliza anticuerpos, proteínas generadas por células plasmáticas, con el objetivo de identificar sustancias antígenas. Cada anticuerpo tiene una única estructura que le permite unirse de forma específica a un único antígeno.[8]

La compleja heurística del SI fue simplificada y descrita por Dasgupta [3] para su uso en la computación como una técnica de Aprendizaje Automático (AA). De los diferentes niveles de complejidad detallados por Dasgupta [3], la teoría de redes inmunológicas se asocia fácilmente con lo que se conoce como Agentes Inteligentes (AI) descrito por Fyfe and Jain [5].

En este trabajo se propone resolver el problema de predicción de estructuras de complejos usando redes de AI que aprenden, a través de un Sistema Inmune Artificial (SIA), la superficie de la estructura de la proteína y su complementario. Toda la información recolectada se usa en un segundo paso para identificar las regiones de unión que permiten determinar el conjunto de complejos más probables de encontrar en la naturaleza.

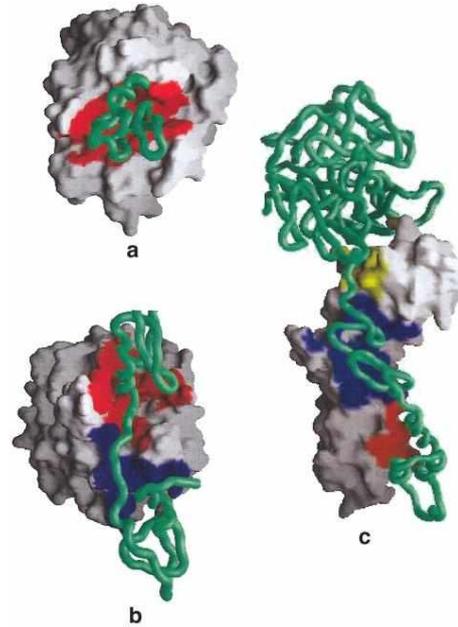


Figura 1: Parches identificados en complejos inhibidores de proteasas. Una de las proteínas del complejo tiene coloreada en la superficie los parches, la otra está representada por su backbone. Las proteínas listadas pueden encontrarse en PDB como A: 2pct, B: 1toc y C:1dan [2]

2. Líneas de Investigación y Desarrollo

La superficie geométrica que quiere reconocerse se la conoce como Superficie Accesible al Solvente (SAS). [9] Esta superficie es la frontera física entre las moléculas del solvente y la molécula en sí. Esta superficie no tiene una topología única y puede variar según la flexibilidad de la molécula. Por ello es vital estudiar características, aunque sean parciales, de estas superficies para poder identificar la interacción entre los cuerpos, y así predecir mejor la estructura de los complejos.

Chakrabarti and Janin [2] logro identificar parches de unión de aproximadamente 800Å^2 (ver figura 1). Cada parche tiene un anillo de átomos accesibles al solvente y un núcleo de átomos no accesibles. Gracias a esta descripción podemos definir lo que para nuestro SIA es un anticuerpo y un antígeno.

Definición 1 *Un anticuerpo es una malla $A_i \in \mathbb{R}^3$ $i \in [0, l]$ capaz de envolver una región de la SAS de aproximadamente 800Å^2 con un cierto grado de tolerancia t .*

Definición 2 *Un antígeno es una superficie $\bar{A}_i \in \mathbb{R}^3$ $i \in [0, n]$ que representa la SAS de la proteína.*

Vease que la tolerancia t es el parámetro de flexibilidad de la afinidad a los antígenos. Cuando t tiende a cero, el anticuerpo es afín a un conjunto menor de proteínas. El valor de t , que consideramos biológicamente correcto es de $1,7\text{Å}$, el diámetro promedio de la molécula de agua.[9]

Para seleccionar nuestra lista de anticuerpos usamos el algoritmo Conalg (Clonación) [4]:

1. Generar una población aleatoria de anticuerpos
2. Presentación a los antígenos
 - a) Evaluación de afinidad
 - b) Selección de clones y expansión
 - c) Maduración de la afinidad
 - d) Reemplazar anticuerpos con otros generados
3. Ciclar el paso 2 hasta que se cumpla la guarda de parada.

Aunque nuestra SIA generará n anticuerpos aleatoriamente, requerimos de superficies iniciales. Dos estrategias válidas para definir las son:

- Si la muestra de aprendizaje es un conjunto de complejos conocidos, las superficies iniciales son los parches descritos por Chakrabarti and Janin [2]. Por ejemplo, si tomamos los complejos de la figura 1 se usarán los 6 parches identificados.
- Si la muestra de aprendizaje es un conjunto de proteínas se toma una parte de la superficie accesible al solvente.

A cada instancia de la nueva población se le incrementa un valor aleatorio $I_i \in \mathbb{R}^3$ $i \in [0, l]$ para generar la próxima generación de anticuerpos. Para que las instancias sean aprobadas hay que verificar que la superficie no exceda los 800Å .

La evaluación de la afinidad de los anticuerpos usa la técnica de correlación para la predecir el docking [10]. Este algoritmo tiene la ventaja de ser exhaustivo y rápido para instancias pequeñas de datos. El resultado es una lista de configuraciones con su correspondiente puntuación dada por correlacionar el anticuerpo con el antígeno. La mejor puntuación es definida como la afinidad del anticuerpo al antígeno.

Del conjunto de anticuerpos generados y presentados, solo se elegirán m de mejor afinidad que se incluirán en el conjunto de anticuerpos ya elegidos. Aleatoriamente se eliminarán algunos y se quedarán con un conjunto reducido a m anticuerpos.

Es fácil de observar que este sistema de aprendizaje sesga su elección de anticuerpos a los cuerpos que les fueran presentados. Es por ello que se requiere de un sistema que flexibilice lo aprendido, distribuyendo el conocimiento entre varios generadores de anticuerpos. Estos generadores son AI.

Definición 3 *Los Agente Inmunes son AI que actúan como repositorios y generadores de anticuerpos.*

Estos agentes inmunes actúan individualmente para seleccionar los anticuerpos más afines para un determinado conjunto de antígenos, y actúan en conjunto para determinar probable interacción entre proteínas. En la figura 2 se presenta un diagrama de proceso de un agente donde se describe el aprendizaje y la comunicación con otros agentes.

Para identificar si dos anticuerpos son complementarios se usa la técnica de correlación descripta anteriormente.

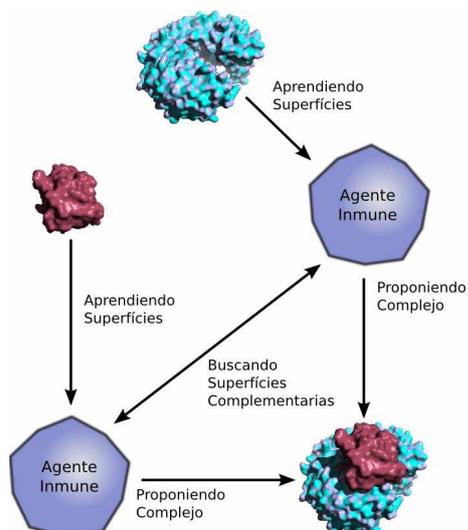


Figura 2: Diagrama de proceso de un agente inmune

3. Resultados Obtenidos/Esperados

Este trabajo se concentra en la utilización de técnicas de AA en el área de Bioinformática Estructural (BE). El desarrollo de estas técnicas están sujetas a no solo al estudio computacional de los algoritmos, sino también a posibles interpretaciones biológicas, permitiendo un intercambio de conceptos e ideas entre ambas áreas. Al ser un modelo teórico preliminar solo se aproximó desde el punto de vista computacional. En una segunda etapa esperamos un constante intercambio entre expertos de Microbiología, Bioquímica, Física y Computación.

El resultado obtenido es una reducción coherente del problema de predicción de estructuras de complejos al problema de identificación de antígenos que se describe en los SIA.

Esta aproximación deja abierta la discusión sobre la utilidad de técnicas de AA en el área de la BE, ya que en una exhaustiva búsqueda bibliográfica no se han encontrado una solución proveniente de esa área.

Se espera poder implementar los algoritmos a la brevedad y testear los resultados usando los datos descritos por Andrusier et al. [1] como fuente de aprendizaje. Aunque la muestra

está acotada a un conjunto reducido de complejos, sesgados a la facilidad de cristalización de las macromolécula, se espera conseguir un comportamiento aproximado al del Sistema Inmune Natural y así comparar resultados artificiales con resultados biológicos ya conocidos.

4. Formación de Recursos Humanos

El siguiente trabajo se desarrollará con la participación de los grupos de Bioinformática del Instituto de Ciencias Básicas y Medicina Experimental (ICBME) del Hospital Italiano (HI), el grupo de Bioinformática Estructural del Departamento de Química Biológica (DQB) y el grupos de Imágenes del Departamento de Computación (DC) ambos de la Facultad de Ciencias Exactas y Naturales (FCEyN) de la Universidad de Buenos Aires (UBA), el grupo de Bioinformática del Centro de Estudio e Investigación (CEeI) de la Universidad Nacional de Quilmes (UNQ) y los grupos de Centro de Tecnología Médica (CTM) y Análisis Matemático de Imágenes (AMI) de la Universidad de Las Palmas de Gran Canarias (ULPGC), bajo la coordinación del Laboratorio de Investigación y Desarrollo de Nuevas Tecnologías (LIDeNTec).

Existe un gran interés de desarrollar la idea, prestar recurso y delinear aquellos experimentos que permitan un estudio exhaustivo de esta técnica en casos reales. Es por ellos que cada grupo tomará una parte del trabajo y los desarrollará bajo un marco de tesis de grado.

5. Bibliografía

- [1] Nelly Andrusier, Efrat Mashiach, Ruth Nussinov, and Haim J. Wolfson. Principles of flexible protein-protein docking. *Proteins: Structure, Function, and Bioinformatics*, 73(2):271–289, 2008. doi: 10.

- 1002/prot.22170. URL <http://dx.doi.org/10.1002/prot.22170>.
- [2] Pinak Chakrabarti and Joël Janin. Dissecting protein-protein recognition sites. *Proteins: Structure, Function, and Genetics*, 47(3):334–343, 2002. doi: 10.1002/prot.10085. URL <http://dx.doi.org/10.1002/prot.10085>.
- [3] D. Dasgupta. An artificial immune system as a multi-agent decision support. In *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, volume 4, pages 3816–3820 vol.4, 1998. doi: 10.1109/ICSMC.1998.726682. URL <http://dx.doi.org/10.1109/ICSMC.1998.726682>.
- [4] Dipankar Dasgupta. *Artificial Immune Systems and Their Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1998. ISBN 3540643907. URL <http://portal.acm.org/citation.cfm?id=552363>.
- [5] Colin Fyfe and Lakhmi Jain. Teams of intelligent agents which learn using artificial immune systems. *J. Netw. Comput. Appl.*, 29(2):147–159, 2006. ISSN 1084-8045. doi: 10.1016/j.jnca.2004.10.003. URL <http://dx.doi.org/10.1016/j.jnca.2004.10.003>.
- [6] I. Halperin, B. Ma, H. Wolfson, and R. Nussinov. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47(4):409–443, June 2002. ISSN 1097-0134. doi: 10.1002/prot.10115. URL <http://dx.doi.org/10.1002/prot.10115>.
- [7] Thomas Lengauer, Raimund Mannhold, Hugo Kubinyi, and Hendrik Timmerman, editors. *Bioinformatics: From Genomes to Drugs*. Wiley-VCH, 1 edition, July 2001. ISBN 3527299882. URL <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/3527299882>.
- [8] Kenneth Murphy. *IMMUNOBIOLOGY 7 PB (Janeway's Immunobiology) (Immunobiology: The Immune System (Janeway))*. Garland Science, November 2007. ISBN 0815341237. URL <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0815341237>.
- [9] A. Shrake and J. Rupley. Environment and exposure to solvent of protein atoms. lysozyme and insulin. *Journal of Molecular Biology*, 79(2):351–364, September 1973. ISSN 00222836. doi: 10.1016/0022-2836(73)90011-9. URL [http://dx.doi.org/10.1016/0022-2836\(73\)90011-9](http://dx.doi.org/10.1016/0022-2836(73)90011-9).
- [10] S. J. Wodak and R. Méndez. Prediction of protein-protein interactions: the capri experiment, its evaluation and implications. *Curr Opin Struct Biol*, 14(2):242–249, April 2004. ISSN 0959-440X. doi: 10.1016/j.sbi.2004.02.003. URL <http://dx.doi.org/10.1016/j.sbi.2004.02.003>.

Inteligencia Artificial en Tiempo Real: Razonamiento Rebatible en Juegos Estratégicos Digitales

Diego C. Martínez

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial - LIDIA
Departamento de Ciencias e Ingeniería de la Computación

Universidad Nacional del Sur – Avenida Alem 1253, Bahía Blanca, Argentina

CONTEXTO

Esta línea de investigación se encuadra dentro de los lineamientos del Proyecto de Investigación “*Representación de Conocimiento, Argumentación y Apoyo para la toma de Decisiones*”, del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur.

RESUMEN

Los Sistemas de Inteligencia Artificial en Tiempo Real son sistemas que interactúan en forma continua con el entorno, tratan con información incompleta, gestionan eventos sincrónicos y asincrónicos y garantizan respuestas en tiempos establecidos. En esta línea de investigación se estudia la aplicación de sistemas de razonamiento rebatible a juegos estratégicos, principalmente aquellos digitales con buen grado de interactividad como los de simulación en tiempo real.

Palabras clave: *inteligencia artificial, juegos, agentes, toma de decisiones.*

1. INTRODUCCION

El desarrollo de programas capaces de jugar juegos compitiendo con los humanos ha sido una costumbre en Inteligencia Artificial, probablemente desde 1950 cuando Claude Shannon propuso una forma de programar las computadoras para jugar ajedrez [1]. Con el transcurso de los años se han agregado otras alternativas atractivas, como las damas, el bridge, el backgammon, y hasta crucigramas. Los juegos se han convertido en un campo de estudio por sí mismo dentro de la Inteligencia Artificial, especialmente aquellos que demandan comportamientos estratégicos. Existe actualmente un interés especial en esta disciplina, debido en parte a la popularidad masiva de los juegos interactivos digitales.

El desarrollo de juegos de entretenimiento interactivo digital es una de las áreas más activas dentro de la industria del software. Se producen anualmente cientos de juegos para satisfacer una audiencia exigente en un mercado altamente competitivo, lo que obliga a los desarrolladores a utilizar los últimos avances tecnológicos. Esta industria es multidisciplinaria, destacándose la computación gráfica, la ingeniería de software y, con mayor fuerza en los últimos años, la inteligencia artificial.

Actualmente existen nuevos congresos y conferencias sobre Inteligencia Artificial aplicada a los juegos, que pretenden fomentar el diálogo entre los investigadores y los desarrolladores en pos del interés común de los desafíos interactivos. Por ejemplo, la AAAI (*Association for the Advancement of Artificial Intelligence*) organiza la Conferencia AIIDE - *Artificial Intelligence and Interactive Digital Entertainment* como punto de interacción entre la industria y la academia.

Los juegos interactivos digitales son campo de interés para diversas áreas de investigación en Inteligencia Artificial, entre ellas el razonamiento de sentido común, la planificación, el modelado cognoscitivo, el razonamiento espacial y temporal, el modelado del oponente y el aprendizaje automatizado [3,5,6,7]. La problemática de la automatización del comportamiento inteligente es constante en los juegos y motiva la investigación científica y tecnológica.

En respuesta a las motivaciones industriales y científicas, han surgido desde el mundo académico una serie de competencias de juegos destinadas a aplicar técnicas de Inteligencia Artificial de diversa índole. Esto permite que cada investigador, con sus formalismos y tecnología particular, puede aplicarla a resolver un problema concreto: ganar un determinado juego bajo reglas específicas. Existe un interés profundo en el desarrollo de agentes que son capaces de desenvolverse en forma inteligente en escenarios dinámicos como los juegos de estrategia, porque básicamente son el punto de partida para la aplicación de nuevas tecnologías a dominios más amplios, como la robótica, la toma de decisiones y la simulación computacional. Actualmente existen varias competencias, como las *Open Real Time Strategy Competition*, basado en un juego de simulación y estrategia en tiempo real organizado por la Universidad de Alberta, o la *AAAI Computer Poker Competition*, donde agentes inteligentes procuran elaborar y seguir estrategias para jugar al póker en diversas modalidades. La AAAI también organiza *General Game Contest*, donde las reglas son presentadas al momento de la partida, y el agente inteligente debe comprenderlas y jugar de manera inteligente.

2. LINEAS DE INVESTIGACION y DESARROLLO

El objetivo general de esta línea de investigación es estudiar técnicas para la construcción de agentes inteligentes para el desenvolvimiento en juegos interactivos digitales que requieren planes estratégicos para alcanzar determinadas metas. Son buenos ejemplos de juegos aquellos planteados en las competencias académicas mencionadas anteriormente. Los agentes deben ser tales que la toma de decisiones, la planificación de acciones y el aprendizaje ocurran bajo restricciones de tiempo y/o espacio críticas. Deben observar ciertas reglas y actuar en forma competitiva o cooperativa para alcanzar objetivos prefijados. Además, el factor de interactividad con el usuario requiere comportamiento dinámico y adaptativo. Estamos especialmente interesados en el uso de formalismos de razonamiento rebatible para la toma de decisiones.

Básicamente existen dos tipos de juegos dinámicos que son exitosos comercialmente y demandan mayores innovaciones tecnológicas en Inteligencia Artificial. Ellos son los juegos de *estrategia* y los juegos de *inmersión*.

Juegos de estrategia

Los juegos digitales son un escenario adecuado y moderno para el desarrollo en inteligencia artificial en tiempo real. Entre ellos son interesantes los juegos de estrategia, que pueden presentarse bajo dos modalidades:

- estrategia *en tiempo real*, en donde los jugadores realizan acciones de manera simultánea,
- estrategia *por turnos*, donde los jugadores realizan acciones de manera alternada.

Las dos modalidades pueden presentarse como *juegos abstractos*, sin temática específica asociada en el mundo real (por ejemplo, el ajedrez), o como *juegos de simulación* donde se recrea una situación realista, con la mecánica del dinamismo del mundo real. En general, es frecuente que los juegos abstractos sean por turnos mientras que lo de simulación funcionen en tiempo real. En estos últimos el tiempo invertido para la toma de decisiones es limitado, pues diversas acciones estarán ocurriendo en forma simultánea y el estado del mundo cambia continuamente. En los juegos de simulación digitales, más conocidos como RTS (*Real Time Strategy*) se modela un terreno con diversos obstáculos y recursos económicos que deben ser administrados correctamente. Esto lo hace particularmente interesante para la investigación en Inteligencia Artificial por la riqueza de los problemas a resolver [2].

Por lo general, en los juegos abstractos el jugador puede tener un conocimiento completo de lo que

ocurre en el escenario, mientras que en los juegos de simulación esa información es limitada. Por ejemplo, parte del escenario está oculto y permanecerá así a menos que diversas acciones sean tomadas para revelar su contenido. Un buen ejemplo de plataforma que implementa un juego de estrategia en tiempo real es la antes mencionada *Open Real Time Strategy – ORTS*, desarrollada exclusivamente para la investigación científica [8].

Juegos de inmersión en primera persona

Otro tipo de juegos que resulta de interés son los denominados Juegos en Primera Persona (*First-person shooters*, o simplemente FPS). En estos juegos la perspectiva del jugador en primera persona da un efecto de inmersión completa en el juego. Aquí la visibilidad del escenario es limitada a la posición específica del jugador y a su punto de vista. El aspecto interesante en este modelo de juego es que el jugador no es omnipresente como en los de tablero, sino que posee un conocimiento parcial y altamente volátil del escenario y del oponente. El conocimiento depende de la perspectiva sensorial según la ubicación del jugador. Además, los FPS son generalmente juegos que buscan el mayor nivel de realismo gráfico y físico de los escenarios simulados, lo que resulta un atractivo para el modelado de agentes inteligentes.

Desarrollo de clientes

Como es de suponer, la mayoría de los productos son comerciales y con restricciones de uso por derechos de copia y secreto industrial. Sin embargo, algunas aplicaciones han sido liberadas y modificadas, o directamente creadas para la investigación científica. Generalmente se organizan en una arquitectura cliente-servidor, donde el servidor simula un escenario de juego y el cliente observa el escenario, decide sus acciones y las ejecuta por medio del servidor. Esa es la arquitectura de la plataforma ORTS (Universidad de Alberta) y de la AAAI Poker Competition, en diferentes versiones. Existe también una adaptación del juego FPS comercial *Unreal Tournament* que permite agregar clientes inteligentes (denominados bots) en la modalidad multijugador, con la particularidad de ofrecer libertad en la implementación de dichos clientes. Es decir, pueden codificarse en lenguajes como Java o Prolog, pues la comunicación con los servidores del juego es por medio de protocolos específicos. Existen algunas librerías que facilitan esta comunicación, como *Gamebots* [9]

En el año 2008 en el LIDIA se construyó un cliente inteligente para la competencia ORTS en su Categoría 2. Aquí se simula un escenario con montañas y obstáculos naturales, donde cada jugador posee cinco bases (edificios) y cincuenta tanques que pueden moverse y atacar bajo órdenes, en

tiempo real. Con ellos cada jugador debe eliminar las bases del oponente. La estrategia utilizada en esta ocasión se basa en una implementación de pathfinding para formaciones de unidades (tanques en este caso) con una política de movimientos ante enfrentamientos con adversarios. Esta primera experiencia fué altamente positiva y se espera avanzar en categorías más complejas.

También se planea participar en las competencias de póker de la AAAI, con clientes que ya están en desarrollo. Existe actualmente una implementación parcial sobre una versión restringida del Texas Hold'em Poker, que utiliza algoritmos genéticos para la determinación de las apuestas y la toma de decisiones.

Estos primeros pasos permiten comprender la complejidad inherente de cada ambiente de juego, lo que facilita desarrollos futuros. Nuestro interés primario, fundamentalmente, es la aplicación de formalismos de razonamiento rebatible en estos escenarios.

El rol del razonamiento rebatible

Como afirmamos anteriormente, el hilo conductor de la línea de investigación en las diversas gamas de juegos es la utilización de mecanismos de razonamiento rebatible. En este contexto es de especial interés el uso de argumentación rebatible, que constituye gran parte de los desarrollos científicos y técnicos del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial. La argumentación rebatible [4] es un mecanismo de razonamiento no monótono en donde la aceptación o el rechazo de una proposición dependen de un análisis entre argumentos a favor y en contra de esa proposición. Usualmente es utilizada bajo diferentes formalizaciones para capturar aspectos del razonamiento del sentido común y la representación de información incompleta y potencialmente inconsistente. En el Laboratorio LIDIA de la Universidad Nacional del Sur se han desarrollado varios formalismos de argumentación, en diferentes niveles. Existe principalmente una implementación concreta de una plataforma para la Programación en Lógica Rebatible (conocido por sus siglas en inglés DeLP, Defeasible Logic Programming) que puede usarse para la elaboración de agentes inteligentes. Se ha explorado levemente la integración de DeLP en una arquitectura BDI, aunque no en el área de juegos estratégicos. Más aún, existen otros modelos cognoscitivos para los cuales la integración con DeLP no ha sido estudiada todavía. Existe en nuestro Laboratorio también un gran desarrollo sobre semántica de argumentación, basada en formalismos abstractos, como los Sistemas Argumentativos Extendidos (Extended Argumentation Framework) o más recientemente los Sistemas Argumentativos de Ataques Variados (Varied-Strength Attacks). Estos estudios

semánticos proveen la base para nuevas ideas e implementaciones en áreas aún no exploradas, como los juegos de simulación y estrategia.

Parte de los objetivos es combinar ideas y mecanismos de argumentación rebatible en agentes inteligentes capaces de jugar a juegos de estrategia. La argumentación rebatible es un área creciente en Inteligencia Artificial, con grandes expectativas de aplicación en diversos ámbitos, como la asistencia para la toma de decisiones o el razonamiento jurídico. Sin embargo, poco se ha estudiado el papel en general de la argumentación rebatible en escenarios con restricciones específicas, como los juegos de simulación estratégica y la toma de decisiones en tiempo real. En estos escenarios los agentes inteligentes deben analizar la información disponible, elaborar un plan estratégico y eventualmente readaptar planes ante cambios en la situación del juego. La argumentación rebatible es un formalismo de cierta complejidad que requiere optimización en los procesos de razonamiento en tiempo real, tal cual ocurre en los juegos de estrategia. Los juegos son aptos también para técnicas de aprendizaje automatizado y modelado del oponente. La consideración de formalismos de argumentación rebatible en este contexto permite evaluar su integración con otras áreas de Inteligencia Artificial.

3. RESULTADOS OBTENIDOS/ESPERADOS

Con el avance de esta nueva línea de investigación esperamos obtener pruebas empíricas de efectividad de los formalismos de argumentación rebatible en el área de juegos estratégicos, principalmente por medio de las competencias académicas tradicionales. Se espera también la especificación de nuevos formalismos para la investigación básica en Inteligencia Artificial y para el desarrollo de aplicaciones en otras áreas.

Esta línea de investigación además marca el comienzo del Grupo de Investigación en Juegos para el Laboratorio de Investigación y Desarrollo en Inteligencia Artificial, que agrupa a los investigadores con intereses comunes sobre la problemática de los juegos en general, siguiendo el perfil de grupos extranjeros similares, como en la Universidad de Alberta o en la Universiteit Maastricht.

4. FORMACION DE RECURSOS HUMANOS

Esta línea de investigación resulta especialmente atractiva para los estudiantes, lo que permite iniciarlos en actividades de investigación científica y tecnológica. Hasta el momento, bajo esta temática, se han realizado tres tesis de grado, dos están

actualmente en desarrollo y al menos otras dos están siendo planificadas para el próximo semestre.

La investigación en inteligencia artificial en tiempo real es una actividad compleja que cobra importancia en el estado actual del arte en juegos interactivos digitales. Por esta razón existen perspectivas de transferencia tecnológica en la industria local y regional.

Los escenarios acotados de juegos, además, son el comienzo de la aplicación de formalismos a dominios más complejos, por lo que resulta atractivo para la formación inicial de recursos humanos de posgrado.

5. BIBLIOGRAFIA

[1] *Programming a Computer Playing Chess*. Claude Shannon. Philosophical Magazine, Ser. 7, vol 41, No. 314. 1950.

[2] *Real Time Strategy. New AI Research*. Michael Buro. Proceedings of the International Joint Conference on AI 2003, Acapulco, Mexico, pp.1534—1535.

[3] *Intelligent Agents in Computer Games*. Van Lent, Laird, et al. Proceedings of the National Conference on Artificial Intelligence, July 1999, Orlando, FL, pp. 929-930.

[4] *Defeasible logic programming: An argumentative approach*. García, A. J., and Simari, G. R. Theory and Practice of Logic Programming 4(1-2):95–138. 2004.

[5] *Adversarial Planning Through Strategy Simulation*, F. Sailer, M. Buro, and M. Lanctot, Computational Intelligence in Games, pp 80-87. Hawaii USA, 2007.

[6] *Efficient Triangulation-Based Pathfinding*. D. Demyen and M. Buro, Proceedings of the AAAI conference, Boston 2006, pp.942-947.

[7] *Learning to be a Bot: Reinforcement Learning in Shooter Games*. Michelle McPartland, Marcus Gallagher. Proc. of the Artificial Intelligence and Interactive Digital Entertainment 2008.

[8] *Open Real Time Strategy – ORTS. A Free Software RTS Game Engine*. Home page: <http://www.cs.ualberta.ca/~mburo/orts/>

[9] *Gamebots: A 3D Virtual World Test-Bed For Multi-Agent Research*. Adobbati, R., Marshall, A., Scholer, A., Tejada, S., Kaminka, G., Schaffer, S., & Sollitto, C. (2001)

Seguridad y gestión del ambiente basado en un Sistema Multi-agentes

Rasjido José Alberto, Vidal Pablo, Lasso Marta, Villagra Andrea, de San Pedro Eugenia, Pandolfi Daniel
Laboratorio de Tecnologías Emergentes (LabTEM)
Unidad Académica Caleta Olivia - Universidad Nacional de La Patagonia Austral
(9011) Caleta Olivia – Santa Cruz - Argentina
e-mail: {jrasjido,pjvidal, mlasso, avillagra, edesanpedro, dpandolfi }@uaco.unpa.edu.ar

CONTEXTO

Esta línea de Investigación es una continuación del proyecto “*Administración de Edificios Inteligentes, Mediante Sistemas Multi-agentes*” que se desarrolla en el marco del Laboratorio de Tecnologías Emergentes (LabTEM) de la Universidad Nacional de la Patagonia Austral.

RESUMEN

Una de las preocupaciones que afectan actualmente al hombre, es la seguridad y la protección de sus pertenencias, como así también la posibilidad de ser reconocido por el ambiente. Hasta el momento se ha logrado desarrollar un prototipo donde se puede mantener información sobre los ingresos y egresos de individuos, como así también la generación de reportes y el mantenimiento de datos históricos de todas las transacciones llevadas a cabo por los usuarios, brindando autonomía a los individuos mediante un código de identificación que poseen. Siguiendo esta línea, el grupo de trabajo se encuentra dedicado a la incorporación de dispositivos capaces de generar condiciones ambientales, tales como iluminación y temperatura, de acuerdo a la preferencia de los individuos, haciendo uso del paradigma de sistemas Multiagentes, como base para la gestión de edificios que permitan optimizar las condiciones del ambiente de trabajo y el ahorro energético.

Palabras clave: *Edificios Inteligentes, Control y Verificación, Sistemas Multi-agentes.*

1. INTRODUCCION

Existe actualmente un gran interés en la automatización de ambientes y/o edificios y términos que han ganado popularidad en estos últimos tiempos, tienen que ver con “domótica” y “edificios inteligentes” entre otras. Se entiende por “domótica” al conjunto de servicios proporcionados por sistemas tecnológicos integrados para satisfacer las necesidades básicas de seguridad, comunicación, gestión energética y confort del hombre y de su entorno más cercano [1] y por “edificio inteligente” como aquel que utiliza tecnología computacional para controlar en forma automática su funcionamiento de manera tal de optimizar el confort del usuario, el consumo de recursos, la

seguridad y la eficiencia del trabajo [2] ambas apuntan básicamente al uso de tecnologías que, sin modificar los hábitos de las personas, permitan un mayor confort y seguridad para los individuos como así también un consumo responsable de recursos energéticos.

Los problemas con estas características han sido abordados usualmente mediante técnicas y enfoques que suelen ser referenciados con el término general de “sistemas inteligentes”.

Se puede definir un agente inteligente como una entidad (de software o hardware) autónoma, conectada directamente al ambiente del problema, capaz de mostrar un comportamiento flexible (reactivo, proactivo y social) tendiente a cumplir con sus objetivos de diseño. Si existe más de un agente interactuando en un ambiente, el sistema multiagente resultante tiene un valor agregado considerable ya que los agentes pueden compartir sus conocimientos y habilidades, pudiendo realizar actividades que van más allá de sus capacidades individuales. [3,4,5,6,7,8,9,10].

La problemática vinculada a la administración de ambientes inteligentes, tiene muchas características que suelen ser consideradas como apropiadas para ser abordadas en el contexto de sistemas multiagentes. Este hecho se ha visto reflejado en la creciente cantidad de trabajos que han comenzado a utilizar a los sistemas multiagentes como metáfora y soporte básico para el análisis, desarrollo e implementación de sistemas para la administración autónoma e inteligente de edificios para uso familiar, oficinas y de tipo industrial. A partir de esta tendencia han surgido un número considerable de desafíos, tanto teóricos como prácticos a los cuales el paradigma de los sistemas multiagentes debe dar una respuesta satisfactoria tomando en cuenta las particularidades y requerimientos de esta área de aplicación.

Algunas de las aplicaciones que podemos identificar a partir de la noción de edificio inteligente son las siguientes:[11]

- *Ahorro de energía:* la automatización de edificios permite un menor consumo de energía, controlando de manera flexible y dinámica la calefacción, refrigeración e iluminación de las distintas partes de un edificio.
- *Servicios personalizados:* es común que en un edificio inteligente sus habitantes puedan expresar sus preferencias respecto a la intensidad de la luz y

calor de su oficina. Este tipo de servicios pueden ser muy útiles y placenteros para las personas en la medida que el edificio pueda satisfacer automáticamente dichas preferencias.

- *Seguridad*: este aspecto abarca tanto la seguridad de los habitantes del edificio como así también la del edificio en sí mismo.

- *Vigilancia*: en este caso, un sistema de detección de intrusos puede ser establecido que cierre todas las puertas y ventanas automáticamente y dé aviso a los servicios de vigilancia tradicionales.

Entre los aspectos que deben ser contemplados por un edificio inteligente podemos citar:

- *Flexibilidad*: Idealmente, el sistema debería tener la capacidad de detectar y adaptarse automáticamente a las extensiones y modificaciones en las políticas que se adoptan en el edificio.

- *Escalabilidad*: el sistema debería funcionar adecuadamente en pequeños edificios como así también en edificios con muchos pisos y habitaciones.

- *Robustez*: las fallas en el sistema, no deberían tener un gran impacto.

- *Amigabilidad*: el sistema debería asistir a sus habitantes y facilitarles sus tareas automatizando gran parte de sus actividades rutinarias.

- *Tiempos de respuestas adecuados*: el ambiente es esencialmente de tiempo real y el sistema debería ser lo suficientemente reactivo como para realizar sus decisiones en un lapso de tiempo acotado.

La mayoría de estas propiedades son características de los sistemas basados en agentes inteligentes.

Dado que una de las aplicaciones identificadas que actualmente preocupan en demasía a los habitantes de hogares, edificios o empresas es su seguridad personal y la posibilidad de que intrusos ocasionen problemas no deseados, es necesario definir algún sistema de identificación.

Cualquier sistema de identificación debe definir las metas de seguridad apropiadas y los atributos dentro de una política de seguridad. Esta política debe identificar el nivel de seguridad apropiado y conmensurado con el valor de cada bien protegido. Una identificación segura debe ser una interfaz entre una forma de acceso individual y la facilidad de acceder a un sistema deseado. Un servicio creíble necesita proveer autenticidad y validación al mismo tiempo. La identificación, una vez autenticada y validada, deberá contener o referenciar información que es usada para verificar no sólo la identidad del individuo sino también sus permisos.

Los sistemas de autenticación se pueden enmarcar en tres grandes grupos: sistemas basados en *algo conocido* (contraseña), sistemas basados en *algo poseído* (tarjeta inteligente), sistemas de identificación por radiofrecuencias (RFID) y sistemas biométricos (basados en características del individuo). Evidentemente un sistema de autenticación puede combinar varios de estos mecanismos para aumentar el nivel de seguridad,

sobre todo si se usa una red de telecomunicaciones. Además, cualquier sistema de autenticación debe ser viable (es decir, económicamente rentable) y aceptado por los usuarios.

2. LINEAS DE INVESTIGACION y DESARROLLO

En los sistemas basados en algo conocido, la utilización de claves secretas, uso de número de identificación personal, (PIN) y/o tarjetas de identificación, uso de máquinas lectoras de tarjetas o códigos de barra, quizás en algunos casos no son suficientes seguros pero si resulta aceptado por el usuario y viable. La autenticación es el mecanismo más básico, y el primero que existe de protección de un sistema y consiste en comprobar que un usuario es quien dice ser, y comúnmente se basa en nombre de usuario/contraseña.

El tema se ha centrado en la forma en que las personas deben interactuar con el ambiente sin que les ocasione grandes cambios en sus rutinas. Confort, gestión de energía y seguridad son las principales aplicaciones en un edificio inteligente. En este contexto, el LabTEM, ha implementado un prototipo de sistemas de identificación de individuos que provee la posibilidad de identificar a las personas y una vez generada la autenticación habilitar los accesos configurados para ese usuario utilizando para ello el enfoque multi-agente.

La aplicación desarrollada ha sido implementada aplicando tecnologías Orientadas a Objetos (OO) y SMA como base para la resolución de problemas de seguridad y gestión de edificios, obteniendo beneficios tales como saber que personas han ingresado y cuanto tiempo han permanecido. Se ha desarrollado básicamente en un módulo que facilita la gestión de usuarios, combinando técnicas de filtrado cognitivo y colaborativo [12] y un módulo de extracción de información basada en el contenido de páginas Web en bibliotecas digitales. Esta combinación ha permitido una mayor flexibilidad en la escalabilidad del sistema.

El sistema se ha desarrollado a partir de tres componentes; un Program Logical Cotroller (PLC) encargado de conectar los dispositivos de ingreso de identificación (teclados) con el servidor a través de un puerto serie RS-232, el servidor donde se encuentra la aplicación Web encargada de la gestión de los usuarios, dispositivos y zonas de acceso y la aplicación SMA se encarga de gestionar los accesos e ingresos, si un usuario abandona el ambiente sin hacer su registración de salida, el SMA se encarga de claudicar los egresos. Este SMA está compuesto por diversos dispositivos, donde cada uno representa una entidad, pudiéndose definir reglas o comportamientos específicos de acuerdo a la responsabilidad de cada agente. El sistema también permite ver información en tiempo real y generar reportes correspondientes a los movimientos que se han registrado en forma diaria.

Con el sistema desarrollado se ha logrado implementar un sistema de identificación segura, el que permite identificar a las personas y una vez generada la autenticación habilitar los recursos configurados para ese usuario utilizando para ello el enfoque multi-agente, a partir de una herramienta con arquitecturas estándares, flexibles y fáciles de mantener. El mismo es un sistema funcional, que permite controlar el acceso a un ambiente determinado, además con la posibilidad para, agregar, quitar o modificar usuarios, zonas o dispositivos hasta donde el hardware lo permita. Como limitante se reconoce la dependencia del protocolo RS-232 considerado por la distancia física entre el servidor y el PLC.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados obtenidos hasta el momento, en el marco del proyecto aquí presentado han sido: Estudio sobre las nuevas tecnologías capaces de administrar recursos en un edificio.

Investigación en tendencias actuales respecto a la confiabilidad y eficiencia de cada una.

Implementación de un sistema de registración de personas que ingresan/egresan a la oficina del LabTem.

Siguiendo en esta misma línea, el grupo se encuentra abocado al desarrollo de una aplicación que una vez autenticada la persona, se permita generar condiciones ambientales de iluminación y temperatura de acuerdo a sus preferencias, de manera de ofrecer una mejora laboral y un consumo responsable de energía.

4. FORMACION DE RECURSOS HUMANOS

Dos integrantes se encuentran realizando su doctorandos.

Dos integrantes están desarrollando su tesis de maestría .

El LabTEM además cuenta con dos alumnos becarios y dos pasantes y la dirección de trabajos finales de la carrera Ingeniería en Sistemas en temas relacionados con este proyecto

5. BIBLIOGRAFIA

[1] Krainier A, Towards smart buildings. Architectural Assian Graduate School. Environment & Energy Studies Program (1996).

[2] S. Sharples, V. Callaghan, and G. Clarke, "A multi-agent architecture for intelligent building sensing and control," *Int'l Sensor Review Journal*, vol. 19, no. 2, 1999.

[3] Poole D., Marchworth A., Goebel R., "Computational Intelligence – A Logical Approach" Oxford University Press, 1998

[4] Wooldridge M., Jennings N. R., "Intelligent agents: Theory and practice". 1994.

[5] M. Huhns and L. Stephens, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, ch. *Multiagent Systems and Societies of Agents*, pp. 79–120. The MIT Press, 1999.

[6] S. Russell and P. Norvig, *Artificial Intelligence - A Modern Approach*. Prentice Hall, second ed., 2003.

[7] S. Kalenka and N. R. Jennings, *Cognition, Agency and Rationality*, ch. *Socially Responsible Decision Making by Autonomous Agents*, pp. 135–149. Kluwer, 1999.

[8] M. Wooldridge, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, ch. *Intelligent Agents*, pp. 27–78. The MIT Press, 1999.

[9] M. Wooldridge, *An Introduction to MultiAgent Systems*. Chichester, England: John Wiley & Sons, 2002.

[10] M. Wooldridge, *An introduction to multiagent systems*, ch. 4. *Practical Reasoning Agents*. John Wiley and Sons, LTD, 2002.

[11] Errecalde M., Lasso M., Villagra A., Pandolfi D., de San Pedro M., "Edificios Inteligentes: el enfoque multi-agente". WICC 2006.

[12] H.S. Nwana, D.T. Ndumu and L.C. Lee. ZEUS: An advanced Tool-Kit for Engineering Distributed Multi-Agent Systems. In: *Proc. of PAAM98*, pp. 377-391, London, U.K. 1998.

Complejidad de la Programación en Lógica Rebatible^{*}

Laura A. Cecchi^{***}
lcecchi@uncoma.edu.ar

Pablo Fillottrani^{**}
prf@cs.uns.edu.ar

Guillermo Simari^{**}
grs@cs.uns.edu.ar

^{***}Depto. de Ciencias de la Computación - Fa.E.A. ^{**}Depto. de Ciencias e Ing. de la Computación

UNIVERSIDAD NACIONAL DEL COMAHUE

Buenos Aires 1400

(8300)Neuquén - Argentina

UNIVERSIDAD NACIONAL DE SUR

Av. Alem 1253

(8000) Bahía Blanca - Argentina

Contexto

La línea de investigación presentada se enmarca en el Proyecto de Investigación “Técnicas de Inteligencia Computacional para el diseño e implementación de Sistemas Multiagentes” del Grupo de Investigación en Robótica Cognitiva de la Universidad Nacional del Comahue, y en los desarrollos del Laboratorio de Investigación y Desarrollo en Inteligencia Artificial y del Laboratorio de Investigación y Desarrollo en Ingeniería de Software y Sistemas de Información, ambos de la Universidad Nacional del Sur.

Resumen

Dando continuidad al desarrollo de la semántica \mathcal{GS} , se ha planteado estudiar la Complejidad Computacional y Descriptiva de la Programación en Lógica Rebatible a través de la semántica declarativa definida. En este sentido se estudiaron diferentes problemas de decisión y se calculó su complejidad computacional. Asimismo se avanzó en el estudio de la Complejidad de Datos, de los Programas y la Combinada. Este análisis es el punto de partida para el estudio en marcha de la expresividad de la teoría.

^{*} Este trabajo está parcialmente financiado por la Universidad Nacional del Comahue (Proyecto de Investigación “Técnicas de Inteligencia Computacional para el diseño e implementación de Sistemas Multiagentes” (04/E062)) y por la Secretaría General de Ciencia y Tecnología de la Universidad Nacional del Sur.

El propósito de este trabajo es presentar los resultados alcanzados en esta línea de investigación, los desarrollos en progreso y los trabajos a futuro.

PALABRAS CLAVE: Sistemas Argumentativos, Razonamiento Rebatible, Programación en Lógica, Semántica basada en juegos, Complejidad Computacional, Complejidad Descriptiva

1. Introducción

La Complejidad Descriptiva [14] estudia la relación entre los lenguajes formales y los recursos computacionales (espacio y tiempo) requeridos para resolver los problemas formulados en esos lenguajes. Por un lado deseamos comprender cómo el poder expresivo de un sistema lógico, tal como lógicas de primer o de segundo orden o lenguajes de consultas de bases de datos basados en lógica como Datalog, está relacionado con sus propiedades algorítmicas. Y por otro, queremos relacionar los niveles de complejidad computacional con el poder expresivo de los lenguajes formales.

Los resultados en este campo muestran que diferentes clases de complejidad tienen caracterización lógica independiente. Por ejemplo, **P** y **NP**, se corresponden con la lógica de primer orden con la definición de un operador de menor punto fijo [17, 13] y la lógica de segundo

orden existencial[11], respectivamente. Así vinculamos a la Complejidad Computacional de un problema con la riqueza necesaria en un lenguaje para especificar dicho problema.

Una de las aplicaciones más interesantes de las técnicas y resultados alcanzados por la Teoría de la Complejidad Descriptiva es el área de Bases de Datos. La Complejidad Descriptiva da una caracterización precisa de cuáles conceptos son definibles como consultas en un sistema, permitiendo evaluar la dificultad de describir consultas.

La Programación en Lógica Rebatible (P.L.R.) [12] es una extensión de la Programación en Lógica (de ahora en más P.L.), cuya teoría de prueba está basada en el análisis dialéctico de argumentos a favor y en contra de un argumento inicial. La semántica declarativa \mathcal{GS} [6] definida a partir de las semánticas de interacción basadas en juegos[1, 18] caracteriza a dicha teoría de prueba.

Algunos cómputos sobre bases de datos necesitan ir más allá del poder expresivo del lenguaje SQL. Así aparecen lenguajes como Datalog. Del mismo modo, puede la P.L.R. ser analizada como una base de datos, calcular su poder expresivo y caracterizar la clase de consultas que podremos realizar. La motivación principal de este estudio es el interés despertado en los últimos años por los sistemas argumentativos como herramienta para el desarrollo de diversas aplicaciones [7, 8, 2, 9] y su posible expansión a otros campos de aplicación.

En este sentido, se ha abierto una línea de investigación en el Proyecto “Técnicas de Inteligencia Computacional para el diseño e implementación de Sistemas Multiagentes” de la Universidad Nacional del Comahue, que corresponde al tema de tesis doctoral del primer autor de este trabajo y en el que se estudia a la P.L.R. como una estructura lógica finita, con el objeto de determinar su poder expresivo y compararlo con otros sistemas. El propósito de este trabajo es presentar los resultados alcanzados y los trabajos a futuro.

2. Resultados Alcanzados

En una primer etapa de la investigación se estudió la Complejidad Computacional de la P.L.R.. Algunos de los problemas de decisión evaluados fueron:

- Determinar si un literal L pertenece a las consecuencias rigurosas[6] de un conjunto de reglas rebatibles es **P**-completo.[4]
- El problema de decisión “dado un conjunto de reglas rebatibles R , es R un argumento para un literal bajo un programa lógico rebatible” es **P**-completo.[5]
- Sean AP el tiempo polinomial requerido para decidir el problema del item anterior, $|Lit|$ la cantidad de literales en el programa y $|\Delta|$ a cantidad de reglas rebatibles. Entonces la cota superior del tiempo para computar todos los argumentos es $|Lit| * 2^{|\Delta|} * AP$.
- (*Existencia de un Argumento*) El problema de decisión “existe un argumento para un literal L bajo un programa lógico rebatible” está en **NP**.[5]
- El problema de decisión “no existe ningún contraargumento para un argumento bajo un programa lógico rebatible” está en **coNP**.

Ya que la P.L.R. no asume como entrada al conjunto de argumentos, los primeros resultados que han sido establecidos están relacionados con los argumentos, los movimientos del juego. Los resultados obtenidos contrastan fuertemente con muchos de los sistemas argumentativos existentes, ya que estos tienen como lógica subyacente a la Lógica Proposicional (por ejemplo, [16]).

En función de los problemas definidos anteriormente se delinearón dos problemas de decisión que se consideraron relevantes [4]:

- **GAMESAT**: Decidir si existe un juego para un literal α ganado por el proponente en el contexto de un programa lógico rebatible.

- **NOWINGAME:** Decidir si no existe ningún juego para el literal α ni para el complemento de dicho literal que sea ganado por el proponente en el contexto de un programa lógico rebatible.

Asimismo, se comenzó el análisis de la P.L.R. como una base de datos [10, 17], estudiando [5] la Complejidad de los Datos (Data Complexity), la de los Programas (Program Complexity) y la Combinada (Combined Complexity). Estos conceptos evalúan la complejidad de aplicar una consulta a una base de datos considerando como variables de entrada a alguno de los dos parámetros o a ambos. Los resultados obtenidos en cuanto a complejidad computacional bajo los enfoques *Data complexity* y *Combined complexity* pueden encontrarse en [4, 5].

El poder expresivo de la P.L.R. fue acotado inferiormente por la Complejidad de Datos en $\mathbf{NP} = \Sigma_1^1$, que coincide con la clase de propiedades de las estructuras expresables en lógica existencial de segundo orden [11].

3. Conclusiones y Trabajo Futuro

En este trabajo se ha presentado la línea de investigación en la que se estudia la Complejidad Computacional y Descriptiva de la P.L.R. junto con los resultados alcanzados.

La Complejidad Computacional mide la cantidad de recursos computacionales (tiempo y espacio) que son necesarios para computar una consulta, como una función del tamaño de la entrada. La Complejidad Descriptiva analiza la sintaxis de las consultas y determina el poder expresivo de la teoría. Con el objeto de determinar el poder expresivo de la P.L.R., actualmente nos encontramos estudiando a la teoría como un lenguaje de consulta de bases de datos y analizando el conjunto de relaciones que pueden ser computadas a través de una consulta.

Entre nuestros trabajos futuros se encuentra estudiar si los resultados obtenidos sobre Data Complexity pueden ser extendido a otros sistemas argumentativos [3, 15] cuya representación

sea similar a la de la Programación en Lógica y cuya teoría de prueba sea análoga.

Referencias

- [1] Samson Abramsky. Semantics of Interaction: an Introduction to Games Semantics. In A. Pitts and P. Dibyer, editors, *Semantics and Logic Computation*. Cambridge, 1997.
- [2] N. Bassiliades, G. Antoniou, and I. Vlahavas. A defeasible logic reasoner for the semantic web. In *Proc. of the Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 49–64, 2004.
- [3] Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. *Artif. Intell.*, 171(5-6):286–310, 2007.
- [4] Laura A. Cecchi, Pablo R. Fillottrani, and Guillermo R. Simari. An Analysis of the Computational Complexity of DeLP through Game Semantics. In *XI Congreso Argentino de Ciencias de la Computación*, pages 1170–1181, Argentina, Octubre 2005. Universidad Nacional de Entre Ríos.
- [5] Laura A. Cecchi, Pablo R. Fillottrani, and Guillermo R. Simari. On the complexity of DeLP through game semantics. In J. Dix and A. Hunter, editors, *XI International Workshops on Nonmonotonic Reasoning*, pages 386–394, Clausthal University, 2006.
- [6] Laura A. Cecchi and Guillermo R. Simari. Sobre la relación entre la Semántica GS y el Razonamiento Rebatible. In *X CACiC - Universidad Nacional de La Matanza*, pages 1883–1894, San Justo - Pcia. de Buenos Aires, 2004.
- [7] C. Chesñevar and A. Maguitman. An Argumentative Approach to Assessing Natural Language Usage based on the Web

- Corpus. In *Proc. of the European Conference on Artificial Intelligence (ECAI) 2004*, pages 581–585, Valencia, Spain, August 2004.
- [8] C. Chesñevar and A. Maguitman. ARGUENET: An Argument-Based Recommender System for Solving Web Search Queries. In *Proc. of the 2nd IEEE Intl. IS-2004 Conference*, pages 282–287, Varna, Bulgaria, June 2004.
- [9] Carlos I. Chesñevar and Ana G. Maguitman. Combining argumentation and web search technology: Towards a qualitative approach for ranking results. *Intl. Journal of Advanced Computational Intelligence & Intelligent Informatics*, 9(1):53–60, 2005.
- [10] Evgeny Dantsin, Thomas Eiter, Georg Gottlob, and Andrei Voronkov. Complexity and expressive power of logic programming. *ACM Computing Surveys (CSUR)*, 33(3):374 – 425, September 2001.
- [11] Ron Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In R. Karp, editor, *Complexity of Computation. SIAM-AMS Proceedings*, volume 7, pages 43–73, 1974.
- [12] Alejandro J. García and Guillermo R. Simari. Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, 4(1):95–138, 2004.
- [13] Neil Immerman. Relational queries computable in polynomial time. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing, STOC82*, pages 147–152, New York, NY, USA, May 1982. ACM Press.
- [14] Neil Immerman. *Descriptive Complexity*. Springer-Verlag, New York, 1999.
- [15] Henry Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese. Special issue on New Perspectives in Dialogical Logics*, 127:187–219, 2001.
- [16] Iyad Rahwan and Leila Amgoud. An argumentation based approach for practical reasoning. In *AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pages 347–354, New York, NY, USA, 2006. ACM.
- [17] Moshe Y. Vardi. The complexity of relational query languages. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing, STOC82*, pages 137–146, New York, NY, USA, May 1982. ACM Press.
- [18] Peter Wegner. Interactive Foundations of Computing. *Theoretical Computer Science*, Feb 1998.

Aprendizaje de Estructuras de Independencia de Modelos Probabilísticos Gráficos

Facundo Bromberg
fbromberg@frm.utn.edu.ar

Federico Schlüter
fschluter@frm.utn.edu.ar

Laboratorio de Inteligencia Artificial
Departamento de Sistemas de Información,
Facultad Regional Mendoza, Universidad Tecnológica Nacional

RESUMEN

Nuestra investigación se enmarca en el problema del aprendizaje, a partir de datos, de estructuras de independencia de modelos probabilísticos gráficos. Es de especial interés el aprendizaje automatizado de estos modelos a partir de datos, debido principalmente a la presencia cada vez más ubicua de datos digitales. El campo del aprendizaje de máquinas en general, y en particular los miembros de nuestro laboratorio, se han concentrado en el aprendizaje del grafo que representa la estructura de independencias de estos modelos. Durante su tesis doctoral el Dr Bromberg (Bromberg 2007) se ha concentrado en el diseño de algoritmos de aprendizaje de estructuras que utilizan un enfoque basado en independencias (Spirtes et. al. 2000), en contraste con los algoritmos basados en puntaje (Lam and Bacchus 1994, Heckerman 1995). Estos últimos recurren a técnicas para aprendizaje de modelos mas establecidas en la estadística como ser por ejemplo la maximización de la verosimilitud (probabilidad del los datos dado el modelo). El enfoque basado en independencias, en cambio, utiliza un enfoque mas directo para aprender la estructura de independencias del modelo, realizando tests estadísticos de independencia entre las variables aleatorias del sistema. Durante su estadía en Iowa State University, y durante el pasado año ya en UTN-FRM, el Dr. Bromberg ha contribuido con varios algoritmos para el aprendizaje de estructuras de modelos Markovianos con el objetivo de reducir la cantidad de tests estadísticos necesarios durante su ejecución. Recientemente el laboratorio se ha enfocado en un problema más exigente y más importante, el diseño de algoritmos que ante la misma entrada de datos, produzcan modelos de mejor calidad. Estos algoritmos son aplicables tanto a redes Markovianas como Bayesianas.

Palabras clave: *Machine Learning, Probabilistic Graphical Models, Independence-based Structure Learning, Probabilistic Reasoning.*

1. INTRODUCCION

La principal línea de nuestra investigación estudia mecanismos de aprendizaje de estructuras de modelos probabilísticos gráficos, con el objetivo de proponer nuevas perspectivas o enfoques al problema, aportando propuestas que mejoren la calidad y la eficiencia de los algoritmos propuestos hasta la actualidad. El aprendizaje de estructuras se efectúa mediante mecanismos que infieren, a partir de los datos, modelos probabilísticos gráficos (Pearl 1988, Lauritzen 1996). Entre éstos, redes Bayesianas y redes Markovianas son importantes modelos que han contribuido con aumentos considerables en la eficiencia del proceso de razonamiento probabilístico, ofreciendo mecanismos de inferencia estadística prácticos para estimación de valores en diversas ramas de la ciencia.

Las redes Bayesianas han encontrado mayor aplicabilidad en la representación de modelos causales (Pearl, 2000), principalmente debido a la direccionalidad del grafo (Friedman et al., 2000). Las redes Markovianas, en cambio, han sido utilizadas principalmente para representar relaciones espaciales simétricas, por lo que históricamente se las ha llamado *Markov Random Fields* (Kindermann and Snell 1980, Geman and Geman 1984, Besag et. al. 1991). En estas aplicaciones ninguna variable puede identificarse como la causa de las otras. Resaltamos los siguientes ejemplos de campos de aplicación de las redes Markovianas: en visión computacional (Besag et. al. 1991, and Anguelov et. al. 2005) para restaurar imágenes ruidosas, clasificar texturas y segmentar imágenes. Recientemente, la comunidad de Data Mining se ha sumado al uso de estos modelos demostrando gran interés en su uso para mineo espacial de datos, con aplicaciones en geografía, transporte, agricultura, climatología, ecología y otros (Shekhar et al., 2004).

Un modelo gráfico está conformado por un grafo que representa una estructura de independencias entre las variables del modelo (representadas por los nodos del grafo), y por un conjunto de parámetros numéricos para modelar las probabilidades. Por ello, el aprendizaje automatizado de modelos gráficos presenta dos desafíos: primeramente, aprender la estructura de independencias, para luego estimar los

parámetros numéricos (Pearl 1988). Para redes Bayesianas el grafo está conformado por aristas dirigidas, mientras que para redes Markovianas, está conformado por aristas no dirigidas. El propósito principal de estos grafos es codificar las independencias condicionales existentes entre las variables aleatorias del sistema. La identificación explícita de estas independencias es la principal causa en la importante reducción de los requerimientos computacionales y espaciales de la inferencia probabilística, evitando cálculos innecesarios entre variables independientes.

A pesar de ser interesante por sí mismo y ser un problema de gran dificultad, no consideramos en nuestra investigación el problema de aprendizaje de parámetros, concentrándonos en el problema del aprendizaje de estructuras. Tradicionalmente, el aprendizaje del grafo a partir de los datos se realiza por medio del método de maximización de algún "puntaje" (e.g., maximización de la verosimilitud), el cual consiste en una optimización en el espacio de todas las estructuras posibles, identificando aquella con mayor puntaje (Lam and Bacchus 1994, Heckerman 1995). Desde un principio, el aprendizaje de la estructura de los modelos Markovianos, en contraste con los Bayesianos, ha presentado importantes dificultades computacionales, principalmente por el hecho de que el cálculo del puntaje para ciertos modelos requiere estimar los parámetros del mismo, un problema NP-completo (Barahona 1982). Esta dificultad computacional ha forzado a los usuarios de estos modelos a recurrir al conocimiento de expertos para estimar las estructuras gráficas.

Los resultados obtenidos en investigaciones recientes (2009, Gandhi et al. 2008, Margaritis and Bromberg 2009) han producido una serie de algoritmos capaces de aprender eficientemente la estructura de modelos Markovianos basándose en el enfoque de independencias (Spirtes et. al. 2000), cuya ventaja radica en que se puede prescindir del aprendizaje de los parámetros numéricos para aprender la estructura de independencias, eliminando así una de las principales causas de ineficiencia de los algoritmos basados en puntaje. El enfoque de aprendizaje basado en independencias consiste en la ejecución de tests estadísticos para estimar independencias condicionales entre las variables aleatorias de un sistema (e.g., es la variable aleatoria X independiente de la variable aleatoria Y dado el conocimiento de los valores del conjunto de variables aleatorias Z ?). Desafortunadamente, la enorme ventaja computacional obtenida por los algoritmos basados en independencias se ve opacada por la sensibilidad de la calidad de las estructuras producidas por estos algoritmos cuando los datos disponibles son insuficientes.

La calidad de las estructuras resultantes de los algoritmos basados en independencias depende fuertemente de la calidad de los tests estadísticos de independencias llevados a cabo durante su ejecución. Se sabe que estos tests pueden resultar en decisiones erradas cuando la cantidad de datos es insuficiente. Si bien la disponibilidad de datos crece continuamente con la innovación en materia de instrumentación digital, para mantener su calidad los tests requieren una cantidad de datos exponencial en la cantidad de variables involucradas en el test (Agresti 2002).

2. LINEAS DE INVESTIGACION Y DESARROLLO

Por lo expuesto, nuestra investigación se ha orientado recientemente en la búsqueda de enfoques novedosos para atacar la problemática que subyace a la falta de certeza de los resultados de los tests, lo que disminuye la calidad de los modelos generados. Hasta el momento hemos detectado dos enfoques complementarios para la mejora de la calidad de las estructuras: (i) diseñar algoritmos para mejorar la calidad de los tests, (ii) ante un test dado, diseñar algoritmos para mejorar la calidad de las estructuras.

En términos generales, el primer enfoque detecta tests incorrectos y corrige sus valores. Esto es posible explotando restricciones existentes entre los tests de un conjunto de variables. Estas restricciones, llamadas *axiomas de independencia* (Pearl, 1988, Dawid 1979) son fundamentales y son resultado directo de los axiomas de Probabilidad, es decir, solo tests erróneos pueden violarlos. Por ello, nuestro enfoque propone buscar violaciones de los axiomas de independencia para detectar los tests erróneos. Dada una restricción violada, una dificultad importante es la de detectar cual (o cuales) de los tests involucrados en esta restricción es el incorrecto. En general la decisión requiere información de las otras violaciones en donde participa cada test. Lamentablemente existe un número exponencial de restricciones por test (con respecto al número de variables del dominio). En (Bromberg and Margaritis, 2009), se propone con éxito la utilización del formalismo de Argumentación (Dung, 95) para decidir eficientemente si un test es errado o no en base al conjunto de restricciones satisfechas y violadas por el test. Bajo el punto de vista de la Argumentación, vemos al problema como una base de conocimiento inconsistente, compuesta por proposiciones de independencias (e.g., es X independiente de Y dado el conjunto de variables Z ?), reglas lógicas de Pearl que las relacionan, y potenciales inconsistencias donde ciertas proposiciones violan una o más reglas. La importancia del formalismo radica en que formalismos tradicionales de razonamiento lógico no son sólidos cuando la base de conocimiento es

inconsistente. Para refinar aún más el formalismo de inferencia por Argumentación, es posible cuantificar las reglas con preferencias (Amgoud and Cayrol, 2002), un valor numérico que permite decidir entre dos reglas conflictivas (una demuestra la negación de la otra) cuando no existe otra información. En nuestro caso, se tomó como preferencias a la probabilidad de que un test estadístico es correcto (calculable a partir de valores numéricos arrojados por tests estadísticos, ver referencia para mas detalles). De esta manera, el formalismo de Argumentación con preferencias se convierte en una instancia de un formalismo para el razonamiento bajo incertidumbre. Los resultados positivos obtenidos con este formalismo demuestran la viabilidad del enfoque de mejora de la calidad basado en restricciones y propone una importante línea de investigación de nuevos algoritmos. Recientemente Federico Schlüter se ha sumado al laboratorio con beca doctoral de UTN-PRH para trabajar en esta línea de investigación, planeando investigar en el corto plazo el uso de las *Markov Logic Networks* (Richardson 2006) como formalismo de razonamiento bajo incertidumbre para la inferencia del valor de independencia de los tests, basado en restricciones. Este formalismo es prometedor por razonar con reglas en lógica de primer orden, reduciendo la complejidad presentada por el formalismo de Argumentación y cualquier otro que requiera de la proposicionalización de las reglas de Pearl.

El segundo enfoque para la mejora de la calidad considera el diseño de algoritmos de aprendizaje novedosos que tomen en cuenta el hecho de que las independencias arrojadas por los tests de independencia pueden estar errados. Este enfoque es el utilizado por PFMN (Margaritis and Bromberg 2009) y por GSMNS, un algoritmo que generaliza GSMN que se encuentra actualmente en la etapa de diseño de heurísticas para lidiar con su costo computacional exponencial, pero cuyos resultados preliminares demuestran mejoras en la calidad de los modelos generados altamente promisorias (ver resultados en Sección 3). Pasamos a explicar el algoritmo GSMNS en detalle en la siguiente sección.

2.1. ALGORITMO GSMNS

Dado un conjunto de variables aleatorias V distribuidas de acuerdo a una distribución \Pr (desconocida), y una muestra estadística de asignaciones conjuntas de las variables V distribuidas de acuerdo a \Pr , llamado simplemente conjunto de datos D , el algoritmo GSMNS (Grow Shrink Markov Network with Search), arroja como resultado un grafo no-dirigido G que representa la estructura de independencias de una red Markoviana M . Cuando los tests son correctos, lo ocurre cuando la muestra de datos es correctamente sampleada de

\Pr y es infinita, es posible demostrar que la distribución de probabilidad representada por la red aprendida M coincide con la distribución subyacente \Pr . En la práctica esta condición rara vez se cumple, produciendo redes incorrectas.

El algoritmo GSMNS al igual que el algoritmo GSMN (Grow Shrink Markov Network) al cual generaliza, aprende la estructura de independencias G infiriendo el *Markov blanket* $MB(X)$ de cada variable X en V . De acuerdo a (Pearl 88), $MB(X)$ consiste en un subconjunto de $V - \{X\}$ de variables aleatorias, vecinas de X en la estructura. En otras palabras, aprender el MB de cada variable en V es suficiente para aprender la estructura. En (Dimitris and Thrun 2000) se presenta el algoritmo GS (Grow Shrink) que dado una variable aleatoria X , y el conjunto de datos D y de variables V , aprende $MB(X)$ realizando tests estadísticos de independencia en dos etapas:

Grow Phase:

for all Y in $V - \{X\}$, if $\neg I(X, Y | S)$ then $S \leftarrow S \cup \{Y\}$

Shrink Phase:

for all Y in S , if $I(X, Y | S - \{Y\})$ then $S \leftarrow S - \{Y\}$

donde $I(X, Y | S)$ indica si la variable X es independiente de la variable Y condicionado en el conjunto de variable S , de acuerdo al test de independencia.

El algoritmo GSMN utiliza una versión del algoritmo GS que preestablece el ordenamiento de variables para recorrer los bucles de ambas fases por medio de una heurística sencilla. Este algoritmo confía plenamente en el resultado de los tests (potencialmente de baja calidad o hasta erróneos), dando indicios de un posible error. Esto produce un algoritmo eficiente (cuadrático en el número de variables) pero que ante falta de datos, y por ende potenciales errores en los tests, el ordenamiento utilizado y la confianza ciega en las independencias obtenidas puede resultar en una estructura errónea.

Por esto proponemos una generalización de GS, el algoritmo GSS, que flexibiliza tanto el ordenamiento de variables, como la asignación de valores de independencias de los tests. Para ello proponemos primero una medida de calidad de un *Markov blanket*, calculada a partir de la medida de calidad de los tests utilizados para aprenderlo (ver apartado 2.1.1), y segundo, la realización de una búsqueda sistemática de todos los posibles ordenamientos y todas las posibles asignaciones (ver apartado 2.1.2.), devolviendo como resultado el *blanket* que maximice la medida de calidad. Por el momento la búsqueda se realiza en la etapa de *grow*, esperando diseñar e implementar una búsqueda en la etapa de *shrink* en el mediano plazo.

2.1.1. Medida de calidad de un Markov blanket

Dado un triplete $(X, Y | \mathbf{Z})$, un test estadístico arroja, además de un valor de independencia, un valor numérico que indica la “calidad” del test. En nuestros experimentos utilizamos un test de independencia Bayesiano propuesto por (Margaritis and Bromberg 2009) que calcula la probabilidad de independencia/dependencia. Si denotamos por $\Pr(X, Y | \mathbf{Z})$ la probabilidad de independencia entre X e Y dado \mathbf{Z} , proponemos calcular la medida de calidad del Markov blanket $\mathbf{MB}(X)$, denotada $\Pr(\mathbf{MB}(X))$, utilizando el conjunto de tests $T(\mathbf{MB}(X))$ realizados durante la etapa de grow de la siguiente manera:

$$\Pr(\mathbf{MB}(X)) = \text{Prod}_{t \in T(\mathbf{MB}(X))} \Pr(t),$$

es decir, como la productoria de las probabilidades de independencia de los tests realizados durante la etapa de grow.

2.1.2. Búsqueda de ordenamiento y asignación óptima

Para encontrar el ordenamiento de $V-\{X\}$ y asignación de independencias que maximicen la medida de calidad $\Pr(\mathbf{MB}(X))$ del blanket de X , proponemos realizar una búsqueda en árboles del tipo A^* (Russell and Norvig, 2002), a través de los distintos estados de la etapa de grow, es decir, el estado de \mathbf{S} y las variables en $V-\{X\}$ ya testeadas, que denotamos \mathbf{P} (por predecesores). Los operadores entre estados son los tests estadísticos que pueden realizarse en ese estado, y el costo de cada operador asociado al test t es $-\log(\Pr(t))$. De esta manera maximizar la calidad de un blanket equivale a minimizar la sumatoria de los costos. Para explorar todos los ordenamientos y todas las asignaciones, proponemos como operadores de un estado $[\mathbf{P}, \mathbf{S}]$ aquellos que nos lleven a los siguientes estados sucesores:

$$\{[\mathbf{P} \square \{U\}, \mathbf{S}] \text{ y } [\mathbf{P} \square \{U\}, \mathbf{S} \square \{U\}] \mid U \square \mathbf{P}\},$$

donde el primer caso corresponde a una asignación de independencia (i.e., $I(X, U | \mathbf{S})$) y el segundo a una asignación de dependencia (i.e., $\neg I(X, U | \mathbf{S})$).

De esta manera, la optimalidad del costo de camino de la solución garantizada por A^* equivale a una maximización de la productoria de la calidad de los tests. En otras palabras, a una calidad del *blanket* máxima. Para completar la descripción de la búsqueda A^* requerimos proponer una heurística que estime, para cada nodo del árbol de búsqueda, el costo de camino óptimo hasta el objetivo, i.e., la calidad óptima de los tests no realizados aún. En nuestros experimentos consideramos como

heurística la probabilidad de los tests que restan calculada en forma voraz (eligiendo para expandir directamente el mejor de entre los hijos de un nodo). Esto garantiza un cálculo eficiente de la heurística.

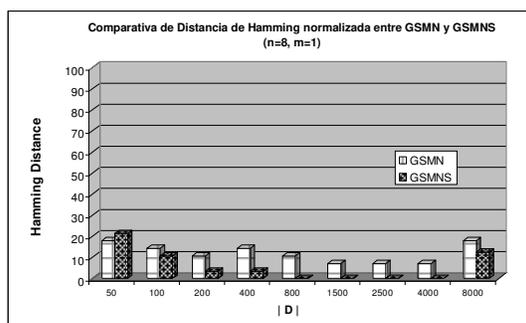
3. RESULTADOS OBTENIDOS

La optimalidad garantizada por las búsquedas en árboles conllevan un costo en tiempo y en memoria exponenciales, salvo para heurísticas muy buenas cuya estimación del costo restante es logarítmicamente cercano al costo real. Esta condición es extremadamente difícil de satisfacer en la práctica y por ello no esperamos, en nuestros experimentos, complejidades espaciales y temporales polinomiales. Nos interesamos, sin embargo, en poder estimar a través de estos experimentos una cota superior (por ser óptimos los algoritmos) en la mejora en la calidad a la cual podemos aspirar. Esperamos en un futuro encontrar heurísticas, o algoritmos de optimización alternativos eficientes, que no reduzcan en demasía las mejoras en la calidad obtenidas con el algoritmo óptimo presentado aquí.

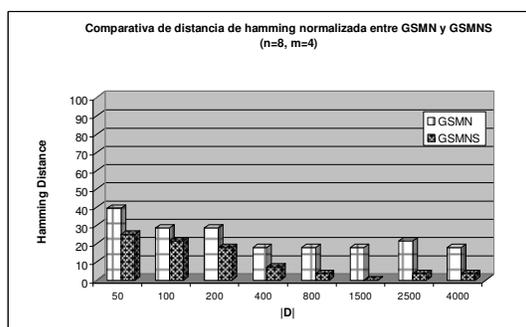
Para evaluar la performance del algoritmo GSMNS comparamos la calidad de las redes que produce con la calidad de las redes producidas por GSMN. Para ello, usamos diversos datasets generados a través de algoritmo de sampleo (Gibbs sampler) a partir de redes Markovianas generadas aleatoriamente. De esta manera conocemos la red subyacente posibilitando una comparación precisa entre el modelo aprendido y el real. Se generaron redes aleatorias de $n=8$ variables, uniendo cada nodo de la red con m variables seleccionadas aleatoria y uniformemente de entre los nodos restantes. Se presentan resultados para $m=1,2,4$. Para estimar la calidad de la red generada medimos la *distancia de Hamming normalizada* entre la red generada y la red real. La distancia de Hamming entre dos redes consiste en la cantidad de aristas existentes en una de las redes e inexistentes en la otra (y viceversa). Se normaliza dividiendo por la cantidad total de pares de variables (i.e., el valor máximo de la distancia de Hamming no normalizada), obteniendo un valor entre 0 y 1. De esta manera, una distancia igual a 0 indica que las dos redes son iguales, mientras que un valor igual a 1 indica que son una el negativo de la otra (siempre que una tiene una arista la otra no).

Dado que la calidad de los tests dependen de la cantidad de datos en el data set, corrimos el algoritmo para subconjuntos de D con un número $|D|$ creciente de renglones, también sampleados aleatoriamente del data set original. Las figuras 1, 2 y 3 comparan la distancia de Hamming normalizada entre la red obtenida utilizando GSMN y GSMNS, y la red real, para los tres valores de m . Puede observarse que en todos los casos la distancia de

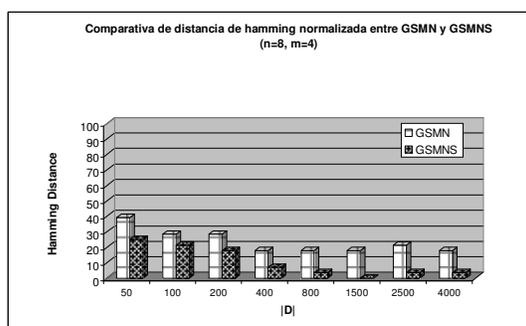
Hamming disminuye para $|D|$ crecientes tanto para GSMN como para GSMNS, lo cual es esperado ya que la calidad de los tests crece con $|D|$. Puede notarse sin embargo que la distancia de Hamming de GSMNS disminuye más rápido que la de su contraparte GSMN, demostrando una mejora en la calidad del primero vs. el segundo. Actualmente los resultados obtenidos para el mismo son prometedores, ya que registran mejores estructuras de independencia, pero aún debemos continuar trabajando sobre la performance, ya que el procesamiento de la solución requiere un tiempo exponencial respecto de la cantidad de variables de la estructura.



-Figura 1: Comparación de la distancia de Hamming normalizada de las redes obtenidas por GSMN y GSMNS para $n=8$, $m=1$, para tamaños crecientes del dataset de entrada.



-Figura 2: Comparación de la distancia de Hamming normalizada de las redes obtenidas por GSMN y GSMNS para $n=8$, $m=2$, para tamaños crecientes del dataset de entrada.



-Figura 3: Comparación de la distancia de Hamming normalizada de las redes obtenidas por GSMN y GSMNS para $n=8$, $m=4$, para tamaños crecientes del dataset de entrada.

4. FORMACION DE RECURSOS HUMANOS Y LINEAS DE INVESTIGACIÓN EMERGENTES

El laboratorio de Inteligencia Artificial en la FRM-UTN ha sido creado a comienzos del 2008 por el Dr. Bromberg tras su incorporación a la Universidad como docente-investigador. Proyectos en curso son:

Aprendizaje automatizado de estructuras de independencias de modelos probabilísticos: En esta investigación colaboran Dr. Bromberg (beca de Fundación UTN y beca post doctoral de reinscripción de CONICET), el Dr. D. Margaritis, profesor asociado en la Iowa State University y ex-director de tesis del Dr. Bromberg, Ing. Federico Schlüter, doctorando bajo la dirección del Dr. Bromberg, financiado con beca UTN, bajo el marco del PRH-2007, y por último Ing. Franco Farinelli (ad-honorem).

GEARS: (<http://ai.frm.utn.edu.ar/gears>) Producto del interés de un grupo de alumnos el laboratorio esta explorando líneas de investigación en el área de la Web Semántica, para lo cual creo el grupo GEARS (Grupo de Estudio de Aplicaciones de Redes Semánticas) en el que participan alumnos de quinto año de la carrera de sistemas A. Edera, M. Garsiolo y M. Pasquier. Este grupo se encuentra desarrollando la aplicación KHIPU para autenticación semántica en el marco de su proyecto final de carrera.

ByOS: (<http://ai.frm.utn.edu.ar/byos>) Búsqueda y Optimización Subjetiva. F. Bromberg, junto con los estudiantes de quinto año de Ing. en Sistemas M. Spertino y S. Perez están investigando técnicas de optimización interactivas basadas en algoritmos genéticos.

De los mencionados arriba, A. Edera, M. Spertino y S. Perez han demostrado interés en continuar con sus estudios realizando un doctorado bajo la dirección del Dr. Bromberg, con intenciones de presentarse a la próxima convocatoria de becas doctorales de CONICET

5. BIBLIOGRAFIA

Agresti, A. (2002). *Categorical Data Analysis*. Wiley, 2nd edition.
 Amgoud and Cayrol (2002). *A reasoning model based on the production of acceptable arguments*. *Annals of Mathematics and Artificial Intelligence*, 34:197-215.
 Angelov, D., Taskar, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G., and Ng, A. (2005). *Discriminative learning of Markov random*

- fields for segmentation of 3D range data. Proc. of CVPR.
- Barahona, F.** (1982). *On the computational complexity of ising spin glass models*. Journal of Physics A: Mathematical and General, **15**(19): 3241-3253.
- Besag, J., York, J., and Mollie, A.** (1991). *Bayesian image restoration with two applications in spatial statistics*. Annals of the Institute of Statistical Mathematics, 43:1-59.
- Bromberg, F., Margaritis, D. and Honavar, V.** (2006). *Efficient Markov Network Structure Discovery using Independence Tests*. Proceedings of SIAM Data Mining 2006. Bethesda, Maryland.
- Bromberg, F.** (2007). *Markov network structure discovery using independence tests*. PhD Thesis, Iowa State University.
- Bromberg, F. and Margaritis, D.** (2009). *Improving the Reliability of Causal Discovery from Small Data Sets using the Argumentation Framework*. Journal of Machine Learning Research, **10**(Feb):301—340.
- Dawid, A. P.** (1979). *Conditional independence in statistical theory*. Jnal. Royal Stat. Society, **41**:1-31.
- Dung, P. M.** (1995). *On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games*. Artificial Intelligence, **77**:321-357.
- Friedman, N., Linial, M., Nachman, I., and Peér, D.** (2000). *Using Bayesian networks to analyze expression data*. Computational Biology, **7**:601-620.
- Gandhi P., Bromberg F. and Margaritis D.** (2008) *Dynamic inference based learning Markov Network Structure*. Proc. of SIAM Data Mining 2008, Atlanta Georgia.
- Heckerman, D., Geiger, D., and Chickering, D. M.** (1995). *Learning Bayesian networks: The combination of knowledge and statistical data*. Machine Learning. **20**:197-243, 1995.
- Kindermann, R. and Snell, J. L.** (1980). *Markov Random Fields and Their Applications*. AMS.
- Lam, W. and Bacchus, F.** (1994). *Learning Bayesian belief networks: an approach based on the MDL principle*. Comp. Intelligence, **10**:269-293.
- Lauritzen, S. L.** (1996). *Graphical Models*. Oxford Statistical Science Series, vol 17. Oxford U. Press.
- Bromberg F. and Maragaritis D.** (2009) *Efficient Markov Network Discovery Using Particle Filters*. This is a preprint of an Article accepted for publication in *Computational Intelligence*.
- Margaritis, D. and Thrun, S.** (2000). *Bayesian network induction via local neighborhoods*. In Solla, S., Leen, T., and Müller, K.R., editors, Advances in Neural information Processing Systems 12, pp. 505-511. MIT Press.
- Pearl, J.** (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*.
- Pearl, J.** (2000). *Causality*. Cambridge Univ. Press.
- Russel, S. and Norvig, P.** (2002). *Artificial Intelligence. A Modern Approach*. Prentice Hall 2nd Edition.
- S. Shekhar, P. Zhang, Y. Huang, and R. Vatsavai.** (2004) Trends in spatial data mining. In H. Kargupta, A. et. al. editors, Data Mining: Next Generation Challenges and Future Directions, **19**, 357–379. AAAI Press / The MIT Press, 2004.
- Richardson R., Domingos, P.** (2006) *Markov Logic Networks*. *Machine Learning*, **62** (2006), pp 107-136.
- Spirtes, P., Glymour, C., and Scheines, R.** (2000). *Causation, Prediction, and Search*. Adaptive Computation and Machine Learning Series. MIT Press, 2nd edition.

OPTIMIZACIÓN DE FUNCIONES MONO-OBJETIVO CON Y SIN RESTRICCIONES, Y FUNCIONES MULTI-OBJETIVO A TRAVÉS DE HEURÍSTICAS BIO-INSPIRADAS

Victoria Aragón[†], Leticia Cagnina[†], Susana Esquivel[†]

[†]Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)
Facultad de Ciencias Físicas, Matemáticas y Naturales - Universidad Nacional de San Luis
Ejército de los Andes 950 - Local 106 - (D5700HHW) - San Luis - Argentina
Tel: (02652) 420823 / Fax: (02652) 430224 - e-mail: {vsaragon, lcagnina, esquivel}@unsl.edu.ar

Resumen

En esta presentación se describen, en forma breve, algunas de las direcciones de investigación que en la actualidad se están desarrollando dentro de la línea "Optimización Mono y Multiobjetivo" del Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC). Uno de los objetivos de esta línea, es el estudio y desarrollo de metaheurísticas aptas para resolver problemas de optimización. En particular, el énfasis está puesto en las heurísticas de inteligencia computacional basadas en los paradigmas de inteligencia colectiva y biológicos, como ser *Particle Swarm Optimization* y Sistemas Inmunes Artificiales.

1. INTRODUCCIÓN

En la actualidad la línea de investigación focaliza su trabajo sobre algunas heurísticas bio-inspiradas, en particular: Sistemas Inmunes Artificiales (SIA) y Particle Swarm Optimizers (PSO), atacando los siguientes problemas: optimización numérica global (SIA), optimización de funciones con restricciones (SIA, PSO), optimización de funciones en ambientes no estacionarios (SIA) y optimización multiobjetivo (PSO).

La optimización de problemas ha sido y es, un área activa de investigación. A medida que los problemas de mundo real, ya sea en el campo de la ciencia o la ingeniería, se vuelven más complejos, es necesario contar con heurísticas capaces de resolver rápida y eficientemente estos problemas. En las últimas décadas algoritmos bio-inspirados han llamado la atención de los investigadores. Uno de ellos es la heurística *Particle Swarm Optimization*, la cual está basada en el comportamiento que muestran ciertas especies como aves, las cuales utilizan la inteligencia colectiva para sacar provecho de actividades realizadas previamente

por algún miembro de la comunidad. En la heurística todos los individuos de la población siguen la trayectoria del mejor integrante del grupo, y también la memoria de su mejor comportamiento previo. De forma análoga sucede con la heurística *Sistema Inmune Natural* y su poderosa capacidad de procesamiento de información [19]. El Sistema Inmune (SI) es un sistema muy complejo con varios mecanismos de defensa contra organismos patógenos. El principal propósito del SI es reconocer a todas las células dentro del cuerpo y categorizarlas con el objeto de inducir un mecanismo de defensa apropiado. EL SI aprende a través de la evolución para poder distinguir entre los antígenos externos peligrosos de las células propias del cuerpo. Un buen conocimiento de cómo trabaja el sistema inmune es una excelente motivación para desarrollar un Sistema Inmune Artificial (SIA) para tratar problemas de optimización, ya sean global, con restricciones, incluyendo problemas de ingeniería del mundo real o funciones en ambientes no estacionarios. Cuando este modelo computacional es comparado con otros modelos bien establecidos, de la inteligencia computacional, como los algoritmos evolutivos, las colonias de hormigas, las redes neuronales, entre otras, puede decirse que los SIAs están aún en los primeros años de su infancia. Con respecto a la aplicación de SIAs en optimización de funciones con restricciones, poco es el trabajo reportado (ver [8]).

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

2.1. Particle Swarm Optimization

En [2] se introdujo el algoritmo *CPSO-shake*, una versión mejorada de *Constrained Particle Swarm Optimization* para optimización de funciones numéricas con restricciones. El algoritmo utiliza un método simple pa-

ra manejar restricciones de diversos tipos: lineales, no-lineales, de igualdad y de desigualdad. Además esta versión incorpora un mecanismo de escape de óptimos locales (*shake*) a través del movimiento de las partículas a nuevas zonas del espacio de búsqueda. También se utiliza una población dual, en vez de una única población que evoluciona en el proceso de búsqueda, para evitar la convergencia prematura. Los detalles de implementación pueden consultarse en [6].

La performance del algoritmo *CPSO-shake* permitió la utilización del mismo en la resolución de problemas más complejos como son los de ingeniería: problema del diseño óptimo de armaduras [17] (*trusses* en inglés), y de diseño de piezas: *welded beam* [20], *pressure vessel* [23], *speed reducer* [14] y, *tension/compression spring* [4]. Estos dos tipos de problemas de ingeniería son difíciles de resolver, como todo problema del mundo real. Por ello, y para finalizar con la parte de optimización de funciones mono-objetivo, se aplicó el mismo algoritmo en la resolución del problema de despacho de cargas eléctricas. Este problema presenta costosas restricciones a satisfacer, además de la restricción de tiempo que es casi fundamental en la solución que debe arrojar el algoritmo. Esto significa que se necesita un algoritmo que sea capaz de resolver bien el problema, pero además de ello, en forma muy rápida. Al algoritmo *CPSO-shake* se le realizó una pequeña modificación: se emplea la población completa en vez de las dos poblaciones que evolucionan en paralelo. Esto es necesario ya que el algoritmo debe arrojar una respuesta en un tiempo mínimo y utilizando la menor cantidad posible de iteraciones. Si se utilizara la población dividida, ambas subpoblaciones necesitarían mayor cantidad de ciclos para obtener una respuesta óptima.

Habiendo probado la heurística PSO con un buen número de problemas mono-objetivo, y habiendo concluido en la buena performance de la misma, se decidió realizar la adaptación de la misma para el caso multiobjetivo. Si bien los problemas multiobjetivos son muy difíciles de resolver, actualmente se está trabajando en algunas aproximaciones que permitirán resolver este tipo de problemas de manera eficiente.

La próxima sección muestra los resultados obtenidos para cada caso experimental.

2.2. Sistema Inmune Artificial

El propósito de esta investigación es explorar las capacidades de búsqueda de un nuevo modelo propuesto de SIA, llamado TCELL, en el contexto de problemas de optimización global, problemas de optimización con restricciones, incluyendo funciones de ingeniería, y optimización de funciones en ambientes no estacionarios. TCELL está basado en las características y procesos que sufren las células T del sistema inmune de los vertebrados.

TCELL es el primer modelo computacional que emplea la filogenia de las células T. Este modelo está com-

puesto por una fase reconocedora y una fase efectora. La primera involucra a las células T vígenes mientras que en la segunda están involucradas las células efectoras y de memoria. La fase de reconocimiento debe proveer dos poblaciones de células efectoras con la diversidad suficiente para que en la próxima fase se pueda encontrar una célula que optimice el problema. Mientras que, la fase efectora es efectivamente la encargada de encontrar, a través de diferentes etapas, dicha célula. TCELL toma en consideración muchos de los procesos que afectan a las células T desde su origen hasta que se transforman en células de memoria, como ser la selección positiva y negativa de las células T vígenes y los procesos de proliferación y distintos tipos de diferenciación para las células efectoras y de memoria. Para realizar estos dos últimos procesos TCELL utiliza señales co-estimulantes, llamadas citocinas.

La premisa de TCELL es que las células T sólo reaccionan ante la presencia de un antígeno junto con señales co-estimulantes, a través del desencadenamiento de una serie de acciones, estas acciones son influenciadas por una serie de señales emitidas por las mismas células T (citocinas). Es decir, las señales determinan el nivel con el cual se desencadenan las acciones: proliferación y diferenciación de las células T.

El modelo trabaja sobre cuatro poblaciones: células vírgenes (CV), células efectoras con denominador de grupo 4 (CD4), células efectoras con denominador de grupo 8 (CD8), este denominador determina las propiedades de las células, y células de memoria (CM). Cada población está compuesta por un conjunto de células T, cuya característica está sujeta a la población que pertenece. Los detalles de implementación se citan en [3].

El pseudo-código de TCELL para problemas de optimización global se presenta en el Algoritmo 1.

3. RESULTADOS OBTENIDOS

3.1. Particle Swarm Optimization

CPSO-shake fue evaluado con el *benchmark* extendido presentado en [15], el cual cuenta con 20 funciones de prueba con diferentes características. Los resultados obtenidos fueron contrastados con el algoritmo *Stochastic Ranking* (SR) [22] el cual sigue representando uno de los mejores en optimización de funciones con restricciones, así como también con una aproximación de una estrategia evolutiva (AESSR) [21] que utiliza también *Stochastic Ranking*.

Los resultados alcanzados con la presente versión de PSO, demostraron ser muy competitivos con respecto a SR y a AESSR. Las figuras 1 y 2 muestran el error promedio de cada algoritmo con respecto al *benchmark*, para cada una de las 20 funciones. Aunque la variabilidad de resultados obtenidos por SR y AESSR fue levemente inferior a la de nuestro algoritmo, argumentamos que esto está relacionado al mecanismo de escape de óptimos (*shake*) implementado para mantener

Algoritmo 1 Pseudo-código de TCELL para problemas de optimización global

```

1: Inicializar_CV();
2: Evaluar_CV();
3: Asignar_Proliferacion();
4: Activar_CDs_Fase_II();
5: Dividir_CDs();
6: Seleccion_Positiva();
7: Seleccion_Negativa_CD4();
8: Seleccion_Negativa_CD8();
9: while un nro determinado de evaluaciones do
10:   while un numero determinado de veces do
11:     Activar_CD4();
12:   end while
13:   Ordenar_CD4();
14:   Comunicacion_CD4_CD8();
15:   while un nro determinado de veces do
16:     Activar_CD8();
17:   end while
18:   Ordenar_CD8();
19:   Insertar_CDs_en_CM();
20:   while un nro determinado de veces do
21:     Activar_CM();
22:   end while
23:   Ordenar_CM();
24:   Estadísticas();
25: end while

```

diversidad en el algoritmo. En [6] se concluyó que *CPSO-shake* es una de las aproximaciones PSO más competitivas en cuanto a técnicas de manejo de restricciones.

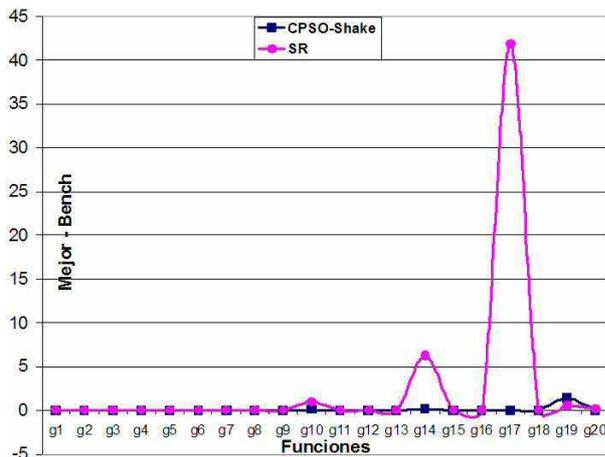


Figura 1: Performance de *CPSO-shake* versus SR

Para las funciones de armaduras se obtuvieron los resultados que se muestran en la tabla 1. Notar que los mejores conocidos hasta ese momento eran los obtenidos en [17]. Nuestro algoritmo PSO [5] obtuvo una leve mejora del óptimo para la función 10-bar, una mejora significativa para 25-bar y una mejora considerable en

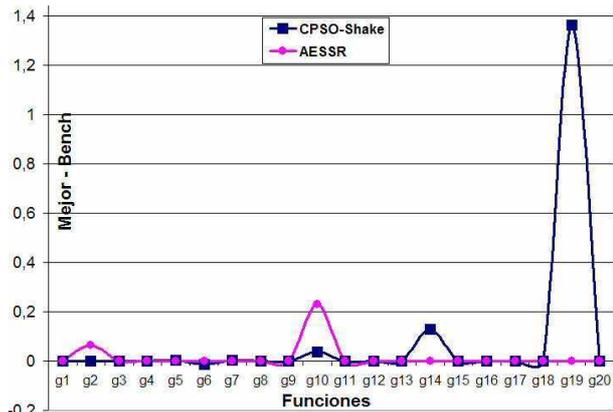


Figura 2: Performance de *CPSO-shake* versus A E

200-bar. Esto demuestra el buen poder exploratorio que ejerce la heurística que concluyó con la mejora de los óptimos conocidos.

Con respecto a los resultados obtenidos para las funciones de diseño de piezas [7], cabe destacar que se compararon los resultados de *CPSO-shake* con los de una aproximación PSO muy elaborada que incluye complejos operadores para evitar convergencia prematura, denominada *COPSO* [1]. También se utilizó en esta sintética comparación, un algoritmo evolutivo (Mezura) [18] que utiliza complejos operadores para manejar las restricciones de los problemas. Los resultados son mostrados en la tabla 2. Se puede observar que los tres algoritmos probados obtuvieron los óptimos en E01, *COPSO* y nuestro algoritmo alcanzaron los mejores conocidos en E02, E03 y E04, mientras que los valores encontrados por Mezura presentan una leve desviación con respecto al óptimo. Esto concluye que *CPSO-shake* presenta un buen funcionamiento y comparable con aproximaciones que utilizan operadores más complejos.

El problema de despacho de cargas puede ser consultado en [24], en donde se presenta un algoritmo PSO que resuelve 3 instancias del problema: Caso A: despacho para 3 generadores con polución, Caso B: despacho para 3 generadores son polución y Caso C: despacho de cargas para 40 generadores sin polución. *CPSO-shake* alcanza los óptimos conocidos para cada instancia, al igual que el algoritmo de [24], pero el aporte de nuestro algoritmo tiene que ver con el tiempo de ejecución insumido para obtener dichos valores. La figura 3 muestran los tiempos de cada algoritmo. Como puede notarse, *CPSO-shake* mejora notablemente el tiempo de respuesta del PSO de referencia, convirtiéndolo en un algoritmo más adecuado para ser utilizado en este tipo de problemas.

3.2. Sistema Inmune Artificial

TCELL fue validado con un conjunto de 23 funciones de prueba tomado de la literatura especializada [26] y comparados con diferentes enfoques bio-inspirados

Tabla 1: Comparación de los mejores valores obtenidos con CPSO-shake, con respecto a los mejores conocidos.

Truss	Mejores Conocidos			CPSO-shake		
	Opt	Std.Dev	Worst	Opt	Std.Dev	Worst
10-bar	4,656.39	0.18	4,656.71	4,656.36	2.84	4,696.06
25-bar	485.33	-	534.84	467.30	0.35	470.87
200-bar	20,319.58	2,492.24	30,269.49	16,468.82	1,476.00	20,373.6719

Tabla 2: Mejores valores alcanzados con COPSO, Mezura and CPSO-shake.

Funciones	Mejores Conocidos	CPSO-skake	COPSO	Mezura
E01	1.724852	1.724852	1.724852	1.724852
E02	6,059.714335	6,059.714335	6,059.714335	6,059.7143
E03	-	2,996.348165	2,996.372448	2,996.348094
E04	0.012665	0.012665	0.012665	0.012689

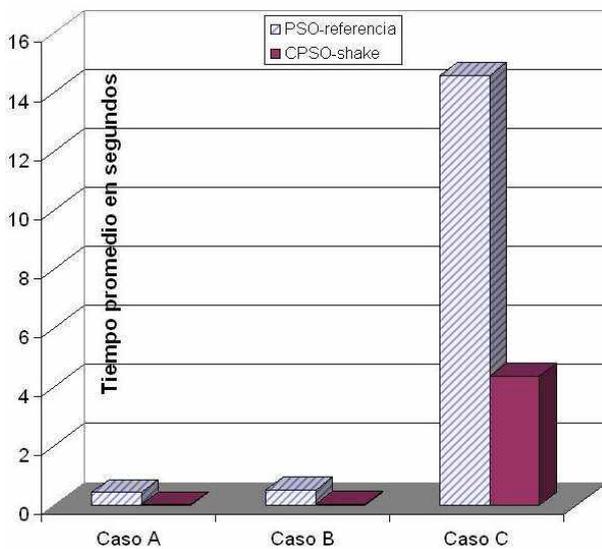


Figura 3: Tiempo insumido por *CPSO-shake* versus PSO [24] para el despacho de cargas.

[3]. Los enfoques considerados fueron: un algoritmo de Evolución Diferencial (DE) [16], un Particle Swarm Optimizer (PSO) [16], un Algoritmo Evolutivo (SEA) [16], un Algoritmo inmune para problemas de optimización global continuos (OPT-IA) [25] y una versión mejorada de este último (OPT-IMM ALG) [9]. Las figuras 4, 5 y 6 muestran los valores medios obtenidos por cada enfoque para cada función.

En la actualidad se están realizando los experimentos correspondientes para los problemas de optimización restringidos y dinámicos. Versiones preliminares del modelo propuesto fueron validadas para el primer tipo de problema en [11, 10, 13, 12].

4. Reconocimientos

El LIDIC reconoce el constante soporte brindado por la Universidad Nacional de San Luis y la ANPYCIT que financian sus actuales investigaciones.

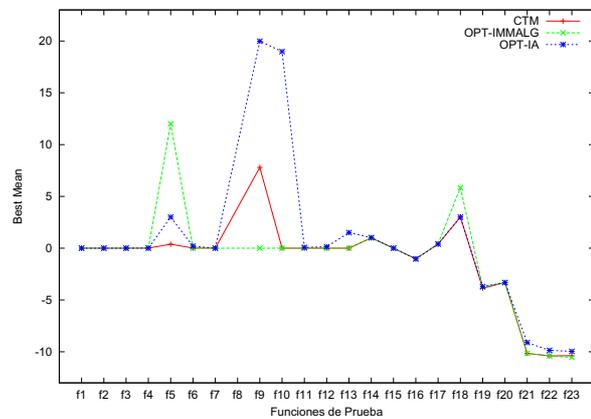


Figura 4: Resultados - Valores Medios Obtenidos por los SIA's

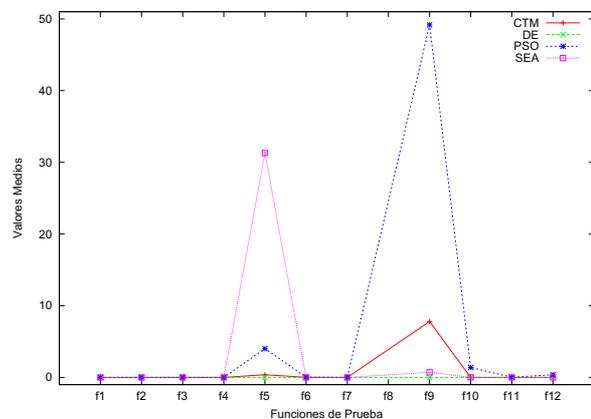


Figura 5: Resultados - Valores Medios Obtenidos por CTM, DE, PSO y SEA

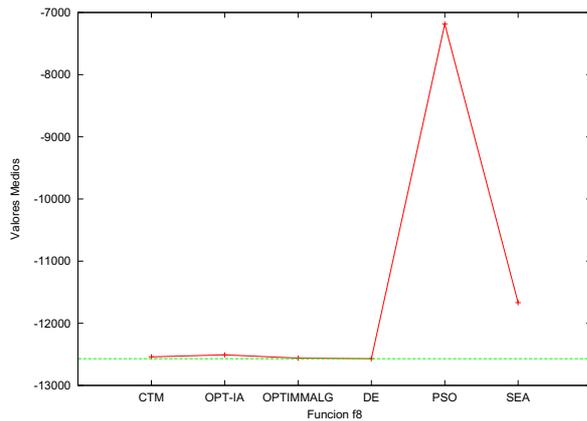


Figura 6: Resultados - Valores Medios Obtenidos por todos los enfoques para f_8

Referencias

- [1] E. Villa Díharce A. Hernandez Aguirre, A. Muñoz Zavala and S. Botello Rionda. Copso: Constrained optimization via pso algorithm. *Center for Research in Mathematics (CIMAT)*.
- [2] Victoria Aragón, Leticia Cagnina, Claudia Gatica, and Susana Esquivel. Metaheurísticas basadas en inteligencia computacional aplicadas a la resolución de problemas de optimización numérica con y sin restricciones y optimización combinatoria. In *IX Workshop de Investigadores en Ciencias de la Computación (WICC 2007)*, pages 124–129, Trelew, Chubut, Argentina, 2007.
- [3] Victoria Aragon, Susana Esquivel, and Carlos Coello Coello. Artificial immune system for solving global optimization problems. In *Congreso Argentino en Ciencias de la Computacion (CACIC08)*, 2008.
- [4] A. Belegundu. A study of mathematical programming methods for structural optimization. *PhD thesis, Department of Civil Environmental Engineering, University of Iowa, Iowa*, 1(1).
- [5] Leticia C. Cagnina, Susana C. Esquivel, and Carlos A. Coello Coello. Solving constrained optimization problems with a particle swarm optimization algorithm. *Artículo enviado (actualmente en proceso de revisión) a: Computer Methods in Applied Mechanics and Engineering (CMAME)*, X(X).
- [6] Leticia C. Cagnina, Susana C. Esquivel, and Carlos A. Coello Coello. A bi-population pso with a shake-mechanism for solving constrained numerical optimization. In *Congress on Evolutionary Computation - CEC 2007*, pages 670–676, Singapur, 2007.
- [7] Leticia C. Cagnina, Susana C. Esquivel, and Carlos A. Coello Coello. Solving engineering optimization problems with the simple constrained pso. *Informatica, International Journal of Computing and Informatics - ISSN 0350-5596*, 32(3):319–326, 2008.
- [8] Nareli Cruz Cortés, Daniel Trejo-Pérez, and Carlos A. Coello Coello. Handling constrained in global optimization using artificial immune system. In Christian Jacob, Marcín L. Pilat, Peter J. Bentley, and Jonathan Timmis, editors, *Artificial Immune Systems. 4th International Conference, ICARIS 2005*, pages 234–247. Springer. Lecture Notes in Computer Science Vol. 3627, Banff, Canada, August 2005.
- [9] Vincenzo Cutello, Giuseppe Nicosia, and Mario Pavone. Real coded clonal selection algorithm for unconstrained global optimization using a hybrid inversely proportional hypermutation operator. In *SAC '06: Proceedings of the 2006 ACM symposium on Applied computing*, pages 950–954, New York, NY, USA, 2006. ACM.
- [10] S. Esquivel, C. Coello Coello, and V. Aragón. Artificial immune system for solving constrained optimization problems. *Revista Iberoamericana de Inteligencia Artificial (AEPIA)*, ISSN: 1137-3601. Vol. 11, Nro. 35, pp. 55-66, 2007.
- [11] S. Esquivel, C. Coello Coello, and V. Aragón. A novel model of artificial immune system for solving constrained optimization problems with dynamic tolerance factor. *VI Mexican International Conference on Artificial Intelligence (MICA107)*, Aguascalientes, México. LNAI 4827, pp. 19-29, year=2007.
- [12] S. Esquivel, C. Coello Coello, and V. Aragón. Optimizing constrained problems through a t-cell artificial immune system, 2008. *Journal of Computer Science & Technology (JCS&T)*, ISSN 1666-6038 Vol. 8, Nro. 3, pp. 158-165.
- [13] S. Esquivel, C. Coello Coello, and V. Aragón. Solving constrained optimization using a t-cell artificial immune system, 2008. *Revista Iberoamericana de Inteligencia Artificial (AEPIA)*, ISSN: 1137-3601. Vol. 12, Nro. 40, pp. 7-22.
- [14] J. Golinski. An adaptive optimization system applied to machine synthesis. *Mechanism and Machine Synthesis*, 8.
- [15] http://www.ntu.edu.sg/home/EPNSugan/index_files/CEC-06/CEC06.html.
- [16] Versterstrom J. and Thomsen R. A comparative study of differential evolution, particle swarm optimization and evolutionary algorithms on numerical benchmark problems. *Congress on Evol. Comp., CEC04*, vol. 1, pp. 1980-1987, 2004.
- [17] Ricardo Landa Becerra and Carlos A. Coello Coello. Cultured differential evolution for constrained optimization. *Computer Methods in Applied Mechanics and Engineering*, 195(33–36):4303–4322, July 1 2006.
- [18] E. Mezura and C. Coello. Useful infeasible solutions in engineering optimization with evolutionary algorithms. In *Proceedings of the 4th Mexican International Conference on Artificial Intelligence, MICA1 2005*, pages 652–662, 2005.
- [19] Leandro Nunes de Castro and Jonathan Timmis. *Artificial Immune Systems: A New Computational Intelligence Approach*. Springer-Verlag, New York, 2002.
- [20] K. Ragsdell and D. Phillips. Optimal design of a class of welded structures using geometric programming optimization. *ASME Journal of Engineering for Industries*, 98(3):1021–1025, 1976.
- [21] T. P. Runarsson. Approximate evolution strategy using stochastic ranking. In *2006 IEEE World Congress on Computational Intelligence*, volume 3, pages 2760–2767, British Columbia, Canada, 2006.
- [22] T. P. Runarsson and X. Yao. Stochastic ranking for constrained evolutionary optimization. In *IEEE Transactions on Evolutionary Computation*, volume 3, pages 284–294, 2000.
- [23] E. Sandgren. Nonlinear integer and discrete programming in mechanical design optimization. *ASME Journal of Mechanical Design*, 112.
- [24] P. Sriyanyong, Y. H. Song, and P. J. Turner. Particle swarm optimisation for operational planning: Unit commitment and economic dispatch. Chapter in *Evolutionary Scheduling*. Keshav P. Dahal, Kay Chen Tan and Peter I. Cowling (Eds.). ISSN: 1860-949X. Studies in Computational Intelligence, Vol. 49, Springer. pp. 313-347, 2007.
- [25] Cutello V, Narzisi G, Nicosia G, and Pavone M. An immunological algorithm for global numerical optimization. *7th International Conference on Artificial Evolution, EA'05*, October 26-28 2005, University of Lille, France, Springer-Verlag, Lecture Notes in Computer Science, vol. 3871, pp. 284-295.
- [26] Xin Yao, Yong Liu, and Guangming Lin. Evolutionary programming made faster. *IEEE Transactions on Evolutionary Computation*, 3:82–102, 1999.

Estimación espectral en ETM+: Infrarrojo Muy Cercano

Paula M. Tristan(1)(2), Ruben S. Wainschenker(1) , Jorge H. Doorn(1)

(1)INTIA, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires
Paraje Arroyo Seco, Campus Universitario (7000), Tandil, Argentina Tel.(02293) 439682 Int. 49.

(2)CONICET, Rivadavia 1917, Ciudad Autónoma de Buenos Aires, Argentina

e-mail: {ptristan,jdoorn,rfw}@exa.unicen.edu.ar

CONTEXTO

La presente propuesta se corresponde con la línea de investigación “Procesamiento de Imágenes Satelitales” en el proyecto de investigación “Procesamiento de Señales” del grupo de investigación en “Procesamiento de Señales y Bases de Datos” de la UNCPBA.

RESUMEN

En los últimos años han aparecido un conjunto de metodologías que permiten realizar operaciones de fusión de datos, entendiendo esto como un proceso en el que, partiendo de imágenes de diferentes resolución espacial, espectral y/o radiométrica, se obtienen imágenes complementarias. En Teledetección Satelital, los mayores esfuerzos de investigación se han concentrado en la mejora de la resolución espacial de las imágenes de alta resolución espectral conocidas como imágenes hiperespectrales. En este trabajo se presenta una alternativa que permite mejorar la resolución espectral de imágenes multiespectrales obtenidas por el ETM+ a bordo de la serie Landsat. Basándose en los datos de cuatro bandas espectrales, de las ocho que son provistas por este sensor, se calcula una nueva banda que, de aquí en adelante se denomina *infrarrojo muy cercano*. Esta nueva banda discrimina aquellos valores de energía que deberían haberse recibido en cierto rango del espectro que esta incluido en una de las bandas pero no considerado por las restantes. Es decir, esta nueva banda contiene información que explícitamente no esta discriminada en ninguna otra banda y que, por su proximidad espectral al infrarrojo cercano puede denominarse como tal. Esta característica que puede extenderse tanto a otros rangos del espectro como a otros sensores.

Palabras clave: Teledetección, Fusión de Datos, Resolución Espectral

1. INTRODUCCION

Desde hace ya varios años en Teledetección, para distinguir diferentes zonas de una escena, se utilizan métodos que permiten la construcción de imágenes combinando matemáticamente los valores digitales de algunas bandas, en los últimos tiempos las técnicas de fusión de datos conforman una de las áreas de investigación en procesamiento de

imágenes mas populares. Debido a que cada vez existe un mayor número de sensores que observan la Tierra, cada uno de ellos con sus características propias, ha hecho evidente la utilidad de desarrollar técnicas que permitan obtener un máximo beneficio de los datos obtenidos. Su utilidad reside en la capacidad de extraer información de las imágenes en si mismas o de sensores que carecen de alguna capacidad mediante el uso de otro sensor que aporta información complementaria, utilizando para esto técnicas matemáticas y modelos físicos [Tapiador et al, 2001].

Existen varias alternativas de fusión de datos que apuntan a la mejora de la resolución espacial de los diferentes sensores conocidas como fusión de imágenes [Liu et al, 2000] las que también se conocen como “Pansharpening”, [Thomas et al, 2008], [Alparone et al, 2007].

Las tablas, figuras y límites que se indican en el resto del documento están referidas al ETM+ del Landsat 7, pero pueden aplicarse por ejemplo a: SPOT, QUICKBIRD, ORBVIEW entre otros.

En particular para el sensor ETM+ Boggione et al, 2003 desarrolló una alternativa de simulación que genera una nueva imagen manteniendo la resolución espacial de la banda pancromática B_p pero incrementando la resolución espectral ya que se obtiene realizando una fusión de las bandas B_2 , B_3 y B_4 (Figura 3). En el presente trabajo se adopta el enfoque opuesto a Boggione, ya que se procura obtener información no explícita en las imágenes satelitales.

Se presenta entonces, un método que permite crear o calcular una imagen correspondiente a una nueva banda del espectro B_{Mc} correspondiente a lo que denominaremos *infrarrojo muy cercano*, y que no esta discriminada por los sensores del ETM+. Para resolverlo se realiza una combinación de datos de las bandas correspondientes al verde B_2 , la rojo B_3 , la del infrarrojo cercano B_4 y la pancromática B_p . La nueva imagen estima los valores que se hubieran recibido con un dispositivo sensible al infrarrojo muy cercano.

El resto del artículo se organiza como sigue: a continuación se presenta el marco teórico en el que se exponen algunos conceptos y se describen algunos aspectos técnicos esenciales para comprender el trabajo, luego se explica la metodología implementada y se muestran algunos ejemplos de los resultados obtenidos y por último se citan las conclusiones y se detallan las diferentes vías de continuidad y expansión del trabajo.

2. LINEAS DE INVESTIGACION y DESARROLLO

Se denomina resolución espectral a la capacidad del sensor para discriminar la radiancia detectada en distintas longitudes de onda del espectro electromagnético. La resolución espectral está determinada por el número de bandas que el sensor puede captar y por la anchura espectral de éstas [Chuvieco et al 1994].

En determinados casos, el sensor será de mayor utilidad cuanto mayor sea el número de bandas que proporcione, ya que algunas cubiertas requieren estudios multispectrales. Por otro lado, conviene que el ancho de cada banda sea lo más reducida posible, con el objeto de no obtener valores medios de regiones espectrales con diferentes significados físicos. Algunos sensores discriminan el espectro en tres bandas mientras que los hiperespectrales oscilan alrededor del centenar.

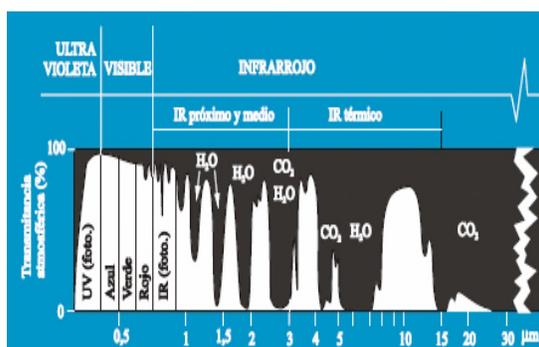


Figura 1. – Situación de absorción y transmisión atmosférica

La radiación del Sol y la reflexión de ésta desde la superficie de la Tierra atraviesan la atmósfera antes de llegar al sensor del satélite la que se comporta como un filtro selectivo en distintas longitudes de onda [Rees et al, 2001]. Debido a su composición química produce, entre otros efectos, la absorción de parte de la radiación.

ventana	banda espectral (μm)
1	0,3 - 1,3
2	1,5 - 1,8
3	2,0 - 2,6
4	3,0 - 3,6
5	4,2 - 5,0
6	7,0 - 15,0

Figura 2.– Ventanas Atmosféricas

Un sensor fuera de la Atmósfera recibirá de la superficie terrestre sólo las longitudes de onda en

que la atmósfera es transparente. Las longitudes de onda en las que la atmósfera es transparente y la radiación puede atravesarla se denominan ventanas atmosféricas, algunas de las cuales se muestran en la Figura 2. [Pinilla et al, 1995]. La atmósfera permite la transmisión de la luz visible y del infrarrojo cercano bloqueando la mayor parte de la radiación infrarroja.

El sensor convierte la radiancia del terreno en una señal continua, luego un convertidor analógico-digital genera el nivel digital o ND que es almacenado para el usuario final. Para conocer el comportamiento físico de la cubierta observada es necesario encontrar la función de transformación entre la radiancia y el ND. Conocida esa función, la aplicación de su inversa a los ND de la imagen permite caracterizar físicamente la cubierta. Se define la radiancia recibida por el sensor $L(b, s)$ como:

$$L(b, s) = G_b * ND_b + B_b$$

Esta es una relación aproximada [Landsat Handbook] en la que se suponen que tanto la distribución espectral de la fuente luminosa s utilizada en la calibración del instrumento, como la de la radiancia que llega de la Tierra son iguales a la banda b , En dónde G_b es la ganancia y B_b es el biase para dicha banda.

$$L(b, s) = \int RSR(b, \lambda) * L(\lambda, s) d\lambda = L(s) * \int RSR(b, \lambda) d\lambda$$

ETM+				
Banda	Nomenclatura	Espectro	Ancho Banda	Resol. Espacial
1	B ₄₅₋₅₂	0,45-0,52 μm	7	30
2	B ₅₂₋₆₀	0,52-0,6 μm	8	30
3	B ₆₃₋₆₉	0,63-0,69 μm	6	30
4	B ₇₆₋₉₀	0,76-0,9 μm	14	30
5	B ₁₅₅₋₁₇₅	1,55-1,75 μm	20	30
6	B ₁₀₄₀₋₁₂₅₀	10,4-12,5 μm	210	60
7	B ₂₀₈₀₋₂₃₅₀	2,08-2,35 μm	270	30
Pan	B ₅₂₋₉₀	0,52-0,9 μm	38	15

Figura 3.- Descripción espectral y espacial de las bandas de ETM+.

Sabiendo que la respuesta espectral RSR de cada sensor varía según la longitud de onda, tal como se muestra en la Figura 3. Cada rango de longitudes de onda representan las distintas bandas, existiendo

también, rangos que no están representados por ninguna banda.

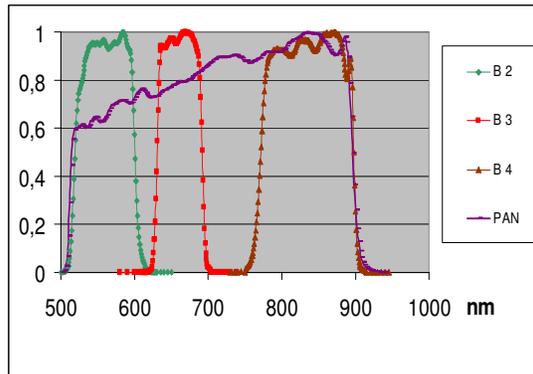


Figura 4.- Curvas espectrales de la banda B₂, B₃, B₄ y B_p del ETM+.

En la Figura 4 se observa el comportamiento espectral de B₅₂₋₆₀, B₆₃₋₆₉ y B₇₆₋₉₀ respecto de la banda pancromática B₅₂₋₉₀, en donde se percibe que existen dos zonas en el espectro para las cuales solo existe información en la banda pancromática. Una pequeña entre 0.60 - 0.63 y otra casi tan importante como la banda B₆₃₋₆₉ entre 0.69 - 0.76 que, por su posición en el espectro se podría denominar como *infrarrojo muy cercano*, B_{Mc} o banda B₆₉₋₇₆.

Como primera aproximación, se podría despreciar la banda más pequeña, entre 0.60 y 0.63, debido a que: a los efectos de los cálculos su incidencia en el resultado final es menor y puede considerarse dentro del error asumido.

3. RESULTADOS OBTENIDOS

Suponiendo que $L(2)$, $L(3)$, $L(Mc)$, $L(4)$, $L(P)$ son los valores de radiancia recibida en las bandas B52-60, B63-69, B69-76, B76-90 y B52-90 respectivamente, se puede definir la radiancia recibida en la banda pancromática B_p como:

$$L(P) = \frac{\sum_x L(x) \cdot \int RSR(x, \lambda) d\lambda}{\int RSR(P, \lambda) d\lambda}$$

Si se define $F_x = \int RSR(x, \lambda) d\lambda$ entonces puede expresarse la relación anterior como:

$$L(P) = \frac{F_2 \cdot L(2)}{F_p} + \frac{F_3 \cdot L(3)}{F_p} + \frac{F_{OC} \cdot L(Mc)}{F_p} + \frac{F_4 \cdot L(4)}{F_p}$$

$$L(Mc) = \frac{F_p \cdot L(P) - F_2 \cdot L(2) - F_3 \cdot L(3) - F_4 \cdot L(4)}{F_{Mc}}$$

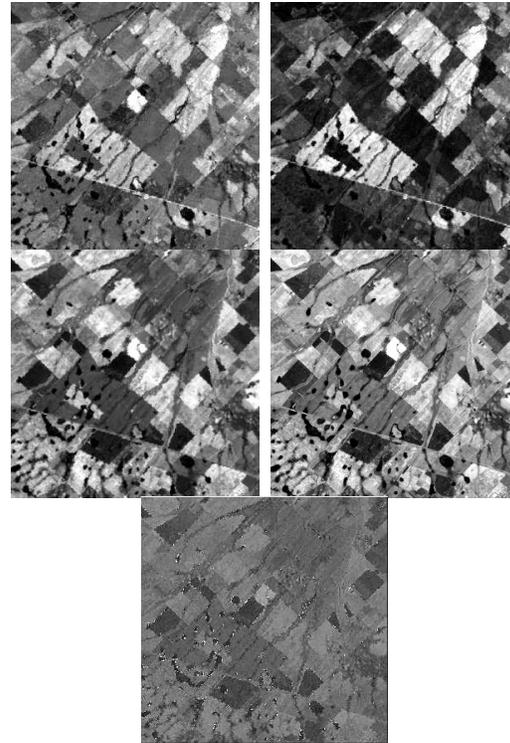


Figura 5.- B₅₂₋₆₀, B₆₃₋₆₉, B₇₆₋₉₀, B₅₂₋₉₀ B₆₉₋₇₆ respectivamente.

En la Figura 5 se muestra una pequeña zona de una imagen en las bandas B52-60, B63-69, B76-90, B52-90 y la banda calculada B69-76. Tal como puede observarse existe una cierta correlación entre la energía recibida por el sensor infrarrojo cercano y los valores calculados para el infrarrojo muy cercano.

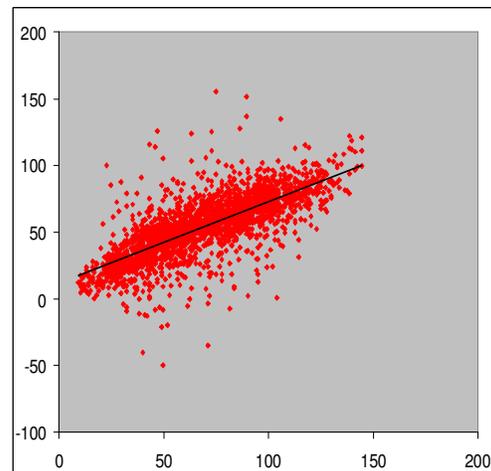


Figura 6.-Relación existente entre la B₄ infrarrojo cercano (abscisas) y la del infrarrojo muy cercano B_{Mc} (ordenadas).

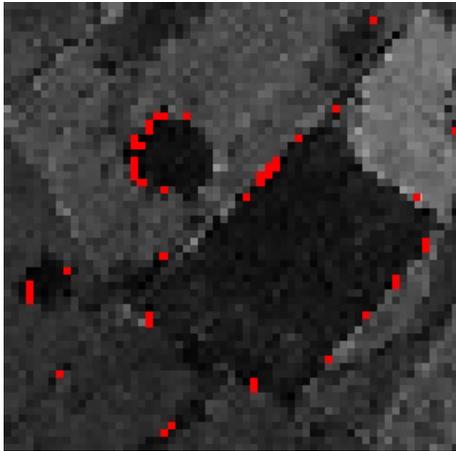


Figura 7.- Outliers encontrados en la B_{Mc}

Sin embargo, también se puede observar en la Figura 6 que existe un número importante de píxeles en los que la correlación no es válida. Se ha comprobado que tanto los puntos que están próximos a la línea de regresión lineal como aquellos que se encuentran a una distancia determinada se agrupan en bloques que coinciden con objetos observables en otras bandas y con objetos reales de la superficie. Esta es una indicación muy fuerte de que la banda calculada de esta manera contiene información útil acerca de la naturaleza de la superficie observada.

Asimismo, se ha detectado la presencia de puntos cuya energía calculada es negativa o adquiere un valor demasiado alto. En la Figura 7 se muestra un fragmento de una imagen calculada para B_{Mc} ampliada para resaltar los píxeles, cuyo valor de energía es inválido. Mayormente, estos puntos se presentan en las fronteras entre áreas de la imagen que muestran gran diferencia entre los valores de gris, ya sea en la banda pancromática B_P como en las bandas B_2 , B_3 o B_4 . La hipótesis que se maneja en este momento, sin ningún tipo de comprobación, es que los procesamientos relacionados con las correcciones geométricas, realizadas en las estaciones terrenas, han sido ligeramente diferentes entre las bandas involucradas.

3.1 CONCLUSIONES Y TRABAJOS FUTUROS

Se ha obtenido una nueva banda para el sensor ETM+ de forma matemática y se ha comprobado que la misma contiene información propia. La imagen resultante contiene algunos píxeles espurios que no han sido estudiados todavía.

Se planifica mejorar las correcciones geométricas para el cálculo de la imagen de la B_{Mc} .

Se espera realizar un estudio que cuantifique la efectividad de los parámetros de cálculo con los límites de los rangos espectrales elegidos.

Se aspira a utilizar la información de los *outliers* como mecanismo alternativo de evaluación de las correcciones geométricas realizadas sobre la imagen

y eventualmente poder opinar acerca de cual debiera ser la estrategia más apropiada para realizar estas correcciones. Por otro lado se aspira a evaluar la posible construcción de la banda B_{Mc} en niveles de procesamiento más bajos para luego realizar las correcciones geométricas sobre la banda ya calculada.

Por último, se espera contrastar las imágenes obtenidas con información provista por sensores hiperspectrales y eventualmente con datos de la superficie.

4. FORMACION DE RECURSOS HUMANOS

El presente proyecto es parte de la tesis doctoral "Mejoras de Resolución de Imágenes Satelitales: Algoritmos de Integración" que está desarrollando la Mg. Paula M. Tristan en la UNCPBA

5. BIBLIOGRAFIA

Alparone, L. Wald, L., Chanussot, J., Thomas, C., Gamba, P. and Mann Bruce, L. "Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data Fusion Contest". IEEE Transactions on Geoscience and Remote Sensing, Vol. 45 Nro. 10, Octubre 2007.

Boggione, G., Pires, E., Santos, P. y Fonseca, L. "Simulation of a Panchromatic Band by Spectral Combination of Multispectral ETM+Bands" International Geoscience and Remote Sensing Symposium, IGARSS'91.

Chuvieco E. "Fundamentos de la Teledetección Espacial". Ed. Rialp S.A., Madrid-España. 1994.

Tapiador, F y J. Casanova. "La fusión de datos en Teledetección". Revista de Teledetección Nro. 15, Junio de 2001.

Liu, J. "Evaluation of Landsat-7 ETM+ Panchromatic Band for Image Fusion with Multispectral Bands". Natural Resources Research, Vol. 9 Nro. 4. pp. 269-276. 2000

Pinilla C. "Elementos de Teledetección". Ed. Ra-Ma., Madrid. 1995.

Rees W. G. "Physical principles of remote sensing". Ed. Cambridge University Press. 2001.

Thomas, C., Ranchin, T., Wald, L., and Chanussot, J. "Síntesis of multispectral Images ti High Spatial Resolution: A Critical Review of Fusion Methods Based on Remote Sensing Physics". IEEE Transactions on Geoscience and Remote Sensing, Vol. 46 Nro. 5, Mayo 2008.

"**Landsat 7 Science Data Users Handbook**". Manual de usuario del Programa Landsat.

<http://landsathandbook.gsfc.nasa.gov/handbook.html>

MODELIZACION DE FENÓMENOS FÍSICOS EN APLICACIONES DE COMPUTACIÓN GRÁFICA EN TIEMPO REAL

Lazo M.^{1,2}, García Bauza C.^{1,2}, Vénere M.^{1,3} y Clause A.^{1,4}

¹ Universidad Nacional del Centro, 7000 Tandil, Argentina,

² Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

³ Comisión Nacional de Energía Atómica

⁴ CONICET

1 INTRODUCCIÓN

Cuando se desarrollan aplicaciones de realidad virtual, uno de los principales desafíos es lograr realismo e inmersión a un costo computacional aceptable. Con las placas gráficas actuales es posible obtener imágenes muy cercanas a la realidad, aunque también se debe incluir comportamiento físico para lograr escenas que brinden una verdadera inmersión.

Para cumplir este objetivo, se han desarrollado en los últimos años varios motores físicos implementados en forma de bibliotecas (ODE, NGD, TPE, BPL¹) o en placas dedicadas a la simulación de modelos físicos (por ejemplo, el producto de nVIDIA[®] denominado PhysX); los cuales permiten agregar comportamiento físico a escenas tridimensionales.

Es posible que los usuarios deseen realizar pruebas de rendimiento y/o precisión a cada motor físico para determinar cual de ellos se ajusta mejor a sus necesidades; o utilizar distintas funcionalidades de cada uno en una misma aplicación, pero ocurre que cada uno de estos motores necesita una configuración distinta de inicialización y muchas veces la funcionalidad implementada difiere entre uno y otro, haciendo que el código implementado sea obsoleto, requiriendo de un esfuerzo considerable para portar la aplicación a otro motor.

Para solucionar dichos problemas, el presente trabajo se centra principalmente en crear una capa de abstracción o PAL (del inglés, *Physics Abstraction Layer*) que sea utilizada por un motor gráfico, la cual brinde una comunicación fluida con diversos motores físicos maximizando la reutilización de código y permita cambiar el uso entre ellos.

2 CAPA DE ABSTRACCION PROPUESTA

Si bien existen esfuerzos en esta área [Boeing 2005, Fischer 2005], estas propuestas se alejan de la idea de facilitar el intercambio de motores físicos de una manera transparente y cómoda para el usuario como la propuesta de [Fischer 2005], o proveen un conjunto acotado de funcionalidad como [Boeing 2005]; exponiendo únicamente la funcionalidad común de los motores físicos que soporte la PAL; es decir, se utiliza un esquema de intersección de funcionalidades; no permitiendo al usuario utilizar algunas de las características de aquellos motores físicos que tengan mayor funcionalidad (Figura 1 Izq.).

En nuestra propuesta, (Figura 1 Der.) se utiliza un esquema de unión de funcionalidades, no sólo proveyendo acceso a la funcionalidad común de los motores físicos, sino también brindando acceso a funcionalidad específica a través del uso de decoradores de diseño (conocidos como *wrappers*). Esto permite que la PAL soporte el conjunto completo de funcionalidades de los motores físicos soportados.

¹ Ver Información adicional en Sección Referencias

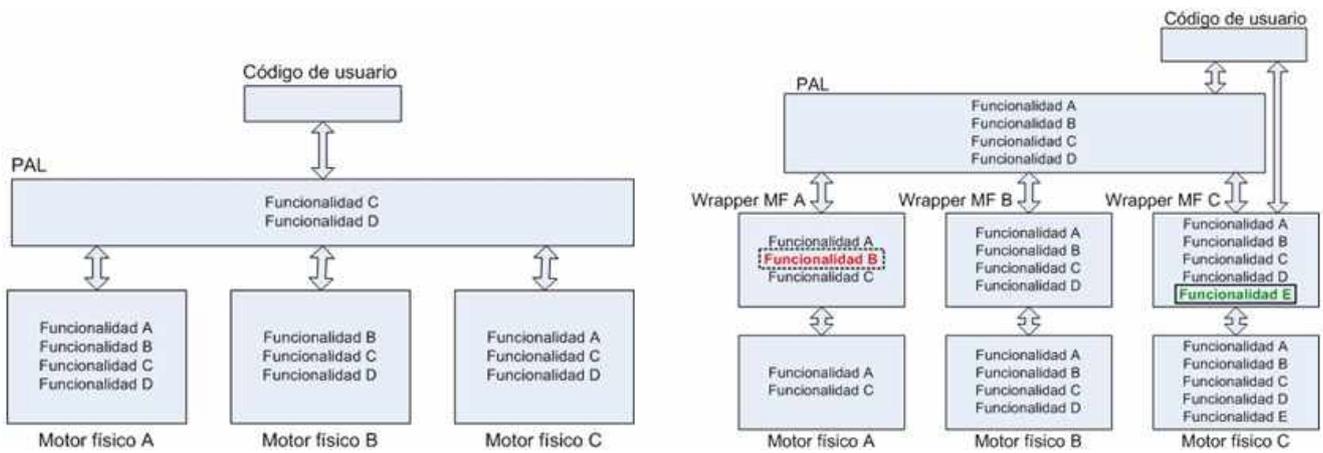


Figura 1: Izq: Capa de abstracción utilizando esquema de intersección de funcionalidades. Der: Capa de abstracción utilizando esquema de unión de funcionalidades.

Tanto las propuestas de [Boeing 2005] como [Fischer 2005] dejan a criterio del usuario la utilización o integración de una biblioteca gráfica; siendo éste un problema no trivial, puesto que comúnmente la interacción entre las dos partes es crucial. Por lo tanto, se implementaron dos mecanismos de sincronización, donde en el primero la simulación está guiada por el motor gráfico y en el segundo, está guiada por el motor físico.

2.1 Sincronización utilizando *framerate* fijo en el motor gráfico

En este enfoque la simulación está guiada por el motor gráfico, el cual indica cuando se realizan las actualizaciones del estado de la escena. Se establece un paso de tiempo para el motor gráfico y un paso de tiempo para el motor físico que represente el tiempo transcurrido entre actualizaciones del motor gráfico. Por ejemplo, estableciendo el paso de tiempo del motor gráfico en X segundos; luego es necesario representar ese tiempo con el motor físico. Dependiendo del motor físico, será necesario realizar varias actualizaciones utilizando pasos de tiempo menores (Figura 3).

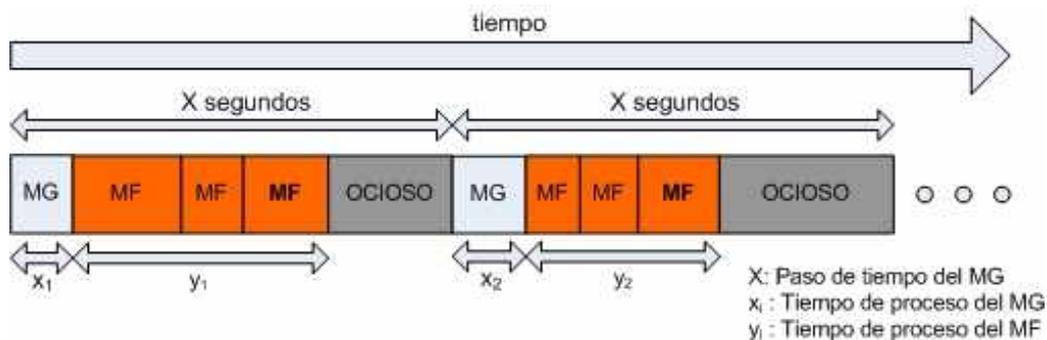


Figura 3: Esquema de interacción utilizando tiempo de actualización fijo en motor gráfico.

2.2 Sincronización utilizando *framerate* variable en el motor gráfico

En este enfoque la simulación está guiada por el motor físico, el cual indica cuando se realizan las actualizaciones del estado de la escena.

El algoritmo de sincronización consta de un acumulador de tiempo, en cada ciclo de la simulación se mide el tiempo transcurrido desde la última actualización y se suma al acumulador; luego, mientras que el valor del acumulador sea mayor o igual al paso de tiempo prefijado del motor físico, se resta este último al acumulador y se realiza la actualización del motor físico. En la Figura 2 se esquematiza un ejemplo, en el cual se acumula tiempo y se realizan actualización conforme avanza

la simulación.

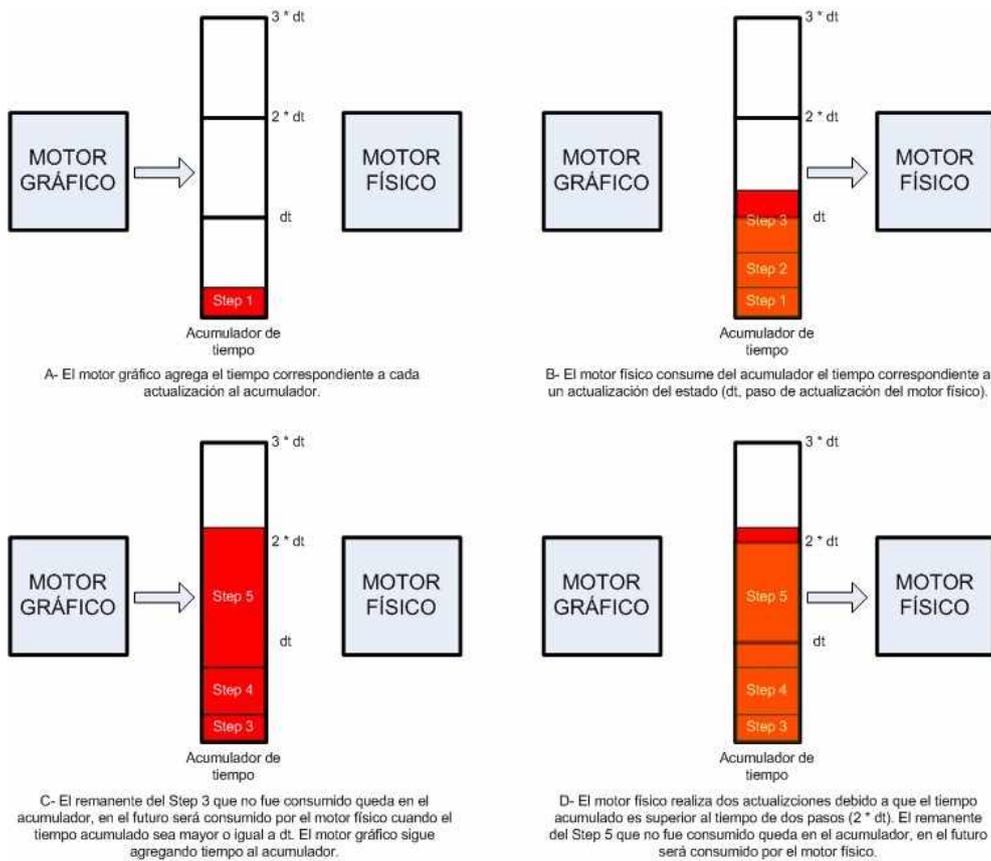


Figura 2: Esquema de interacción utilizando tiempo de actualización variable en motor gráfico.

Ambos algoritmos mantienen la sincronía entre el motor gráfico y el motor físico, asegurando que el tiempo transcurrido en la simulación represente el tiempo transcurrido en la realidad.

3 EJEMPLOS UTILIZANDO LA PAL

A continuación se presentan algunas de las aplicaciones implementadas utilizando la capa de abstracción. Todas ellas fueron implementadas utilizando el motor gráfico Impromptu [D'amato]. En la Figura 4 Izq. se visualiza una escena con 4000 esferas (320.000 polígonos) con movimiento dinámico. Para el modelo estático de la estatua se utilizó una geometría detallada para visualizar y otra aproximada para simular comportamiento. La Figura 4 Der. muestra una escena con 90 cajas con movimiento dinámico que interactúan con un vehículo en una escena compleja.



Figura 4: Escenas complejas con comportamiento físico implementadas utilizando la PAL.

La capa de abstracción provee una interfaz independiente de las bibliotecas físicas, sin agregar sobrecarga de tiempo, reduciendo significativamente el esfuerzo del programador para portar sus aplicaciones a otros motores físicos. Esto es importante, dado que cada motor físico brinda a las simulaciones velocidad de cálculo, estabilidad o precisión.

4 TRABAJOS FUTUROS

En este último tiempo, existe una tendencia de desarrollo e investigaciones de motores específicos y no de propósito general, por ejemplo se pueden citar, aquellos que simulan tejidos, objetos que se rompen, fluidos, vehículos, *ragdolls*, etc. Cada uno de estos motores se encarga del comportamiento que simula, proveyendo precisión, calidad y rendimiento. Por lo tanto es necesario proveer un mecanismo de comunicación entre estos motores específicos. La Figura 7 es una secuencia de imágenes en las cuales un motor físico de propósito general, a través de la capa de abstracción implementada, se comunica con un motor físico específico de simulación de fluidos basado en un modelo de Lattice-Boltzmann, permitiendo al personaje y las pelotas interactuar con la superficie de agua.

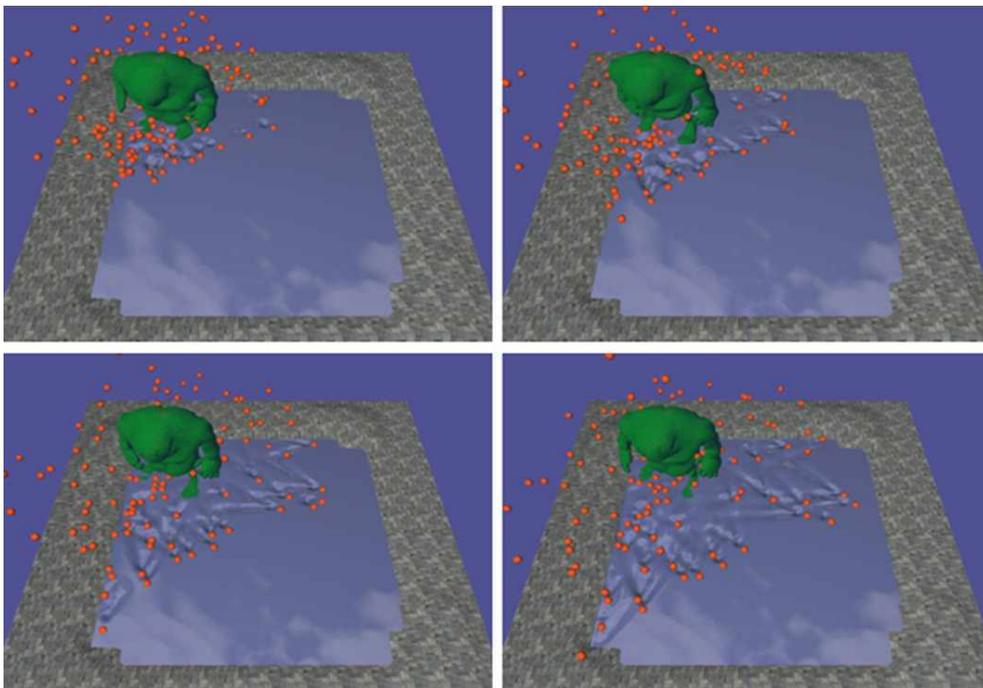


Figura 6: La secuencia muestra distintos momentos de una simulación en la cuál interactúan un motor físico y un modelo basado en Lattice-Boltzmann para la simulación de fluidos.

Actualmente se está trabajando en la simulación de efectos complejos, como un tornado interactuando con objetos del ambiente, mediante la simulación de vórtices, calculando fuerzas de rotación y atracción/repulsión y utilizando un motor físico, a través de la capa de abstracción implementada, para el cálculo de trayectorias y colisiones. En la secuencia de imágenes de la Figura 8 se muestra la simulación del inicio de un tornado.

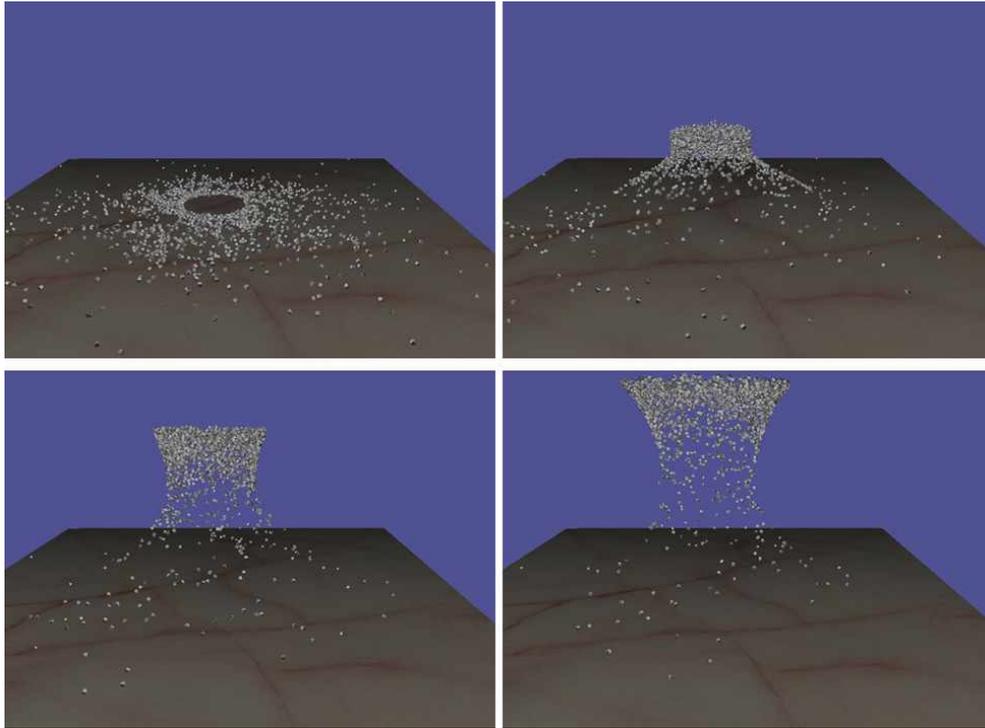


Figura 7: La secuencia muestra distintos momentos de una simulación de un tornado interactuando con 1500 objetos.

5 REFERENCIAS

BOEING A. PAL, Physics Abstraction Layer Home Page. [Online] Disponible en: <http://www.adrianboeing.com/pal/>. 2005

BPL, BULLET PHYSICS LIBRARY. [Online]. Disponible en: <http://www.bulletphysics.com>

CRISTIAN GARCÍA BAUZA, MARCOS LAZO, MARCELO VÉNERE. “Incorporación de comportamiento físico en motores gráficos”. Mecánica Computacional, Vol. XXVII, p. 3023-3039. ISSN 1666-6070.

D’AMATO J.P, GARCÍA BAUZA C.. Simulación de Escenarios Tridimensionales Dinámicos. Tesis de grado de Ingeniería de Sistemas, Universidad Nacional del Centro. 2004.

NGD, Newton Game Dynamics Home Page. [Online]. Disponible en: <http://www.newtondynamics.com/>

NVIDIA, NVIDIA PhysX Home Page. [Online]. Disponible en: http://www.nvidia.com/object/nvidia_physx.html.

ODE, Open Dynamics Engine Home Page. [Online]. Disponible en: <http://www.ode.org/>

FISCHER A., REINOT A., STREETER T.. OPAL, Open Physics Abstraction Layer Home Page. [Online] Disponible en: <http://opal.sourceforge.net/>. (2005)

TPE, Tokamak Physics Engine Home Page. [Online]. Disponible en: <http://www.tokamakphysics.com/>

Realidad Aumentada en un entorno gráfico de alta performance

Schneider, José – Martig, Sergio – Castro, Silvia

Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica
Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur

Av. Alem 1253 – 8000 – Bahía Blanca – Buenos Aires – Argentina

Tel: 4595101 - Interno: 2618

{jis, srm, smc}@cs.uns.edu.ar

Contexto de la Investigación

Este trabajo se desarrolla en el marco del proyecto Interfaces No Convencionales y su impacto en las interacciones. Siendo Sergio Martig el director de dicho proyecto. Además, este trabajo forma parte de la tesis Realidad Espacial Aumentada del Magister de Schneider, José.

El trabajo se lleva a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur. Cabe señalar que en el Laboratorio se están desarrollando distintos proyectos vinculados con Visualización y se pretende que, a mediano plazo, se puedan integrar los resultados obtenidos en el marco del proyecto actual.

Resumen

La Realidad Aumentada (RA) tiene un gran auge en el mundo gráfico actual. Sin embargo, esta área de la Computación Gráfica carece de madurez en prácticamente todos sus aspectos. Actualmente, hay pocos entornos que permitan utilizar la RA de manera sencilla y que adicionalmente permitan incorporar las últimas tecnologías gráficas. El objetivo de este proyecto consiste en integrar distintas técnicas de realidad aumentada en un entorno con las características mencionadas.

Palabras clave: Realidad Aumentada, Computación Gráfica, Dispositivos No Convencionales, Visualización

Introducción

Los términos Realidad Virtual (RV) y ciberespacio se han convertido en términos muy populares en las últimas dos décadas en todos los ámbitos fuera de la comunidad de investigación. Las películas de ciencia ficción, por ejemplo, no sólo acercaron estos conceptos al público en general sino que también influenciaron fuertemente la comunidad de investigación. La mayoría de nosotros asocia estos términos con la posibilidad tecnológica de sumergirse en un mundo sintético completamente

generado por computadora – referido habitualmente como mundo virtual. En un ambiente virtual nuestros sentidos, tales como la visión, el oído, el olfato, el tacto, etc., se controlan por medio de una computadora en tanto nuestras acciones influyen el estímulo producido.

Como en el caso de la realidad virtual, existen varias definiciones y clasificaciones formales para la realidad aumentada (por ejemplo, [Mil94a], [Mil94b]). Algunos autores definen la Realidad Aumentada como un caso especial de la RV; otros argumentan que la RA es un concepto más general y ven la RV como un caso especial de la RA. El hecho es que, en contraste con la RV tradicional, en RA el ambiente real no se suprime completamente, sino que juega un papel dominante. En lugar de sumergir una persona en un mundo completamente sintético, la RA intenta embeber elementos sintéticos en el ambiente real. Esto plantea un problema fundamental: el ambiente real es mucho más complejo de controlar que el completamente sintético. La información aumentada tiene una relación mucho más fuerte con el ambiente real y ésta es mayormente una relación espacial entre la aumentación y el ambiente real; esto se conoce como registración.

Así como la Realidad Virtual (RV), la Realidad Aumentada (RA) está emergiendo como una tecnología con grandes posibilidades de aplicación en diferentes ámbitos. Nuevas alternativas plantean conducir hacia aplicaciones de RA nuevas, flexibles y de gran eficiencia en muchos dominios de aplicación no móviles, habilitando nuevas áreas de aplicación para museos, educación, entretenimiento, investigación, industria, obras artísticas, etc.

Por otro lado, la visualización es una herramienta muy fuerte de análisis y presentación de datos en áreas como ciencia, ingeniería y medicina. El procesamiento, la consulta, la exploración y la visualización de distintos conjuntos de datos presenta una serie de interesantes desafíos computacionales, visuales y de interacción.

Una alternativa para proveer una Visualización de Datos adecuada exige además alternativas más demandantes en las que el escritorio ya no es suficiente; salir del escritorio se puede lograr usando tecnologías de RA que pueden variar en función de las necesidades planteadas por los distintos problemas y de los recursos disponibles por los usuarios.

Líneas de investigación y desarrollo

El presente proyecto se encuadra en un proyecto más general que tiene como objetivo el desarrollo de *Interfaces de Usuario* que trasciendan los escritorios. Estas interfaces emergentes emplean técnicas novedosas no sólo para el ingreso de los datos sino que se pretende que la presentación de la información se realice a través de una gran variedad de dispositivos tendientes a embeberse en el entorno del usuario. Los objetivos generales de este proyecto consisten en:

- Analizar los aspectos tecnológicos de las mismas enfocándonos en los requerimientos de hardware y en los desafíos que implican su implementación.
- Detectar las interacciones soportadas por las distintas interfaces, revisando la validez de los distintos estilos de interacción y procediéndose a su adecuación en caso de ser necesario.

Basándonos en los resultados del relevamiento antes mencionado se buscarán las maneras de lograr interfaces de este tipo utilizando tecnología de bajo costo, dentro de nuestro alcance, para poder implementarlas y lograr de esta manera obtener un entorno en el cual validar las conclusiones del análisis de los paradigmas de interacción.

En este contexto general se pretende estudiar la aplicación de estas interfaces en dos áreas donde la interacción juega un rol esencial.

Las aplicaciones de visualización constituyen, en este caso, un espacio de análisis y experimentación muy rico; pues las propias visualizaciones se convierten en las interfaces que posibilitan que el usuario, a través de las interacciones adecuadas, pueda explorar el espacio de información subyacente logrando el *insight* buscado de los datos.

Otro campo donde las interacciones tienen un rol preponderante es el de los juegos, en este campo tanto la visualización como las interacciones se integran naturalmente en la interfaz, abriendo todo un abanico de posibilidades y desafíos tendientes a lograr que la comunicación hombre-máquina se establezca de una manera natural y efectiva.

Estas son dos de las áreas que se verán beneficiadas con la introducción de la RA como una forma de salir del escritorio y los estilos de interacción humano-computadora convencionales.

Por esta razón se llegó a la conclusión que resultaría muy útil y indudablemente atrayente contar con un entorno de RA que sea fácil de utilizar y que tenga incorporadas las últimas tecnologías gráficas, en particular el pipeline programable. Este entorno permitirá el desarrollo de aplicaciones de alta performance en tiempo real.

Resultados obtenidos/esperados

En las primeras etapas del proyecto se realizó una evaluación de forma aislada de las distintas tecnologías necesarias para cumplir los objetivos particulares propuestos tendientes a contar con el mencionado entorno de RA mencionado en la sección anterior, comenzando a su vez a integrar algunas de estas tecnologías entre sí. Los resultados a pesar de ser positivos aún no son concretos.

En lo que respecta al entorno de RA se están integrando los sistemas no comerciales más relevantes de RA, ARTag y Artoolkit Plus, sobre un motor gráfico perteneciente a un trabajo previo del tesista Schneider, José. Este motor gráfico surgió del trabajo final de su tesis de grado, junto con un proyecto personal. Este motor está programado en C#, XNA y HLSL.

Como trabajo futuro se contempla la posibilidad de incorporar el Wiimote al entorno como dispositivo de entrada alternativo a los ya existentes. Esto permitirá no sólo enriquecer las capacidades del motor, sino también incorporar al Wiimote como un rico dispositivo de interacción en un esquema de realidad aumentada alternativo o complementario.

Formación de Recursos Humanos

En lo que respecta a la Formación de Recursos humanos el trabajo desarrollado contribuirá a la formación de tesis de Posgrado así como de Becarios de Investigación. Por otro lado también contribuirá a la organización de una materia de Posgrado que se dictará en tentativamente en el segundo cuatrimestre del corriente año. Asimismo, este trabajo pretende contribuir a los trabajos de Computación Gráfica, materia curricular de nuestro departamento en la que, como tarea final, sus alumnos realizan un proyecto utilizando un enfoque simple de Realidad Aumentada a elección.

Actualmente uno de los miembros del Laboratorio tiene una beca de Investigación de la CIC (Comisión de Investigaciones Científicas de la Provincia de Buenos Aires) para realizar estudios sobre técnicas básicas de RA y su aplicación en el ámbito de Visualización; y realiza en paralelo su tesis de Magister sobre el mismo tema. Asimismo, a partir de abril se integrará a este proyecto un alumno egresado de la Ingeniería en Ciencias de la Computación, que además posee una beca de la ANPCyT y cuyo tema se relaciona directamente con lo propuesto en el presente proyecto de Investigación.

Por otro lado, distintos miembros del Laboratorio están participando en proyectos de Visualización y de Interacción Hombre Máquina. Estos se verán enriquecidos con los resultados a obtener en el presente proyecto dado que se considera que la RA puede contribuir al campo de la Visualización al integrar naturalmente interacciones tendientes a lograr que la comunicación hombre-máquina se establezca de una manera natural y efectiva.

Bibliografía

- [Bim04] O. Bimber & R. Raskar. *Modern Approaches to Augmented Reality*. Conference Tutorial Eurographics, 2004.
- [Bim05] O. Bimber & R. Raskar. *Spatial Augmented Reality. Merging Real and Virtual Worlds*. A K Peters, Wellesley, Massachusetts, 2005. ISBN 1-56881-230-2
- [Bro04] K.W. Brodlie, D.A. Duce, J.R. Gallop, J.P.R.B. Walton, J.D. Wood, “*Distributed and Collaborative Visualization*”, The Eurographics 2004
- [Caw07] Stephen Cawood, Mark Fiala, “*Augmented Reality. A practical Guide*”.
- [Mar03] Martig, S., Castro, S., Fillotrani, P. & Estévez, E , “*Un Modelo Unificado de Visualización*”. Proceedings, pp. 881-892, 9º Congreso Argentino de Ciencias de la Computación. 6 al 10 de Octubre de 2003. La Plata. Argentina.
- [Fia04] Mark Fiala, “*ARTag Revision 1. A Fiducial Marker System Using Digital Techniques*”.

- [Mil94a] P. Milgram and F. Kishino. *A Taxonomy of Mixed Reality Visual Displays*. IEICE Transactions on Information Systems E77-D 12 (1994), 1321–1329.
- [Mil94b] P. Milgram, H. Takemura et al. *Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum*. In Proceedings of SPIE: Telem manipulator and Telepresence Technologies 2351 (1994), 282–292.
- [New01] M. Newman, “*The Structure of Scientific Collaboration Networks*”, Proc. Nat'l Academy of Science USA, vol. 98, pp. 404-409, 2001.
- [Stri03] D. Stricker, W. Kresse, J. F. Viguera s-Gomez, G. Simon, S. Gibson, J. Cook, P. Ledda, and A. Chalmers. *Photorealistic Augmented Reality*, Conference tutorial ISMAR 2003 (Second International Symposium on Mixed and Augmented Reality). Los Alamitos, CA: IEEE Press, 2003.
- [Str01] S. Strogatz, “*Exploring Complex Networks*”, Nature, vol. 410, no. 8, pp. 268-276, Mar. 2001.
- [Sut65] I. E. Sutherland. *The Ultimate Display*. Proceedings of IFIP'65, pp. 506–508, 1965.
- [Sut68] I. E. Sutherland. *A Head Mounted Three Dimensional Display*. Proceedings of the Fall Joint Computer Conference (AFIPS) 33:1 (1968) 757–764.
- [Woo00] “*Volume Graphics and the Internet*”, Ken Brodlie, Jason Wood, 2000.

SISTEMA AUTOMÁTICO NO DESTRUCTIVO PARA LA CLASIFICACIÓN DE LA CALIDAD DE FRUTAS BASADO EN VISIÓN POR COMPUTADORA

Guillermo M Sampallo - Arturo O González Thomas
Facultad Regional Resistencia, UTN
Resistencia, Chaco, Argentina

CONTEXTO

La línea de I+D presentada es una continuación y profundización de proyectos desarrollados por el grupo previamente. Todos dentro del área de procesamiento digital de imágenes y la utilización de la Inteligencia artificial para el procesamiento de los datos obtenidos por ese medio.

RESUMEN

El objetivo de este proyecto es determinar empleando procesamiento digital de imágenes, un conjunto de descriptores morfológicos y de aspecto, y a partir de ellos seleccionar el subconjunto más apropiado para la evaluación de la calidad de frutas. Estos descriptores son el insumo que emplea un sistema experto para la clasificación según estándares de calidad preestablecidos.

El método desarrollado calcula los descriptores morfológicos en forma automática, independientemente de la posición del producto en la escena. Los descriptores son firma, alto, ancho, perímetro, área, factores de forma, desviación cuadrática media, distancia significativa, superficie y volumen. La aplicación de este método está restringida a frutas esféricas con elevada simetría respecto de un eje. Para comprobar la validez de los algoritmos de cálculo se los aplicó primero a modelos y se valuó su performance. Finalmente se los aplicó a los siguientes productos: naranja y quinoto, y se realizó un análisis de los errores del método.

Los descriptores de aspecto analizados son: de textura (descriptores que surgen de la matriz de coocurrencia), color (análisis de las propiedades por banda color) y manchas (reconocimiento de la región manchada y calculo porcentaje manchado en la cara visible).

Palabras clave: visión por computadora, segmentación, descriptores, calidad.

INTRODUCCION

El consumidor de frutas es cada vez más selectivo y exigente al momento de comprar. Analiza la apariencia a través de la forma, la uniformidad y los defectos, y madurez a través del color, tamaño y textura. Esto trae como consecuencia la necesidad de optimizar las tareas de control de calidad para satisfacer dicha demanda.

La valoración de la calidad de productos alimenticios es a menudo subjetiva, y basada en cualidades tales como aspecto (color y forma), olor, textura y sabor. Frecuentemente las mismas son evaluadas por inspectores humanos. Se ha demostrado que la opinión humana puede ser engañada fácilmente. Tanto el costo de la labor como la subjetividad de los resultados acentúan la necesidad de instrumentar sistemas objetivos de control de la calidad.

Una alternativa es la aplicación de la inspección visual automatizada, debido a su bajo costo relativo, robustez, adaptabilidad, superior velocidad y exactitud en los resultados.

El empleo de la visión por computadora y software asociado para la clasificación de objetos es una herramienta cada vez más ampliamente utilizada en múltiples ámbitos de la industria y los servicios. Particularmente en la clasificación de frutas su empleo es creciente, debido a su bajo costo relativo y objetividad.

Los sistemas de inspección visual automática (Cámara-PC-Software) realizan tareas de adquisición, procesamiento y análisis de imágenes para la determinación de los descriptores visuales que definen la calidad de un producto agrícola y/o alimenticio. Tienen la ventaja de realizar un análisis no destructivo. Por ello, los algoritmos que realizan el cálculo de los descriptores correspondientes a partir de una imagen están en permanente revisión para mejorar tanto la velocidad de cálculo como reducir el error en sus resultados.

La limitación de este método es que mediante el mismo no se puede analizar ni el aroma ni el sabor y textura interna de la fruta, aunque las características que se obtienen del análisis del aspecto están fuertemente correlacionadas con los mismos: un color adecuado indica el grado de maduración, y por lo tanto un adecuado sabor y textura interna. La ausencia de manchas indica un producto sano y fresco.

LINEAS DE INVESTIGACION Y DESARROLLO

A partir de una imagen color de una escena donde están en posición fija relativa un número determinado de frutas de la misma especie claramente diferenciadas del fondo, lo que se busca básicamente es desarrollar algoritmos que aplicados a ese tipo imagen permitan determinar descriptores morfológicos (tamaño y forma) y de aspecto (textura, color y manchas) que permitan realizar un adecuado reconocimiento y evaluar la calidad del producto.

Los ejes temáticos sobre los que se está trabajando son básicamente:

- 1) La segmentación de la imagen. Es un tema sensible porque presenta varios niveles de dificultad:
 - a. Diferenciación entre el fondo y los objetos de interés.
 - b. Individualización de cada objeto de interés.
 - c. En cada objeto de interés se debe detectar características propias.
- 2) Los descriptores morfológicos bajo análisis son: firma, alto, ancho, factor de forma 1, área, perímetro, factor de forma 2, desviación cuadrática media, distancia significativa, superficie de revolución y volumen de revolución. Este conjunto de descriptores se evalúa para cada una de las frutas de la escena y el objetivo es reducir significativamente el tiempo de cálculo de los mismos. Para ello se está empleando solo datos obtenidos de la firma.
- 3) Los descriptores de aspecto bajo análisis son: textura (descriptores que surgen de la matriz de coocurrencia) color (análisis de las propiedades por banda color), manchas (reconocimiento de la región manchada y calculo porcentaje manchado en la cara visible).
- 4) Construcción de una base de conocimiento a partir de los descriptores morfológicos y de aspecto, de modo que cada vez que una imagen nueva se procese, incorpore y actualice los datos en forma eficiente.
- 5) Construcción de un sistema experto que use la base para realizar la evaluación de la calidad de cada una de las frutas presentes en la imagen.

RESULTADOS OBTENIDOS/ESPERADOS

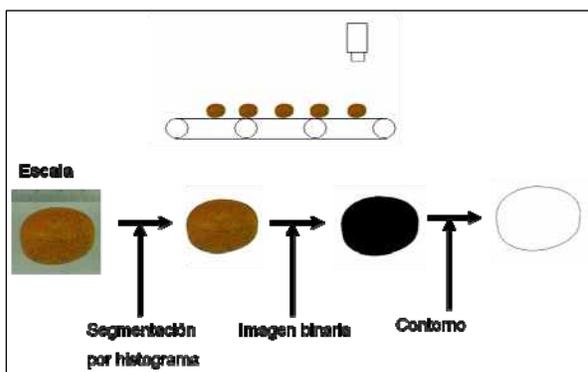


Fig 1: Esquema y secuencia de las operaciones de preprocesamiento

Determinación y validación de los descriptores morfológicos. Algunos resultados alcanzados

El esquema de la figura 1 muestra la forma en que se hicieron el registro de las imágenes y las operaciones de preprocesamiento que permiten obtener el contorno del producto, que consiste en los píxeles del borde de la imagen binaria. Su ancho es de un píxel. Se desarrollaron los algoritmos que toman como dato el contorno para calcular la Firma¹, y a partir de ella los siguientes descriptores:

1. Área.
2. Perímetro.
3. Alto.
4. Ancho.
5. Factores de forma.
6. Superficie de revolución.
7. Volumen de revolución.

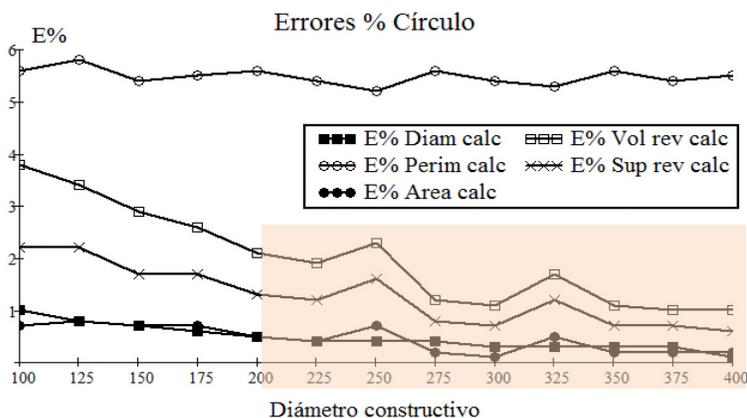


Gráfico 1 Representación de los errores porcentuales de los descriptores en función del diámetro constructivo en píxeles

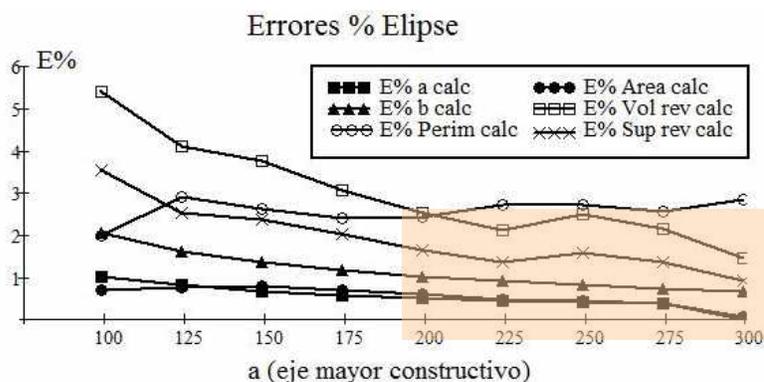


Gráfico 2: Representación de los errores porcentuales de los descriptores en función del eje mayor constructivo en píxeles

Para evaluar el cálculo de los descriptores mencionados, se construyó con un editor dos conjuntos de imágenes de dimensiones conocidas. El primero está compuesto de imágenes con un círculo negro sobre fondo blanco y el segundo conjunto con una elipse negra sobre un fondo blanco. Estos conjuntos pueden considerarse imágenes binarias de frutas de forma esférica o esferoidal. Los descriptores de tamaño de estas figuras

¹ La firma de una fruta es la representación en cartesianas de las coordenadas polar del contorno referidas al baricentro

geométricas se calculan en forma analítica a partir de sus valores constructivos, y se emplean como valores verdaderos para calcular el error del método.

Aplicando los algoritmos desarrollados a las imágenes de las figuras se calcularon los descriptores. Los mismos se emplean como valores medidos en el cálculo del error del método.

En el Gráfico 1 se representan los errores porcentuales del diámetro, perímetro, área, volumen y superficie de revolución en función del diámetro constructivo en píxeles, para las figuras circulares.

En el Gráfico 2 se representan los errores porcentuales en función del eje mayor a en píxeles, para las figuras elípticas. Se observa que los errores porcentuales del método disminuyen al aumentar a . Para todos los descriptores, si $a > 100$ píxeles el $|E\%|$ se mantiene por debajo del 5.5 %, y si $a > 200$ píxeles por debajo del 3.3%.

Para evaluar la precisión del método en la determinación de alto a y ancho b , se compararon los valores calculados procesando las imágenes de naranjas con los correspondientes valores medidos con un calibre de apreciación 0.02 mm. Los resultados se representan en el Grafico 3.

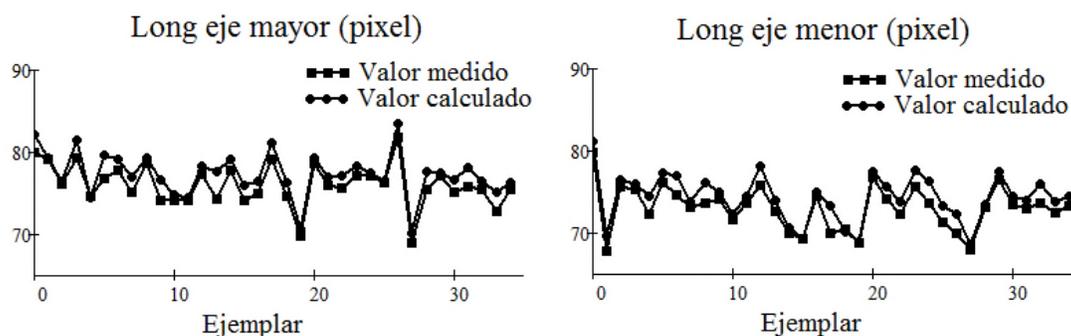


Gráfico 3: Representación de los valores medidos y calculados de las longitudes de los ejes mayor (izquierda) y menor (derecha) en función de los ejemplares.

Construcción del vector característico.

Consideramos un conjunto de M ejemplares de un producto y se quiere seleccionar de los mismos, aquellos que cumplan con determinados requerimientos de forma y tamaño. Al efecto se realizan las siguientes operaciones:

- a) Ajuste y normalización de la firma.
- b) Cálculo de la firma promedio.

Con estos datos se obtiene la desviación cuadrática media (DCM) y la distancia significativa (DS), que son descriptores de tipo estadísticos relacionados con la forma.

Para cada ejemplar de un producto se construye un vector característico formado por los siguientes elementos: DCM, DS, área, perímetro, factor de forma1 ($\text{Área} / \text{Perímetro}^2$), alto, ancho, factor de forma 2 ($\text{alto} / \text{ancho}$), superficie y volumen. No todos los elementos de este vector característico son linealmente independientes entre si. Es posible seleccionar una terna de , los menos correlacionados, y con ellos construir un espacio de clasificación de tres dimensiones. En este espacio los requerimientos determinan una región. El candidato que pertenezca a la región cumplirá con la exigencia solicitada.

Desarrollo del próximo objetivo del proyecto

La determinación y validación de los descriptores de aspecto (textura, color y manchas).

FORMACION DE RECURSOS HUMANOS

El equipo de trabajo consta de tres investigadores categorizados estables y tres becarios de grado y uno posgrado.

La relación con otros grupos del país se realiza a través de presentaciones a Congresos (JAIIO 2008 JII- Agroindustria) y vinculación a través de convenios institucionales.

BIBLIOGRAFIA

1. Duda R. O. and Hart P. E. Pattern Classification and Scene Analysis, John Wiley and Sons, 1973.
2. González Rafael C y Woods E. Richard Tratamiento Digital de Imágenes. Addison-Wesley / Diaz de Santos (1996).
3. Computer vision systems for rapid quality inspection of agricultural and food products. Wang Lu, Sun Da-Wen. ICETS2000 - Session 6: Technology Innovation and Sustainable Agriculture.
4. Artificial intelligence in agriculture. Editorial Computers and Electronics in Agriculture 40 (2003) 1 _ 3.
5. Improving quality inspection of food products by computer vision: a review. Jadhg Brosnan, Da-Wen. Journal of Sun Food Engineering 61 (2004) 3–16.
6. Machine Vision System for Automatic Quality Grading of Fruit. J. Blasco, N. Aleixos and E. Moltó Biosystems Engineering Volume 85, Issue 4 , August 2003, Pages 415 – 423
7. Citrus sorting by identification of the most common defects using multispectral computer vision. J. Blasco a, N. Aleixos, J. Gómez, E. Molto. Accepted 12 March 2007 Journal of Food Engineering
8. Computer vision detection of peel defects in citrus by means of a region oriented segmentation algorithm. J. Blasco, N. Aleixos, E. Molto. Journal of Food Engineering 81 (2007) 535–543
9. A classification system for beans using computer vision system and artificial neural networks. Kivanc Kılıc, Ismail Hakki Boyacı, Hamit Koksel, Ismail Kusmenoglu. Journal of Food Engineering 78 (2007) 897–904.
10. A simple digital imaging method for measuring and analyzing color of food surfaces. Kit L. Yam, Spyridon E. Papadakis. Journal of Food Engineering 61 (2004) 137–142.
11. Comparison of three algorithms in the classification of table olives by means of computer vision. R. Diaz, L. Gil, C. Serrano, M. Blasco, E. Molt, J. Blasco. Journal of Food Engineering 61 (2004) 101–107.
12. Computer vision detection of peel defects in citrus by means of a region oriented segmentation algorithm. J. Blasco, N. Aleixos, E. Molto. Journal of Food Engineering (2007).
13. Development of a computer vision system to measure the color of potato chips. Franco Pedreschi, Jorge Leo'n, Domingo Mery, Pedro Moyano. Food Research International 39 (2006) 1092–1098.
14. Low cost sensor for volume and surface area computation of axi-symmetric agricultural products. Ta Yuan Wang, Sing Kiong Nguang. Journal of Food Engineering 79 (2007) 870–877.
15. Segmentation of colour food images using a robust algorithm. Domingo Mery, Franco Pedreschi. Journal of Food Engineering 66 (2005) 353–360

Medición de diámetro pupilar ocular

Arturo Bianchetti

Coordinación de Innovación Técnica – Gerencia de Sistemas y Telecomunicaciones (GSyT) – ANSES
abianchetti@gmail.com

CONTEXTO

El presente trabajo, forma parte de la línea de investigación sobre biometría de la Coordinación de Innovación Técnica, dependiente de Gerencia de Sistemas y Telecomunicaciones (GSyT) de ANSES.

RESUMEN

La pupila natural del ojo humano suele ser aproximadamente circular y, en determinadas condiciones externas, su diámetro depende del sujeto y de su estado psicofísico. Para un dado sujeto, el patrón de aberraciones, la difracción, la profundidad de campo y la iluminación retiniana dependen del diámetro pupilar, el cual a su vez varía en función de la iluminación ambiente. El conocimiento de este diámetro es importante tanto en pruebas objetivas como subjetivas en las cuales se determina la calidad visual. La medición del diámetro pupilar también se utiliza como indicador del nivel de vigilia o anestesia o para detectar la presencia de narcóticos. En este trabajo se presenta un software desarrollado en MATLAB (denominado AS08) capaz de procesar imágenes y de medir el diámetro de una pupila circular. Se consideran imágenes registradas dentro del espectro visible y también en el infrarrojo cercano. En el software AS08 se utilizan varias técnicas de procesamiento de imágenes, entre las cuales cabe destacar la transformada de Hough, por sus conocidas virtudes referentes al reducido consumo de recursos computacionales para detectar círculos.

Palabras clave: diámetro pupilar, estado psicofísico, calidad visual, filtrado espacial, transformada de Hough.

1. INTRODUCCION

Cuando se determina la calidad visual mediante pruebas objetivas (aberrometría ocular y corneal, medición de la función de punto extendida, etc.) o subjetivas (agudeza visual, deslumbramiento, sensibilidad al contraste, etc.) suele ser de interés¹⁻⁸ conocer el diámetro pupilar ocular pues este afecta las aberraciones, la difracción, la profundidad de foco, la iluminación retiniana, etc. La determinación del diámetro pupilar también es importante en ciertos tratamientos para corregir la visión como es el caso de la cirugía refractiva (Lasik). Para identificar pacientes que no son aptos para la cirugía, se determina su diámetro pupilar en un cuarto oscuro⁹⁻¹¹ y aquellos que presentan diámetros grandes son advertidos de que corren un mayor

riesgo de padecer de distorsiones visuales nocturnas permanentes tales como halos, imágenes múltiples y pérdida de sensibilidad al contraste. Por otra parte, aunque algunos sujetos pueden tener pupila elíptica o padecer de coloboma o deformaciones de iris causadas por cirugías, la mayoría de las pupilas pueden aproximarse a circulares.

2. LINEAS DE INVESTIGACION y DESARROLLO

En este trabajo presentamos el software AS08¹² el cual, basándonos en Ref.13 pero realizando algunas adaptaciones, desarrollamos en MATLAB para detectar la posición y el diámetro de una pupila ocular circular presente en una imagen. El software puede procesar imágenes en formato RGB convencional o en escala de grises, registradas en el espectro visible o en el infrarrojo cercano. Adicionalmente AS08 contiene un proceso de auto-calibración que permite medir el diámetro pupilar no solo en píxeles sino también en milímetros. Esto brinda una considerable ventaja respecto a otros dispositivos capaces de medir el tamaño pupilar solamente en píxeles tales como el Eyetracker de Arrington Research. El software AS08 puede operar en modo autónomo (AU) o en modo de comandos extendidos (EX) según lo desee el usuario. En modo AU no hay intervención del usuario por lo que AS08 puede integrarse a procesos automatizados mientras que en modo EX el usuario elige algunos parámetros con lo cual facilita el procesamiento de la imagen.

AS08 contiene varios módulos de pre-procesamiento para mejorar y adecuar la imagen ocular adquirida al módulo de detección de círculos basado en la transformada de Hough.

La imagen ocular que se presenta en la parte izquierda Figura 1 presenta bajo contraste y los niveles de gris presentes en ella corresponden a un rango dinámico reducido.

Para mejorar la calidad de una imagen de rango dinámico reducido realizamos una expansión del mismo. Caracterizamos el nivel de intensidad en todos los puntos de la imagen al finalizar este proceso mediante la función de expansión de contraste¹³, $g(x, y)$, definida mediante:

$$g(x, y) = \frac{255}{(J_{MAX} - J_{MIN})} \cdot (f(x, y) - J_{MIN}) \quad (1)$$

Luego de aplicar la expansión, se utiliza al máximo el rango dinámico del dispositivo de

presentación (0-255). La imagen resultante se muestra en la Figura 1 y, debido a que se ha aumentado la separación de los niveles de gris que la componen, hay una mejoría de contraste considerable.



Figura 1 – Expansión de contraste de la imagen SC1.

Aunque en una imagen, aprovechando al máximo el rango dinámico del dispositivo de presentación, J tome valores entre 0 y 255, en varios casos es necesario incrementar el contraste, dependiendo de las características de la imagen, para niveles de gris oscuros (cercanos a 0) o claros (cercanos a 255). Para ello, suele utilizarse¹³ una función de corrección $gamma$, $G[f(x, y)] = g(x, y)$, tal que $g(x, y)$ es la distribución de niveles de gris en todos los puntos de la imagen obtenida al finalizar este proceso. Se tiene¹³

$$g(x, y) = \frac{255}{(255)^{\gamma}} \cdot (f(x, y))^{\gamma} \quad (2)$$

En la Figura 2 se muestra el resultado de aplicar corrección de contraste y luego corrección gamma a una imagen capturada en el infrarrojo extraída de CASIA¹⁴.

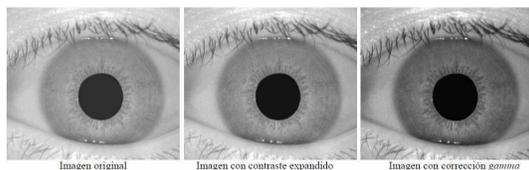


Figura 2 – Expansión de contraste y corrección $gamma$ de la imagen IR1.

Además, la imagen suele tener ruido que es necesario reducir mediante algún tipo de filtro para que las etapas posteriores del procesamiento funcionen adecuadamente. En AS08, el ruido se reduce empleando *filtrado espacial* con la máscara de promediado que se muestra en la Figura 3.

$$\frac{1}{9} \times \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array}$$

Figura 3 – Máscara de promediado

En la Figura 4 mostramos el efecto que causa el filtro de promediado en la imagen real IR1 con un agregado de ruido. Se puede apreciar que la reducción de ruido se hace a costa de una pérdida en el nivel de detalle pero esto no afecta al proceso de detección pupilar ocular.

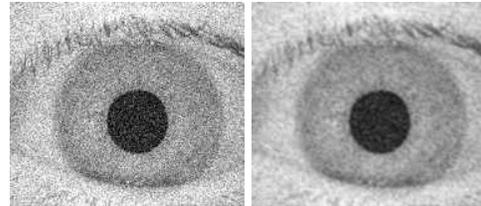


Figura 4 - Aplicación de máscara de promediado a imagen IR1 extraída de CASIA.

Dado que solamente se desea conocer la posición y el diámetro pupilar, se realiza un umbralizado para remover información irrelevante de la imagen que resulta del suavizado. El umbralizado limita los niveles de gris posibles en cada píxel de manera que, mientras que en la imagen de entrada cada píxel puede tener 256 niveles de gris diferentes (8 bits), a la de salida solo puede tener dos (1 bit). Todos los niveles de gris superiores a un nivel de gris umbral, que denotamos T , son transformados al valor 255 y los inferiores al valor 0.

El valor de T se elige de manera tal que la imagen umbralizada contenga información suficiente para detectar la pupila pero excluya información irrelevante que podría interferir en la detección. En la Figura 5 mostramos el perfil de niveles de gris en función de la coordenada x para la imagen IR1.

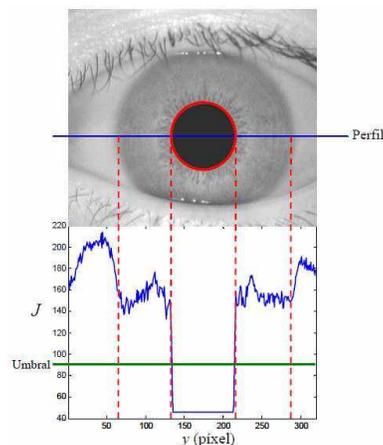


Figura 5 - Perfil de nivel de intensidad de la imagen IR1.

En general para una imagen ocular registrada tanto dentro del espectro visible como en el infrarrojo cercano, el área de la pupila corresponde a una de las áreas de menor nivel de gris. Un nivel umbral adecuado debe ser mayor que el nivel de gris de la pupila y menor que el del iris. Mientras más oscuro sea el iris más parecido será su nivel de gris a

la pupila y más difícil será escoger un valor de umbral adecuado. En base a pruebas realizadas con AS08, si J_{MAX} e J_{MIN} son los valores máximo y mínimo de J teniendo en cuenta toda la imagen (la cual ya ha sido suavizada de manera que se ha eliminando ruido que podría interferir en la determinación de J_{MAX} y J_{MIN}) y si definimos un coeficiente de umbralizado, que denotamos P y es tal que $0 < P < 1$, entonces evaluamos T mediante:

$$T = J_{MIN} + (J_{MAX} - J_{MIN}) \times P \quad (3)$$

donde J_{MAX} e J_{MIN} corresponden a valores máximos y mínimos de intensidad y definimos P como el coeficiente de umbralizado con $0 < P < 1$.

En modo AU, AS08 considera $P = 0.1$ mientras que en modo EX, el usuario puede escoger varios valores de P hasta encontrar un umbral adecuado. En la Figura 6, mostramos el resultado del umbralizado para la imagen de la Figura 5.

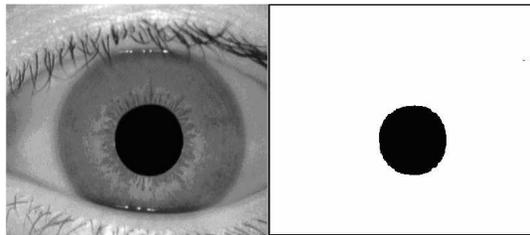


Figura 6 - Umbralizado de la imagen IR1.

Con el objetivo de simplificar aún más la imagen sobre la cual se va a realizar la detección de círculos, utilizamos¹³ la conocida máscara de Laplace para detección de bordes. El procedimiento es similar al de promediado pero los pesos de la máscara son diferentes (Figura 7). Esta máscara tiene la propiedad de resaltar las transiciones o bordes y eliminar las zonas de nivel de gris constante.

$$\frac{1}{9} \times \begin{array}{|c|c|c|} \hline -1 & -1 & -1 \\ \hline -1 & 8 & -1 \\ \hline -1 & -1 & -1 \\ \hline \end{array}$$

Figura 7 - Máscara de Laplace.

El resultado de aplicar la máscara de Laplace a la imagen IR1 umbralizada, se muestra en la Figura 8. La imagen umbralizada presenta una transición de J más abrupta (por lo tanto más simple) que la imagen de ejemplo presentada en la Figura 5. La imagen a la salida solo contiene un número reducido de píxeles en con niveles de gris o muy altos o muy bajos, es decir que puede contener más de dos niveles de gris. Para que poder detectar círculos en esta imagen utilizando la transformada de Hough debemos

binarizar la imagen, es decir limitar la imagen a solo dos niveles de gris posibles, 0 y 255.

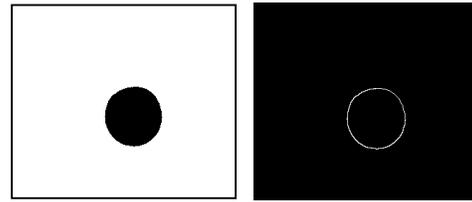


Figura 8 - Resultado de la aplicación de la máscara de Laplace a la imagen IR1. Panel izquierdo: imagen a la salida del umbralizado. Panel derecho: imagen a la salida de la aplicación de la máscara

El procedimiento de detección de círculos lo hemos desarrollado adaptando la metodología presentada en Ref.13 para detección de líneas rectas a detección de círculos. En este trabajo se hace uso de la conocida ecuación paramétrica de un círculo de centro $(x, y) = (a, b)$ y radio R :

$$(x - a)^2 + (y - b)^2 = R^2 \quad (4)$$

La transformada de Hough opera del siguiente modo:

I. Tomando como entrada la imagen binaria obtenida al finalizar el pre-procesamiento (que en el caso ideal contiene una circunferencia ubicada en el lugar donde está la pupila), encontramos todos los píxeles de interés que son tales que tienen $J_{IMAGEN}=1$ y coordenadas (x_q, y_q) (con $q=1, 2, \dots, Q$).

II. Teniendo en cuenta un radio inicial $r_{min} = 1$, aplicamos la transformada de Hough de manera que para cada punto (x_q, y_q) del espacio imagen, en el espacio de parámetros se tiene un círculo de radio 1 y, considerando los Q puntos del espacio imagen, en el espacio de parámetros se tienen Q círculos que se intersectan en algunos puntos del plano (a, b) dando lugar a máximos absolutos o no de J_{HOUGH} (Figura 9).

III. Repetimos el procedimiento descrito en el paso II para cada uno de los radios r_k de interés, siendo $1 \leq r_k \leq 60$, y evaluamos el nivel de gris $J_{MAX,k}$ correspondiente al máximo (absoluto o no) de J_{HOUGH} . Los resultados obtenidos para cada r_k se guardan en diferentes capas del array acumulador.

IV. Considerando todas las capas del array acumulador, que se muestra en la buscamos el píxel del plano (a, b) que contiene el máximo valor de $J_{MAX,k}$ y denominamos r_{PICO} al valor de r_k en esta capa. Las coordenadas (a_{PICO}, b_{PICO}) de este píxel en el plano (a, b) coinciden con las coordenadas del centro del círculo en el plano (x, y) y el radio r_{PICO} correspondiente a esta capa es el radio del círculo detectado (Figura 10).

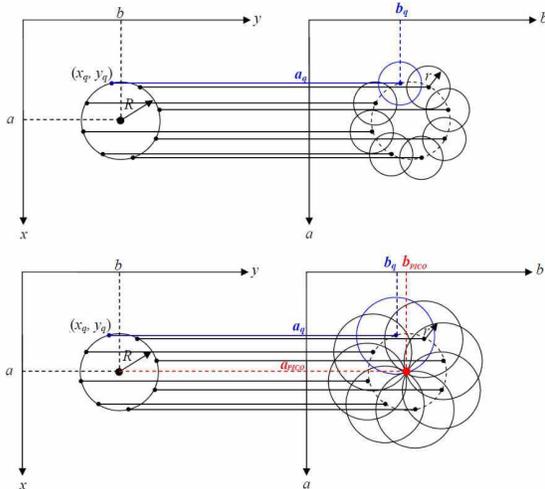


Figura 9 - Arriba: transformada de Hough para $r < R$.
Abajo: transformada de Hough para $r = R$.

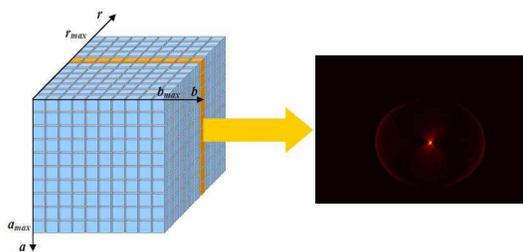


Figura 10 - Panel izquierdo: array acumulador de la transformada de Hough. Panel derecho: transformada de Hough de la imagen de Laplace de la Figura 8.

En la Figura 11 se muestra el resultado de la detección pupilar para la imagen IR2. Se puede apreciar que la detección es satisfactoria pero no perfecta debido a que la pupila no es estrictamente circular.

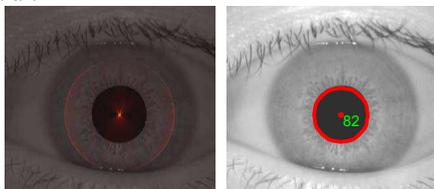


Figura 11 - Detección pupilar exitosa para la imagen IR2. Panel derecho: superposición de imagen original y transformada de Hough. Panel izquierdo: salida de AS08.

El resultado presentado en la Figura 11 corresponde a diámetro en píxeles. Para convertir este resultado a milímetros se utiliza un sticker de diámetro conocido que se utiliza de referencia y que también es detectado por AS08 utilizando la transformada de Hough. Luego utilizando ec.(5) es posible presentar el diámetro detectado en milímetros,

$$D_{pupila} = d_{pupila} \cdot \left(\frac{D_{calibración}}{d_{calibración}} \right) \quad (5)$$

donde D_{pupila} es el diámetro en milímetros de la pupila, d_{pupila} es el diámetro en píxeles de la pupila

$D_{calibración}$ es el diámetro conocido del sticker y $d_{calibración}$ es el diámetro medido en píxeles del sticker. En la Figura 12 se muestra la detección pupilar y de sticker para la imagen SC2, en este último caso el resultado se presenta en milímetros.

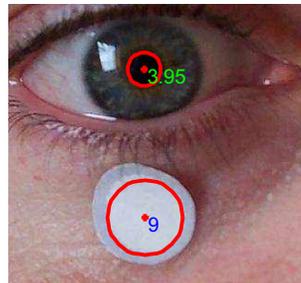


Figura 12 - Detección de diámetro pupilar en milímetros.

3. RESULTADOS OBTENIDOS/ESPERADOS

En el presente trabajo se implementa un software desarrollado en Matlab (denominado AS08) que, a partir de una imagen conteniendo una pupila ocular circular, puede detectar eficientemente la posición de la pupila y medir su diámetro tanto en píxeles como en milímetros. El software puede operar en modo autónomo (AU) o en modo de comandos extendidos (EX) según lo desee el usuario. La ventaja del modo AU es que no requiere esfuerzo del usuario y su desventaja es que, para imágenes con contraste pobre o con regiones oscuras no correspondientes a la pupila (tales como sombras ocasionadas por la ubicación de la fuente de iluminación), puede fallar la detección. Por otro lado, el modo EX requiere que el usuario esté dispuesto a ingresar una serie de parámetros pero tiene la ventaja de se puede medir el diámetro pupilar para imágenes con condiciones de contraste pobres o con sombras. Para cualquier imagen capturada, la probabilidad de detectar con éxito la región pupilar es mayor cuanto mayor sea el contraste visual iris-pupila.

Para determinar el diámetro pupilar, el software AS08 utiliza la transformada de Hough, la cual al ser apta solo para patrones circulares, no permite detectar satisfactoriamente pupilas elípticas, con coloboma o con deformaciones causadas por cirugías. Esta es la principal limitación del software aunque la mayoría de las pupilas del ojo humano en condiciones cotidianas son casi circulares.

4. FORMACION DE RECURSOS HUMANOS

Se está articulando horizontalmente con la Universidad de Buenos Aires.

5. BIBLIOGRAFIA

1. S.A.Comastri, G.Martin and T.Pfortner "Analysis of pupil and corneal wave aberration data supplied by the SN CT 1000 topography system", Optik (International Journal for Light and Electron

- Optics, Elsevier, Amsterdam) 117 (11), 537-545 (2006).
2. S.A.Comastri, R.Echarri and T.Pfortner, "Correlation between visual acuity and pupil size", 5th Iberoamerican Meeting on Optics and the 8th Latin American Meeting of Optics, Lasers and their Applications, Isla Margarita-Venezuela 3-8 October 2004, Proceedings of SPIE (International Society for Optical Engineering, Washington) 5622,1341-1346, (2004).
 3. P. L. Kaufman, F. H. Adler, A. Alm, Fisiología del ojo: Aplicación clínica, Elsevier España, 2003.
 4. S.A.Comastri, L.I.Perez, G.D.Pérez, G.Martin and K.Bastida, "Zernike expansion coefficients: rescaling and decentering for different pupils and evaluation of corneal aberrations", Journal of Optics A: Pure and Applied Optics (Institute of Physics Publishing, London) 9, 209-221 (2007).
 5. E. Colombo, S.A. Comastri, L. Issolio, R. Echarri, "Pupil light reflex produced by glare under mesopic adaptation", Journal of Light & Visual Environment (the Illuminating Engineering Institute of Japan, Tokyo) 31 (2), 70-79, (2007).
 6. S.A.Comastri, Liliana I.Perez, Gervasio D.Pérez, Karina Bastida, and G. Martin, "Wavefront aberrations: analytical method to convert Zernike coefficients from a pupil to a scaled arbitrarily decentered one", 6th Iberoamerican Meeting on Optics and 9th Latin American Meeting of Optics, Lasers and their Applications, Campinas-Brasil 21-26 October 2007, Edited by N.U.Wetter and J.Frejlich (American Institute of Physics, New York), Proceedings American Institute of Physics 978, 69-74 (2008)
 7. S.A. Comastri, K. Bastida, G. Martin y A. Bianchetti, "Aberrometrías oculares y de otros sistemas ópticos: transformación de coeficientes Zernike al contraer y desplazar transversalmente la pupila", Edición Universidad de Belgrano (Buenos Aires) Documento de Trabajo No 208,1-17, 2008 http://www.ub.edu.ar/investigaciones/dt_nuevos/208_comastri.pdf
 8. S.A. Comastri, G.D. Pérez, G. Martin, A. Bianchetti, "Aberraciones Zernike corneales para diferentes pupilas: precisión en los parámetros y aplicaciones", Óptica Pura y Aplicada (Journal Sociedad Española de Óptica, Madrid). Aceptado septiembre 2008.
 9. L.J. Maguire, "Keratorefractive surgery, success, and the public health", A.M. J. Ophthalmol, March 1994, (3):177-394.
 10. J.J. Salz, W. Trattler, "Pupil size and corneal laser surgery", A.M. J. Ophthalmol, August 2006, (4):373-9.
 11. M.J. Mannis, W.A. Segal, J.K. Darlington, "Making sense of refractive surgery in 2001: why, when, for whom, and by whom?", Mayo Clinic Proc., August 2001, 76(8):823-9.
 12. Bianchetti, Arturo y Comastri, Silvia A. (2008). Desarrollo de una metodología para medir el diámetro pupilar ocular a partir del procesado de imágenes conteniendo el ojo. Documento de Trabajo N° 221, Universidad de Belgrano. http://www.ub.edu.ar/investigaciones/dt_nuevos/221_bianchetti.pdf
 13. R.C. Gonzalez, R. E. Woods, Digital Image Processing, Second Edition, Prentice Hall (2001).
 14. Chinese Academy of Sciences (CASIA) <http://www.cbsr.ia.ac.cn/english/Databases.asp>

Interfaces No Convencionales. Su impacto en las interacciones

Sergio Martig – Silvia Castro - Martín Larrea – Sebastián Escarza
Maximiliano Escudero – María Luján Ganuza - Dana Urribarri
José Schneider – Damián Flores Choque
{ srm, smc, mll, se }@cs.uns.edu.ar

Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab)

Departamento de Ciencias e Ingeniería de la Computación

Universidad Nacional del Sur

Tel. 0291-4595135 Fax 0291-4595136

Bahía Blanca, CP 8000, Argentina

Contexto

El trabajo se lleva a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur. En el Laboratorio se están desarrollando distintos proyectos que abordan distintos aspectos vinculados con Visualización, Modelado, Interfaces, Interacciones, entre otros y se pretende que, a mediano plazo, se puedan integrar los resultados obtenidos en el marco del proyecto actual.

Resumen

La comunidad en Interacción Humano Computadora (HCI) ha mostrado que las Interfaces de Usuario están trascendiendo los escritorios. Estas interfaces emergentes emplean técnicas novedosas para el ingreso de los datos, ejemplo de esto lo constituyen las *interfaces hápticas, las tangibles, las basadas en cámaras* entre otras. Por otro lado la presentación de información se realiza a través de una variedad de dispositivos cada vez más amplia y tendientes a embeberse en el entorno del usuario. Todo esto demanda una profunda revisión y adecuación de los paradigmas de interacción tradicionales y seguramente la formulación de nuevos modelos que potencien su utilidad.

Palabras clave: *Visualización, Interacción Humano Computadora, Interfaces No Convencionales*

1. Introducción

Es sabido que la manera en que las personas interactúan con las computadoras está fuertemente determinada no solo por la tecnología de base disponible, sino también por las herramientas y facilidades para construir soluciones sobre dicha tecnología. Las interacciones basadas en gráficos surgieron casi una década antes de la aparición de la Xerox Star y de que la Macintosh proveyera un conjunto de herramientas que puso a la programación de tales interfaces al alcance del programador promedio, junto con la provisión de un ambiente en el cual dichas soluciones pudieran interoperar efectivamente. La Internet existió por décadas antes de que la World Wide Web proveyera un conjunto de herramientas interoperables que hizo que la interacción con la red fuera posible para los programadores promedio y no solo para los especialistas como era hasta ese momento.

La tecnología computacional ha cambiado radicalmente desde los '80. Las computadoras personales han incrementado en varios órdenes de magnitud su velocidad de procesamiento y su capacidad de almacenamiento. Virtualmente todas se

encuentran conectadas a la Internet. La revolución en la tecnología ha generado una amplia gama de dispositivos de procesamiento de información, a la PC de escritorio se le suman dispositivos móviles, capacidades en constante aumento en la telefonía celular, además surgieron nuevos dispositivos de Entrada/Salida que permiten que la computadora trascienda el escritorio embebiéndose en el ambiente del usuario. Los modelos de computación interactiva creados en los '80 no aprovechan las capacidades computacionales actuales. La tecnología actual nos permite asegurar que nuevos paradigmas de interacción no son solamente requeridos, sino también viables.

Para lograr avances significativos en esta dirección se hace necesaria una profunda revisión de los estilos de interacción tradicionales, evaluando su adecuación ya no solo a la nueva tecnología disponible sino también a las expectativas que esta ha generado en el usuario. Como consecuencia de este proceso pueden surgir nuevos estilos de interacción que permitan a través de un uso eficiente de los recursos computacionales lograr las metas del usuario de forma efectiva.

2. Líneas de Investigación y Desarrollo

En la última década han surgido un gran número de nuevos paradigmas que incluyen, en distinto grado, elementos físicos aumentados o potenciados por poder computacional. Las interfaces Tangibles (TUI), las interfaces de realidad aumentada (AR), las de realidad mixta (MR), entre otras, son algunos ejemplos de diseños que plantean formas no convencionales de interacción con información digital. De estos abordajes surgen como cuestiones relevantes a considerar: las tecnologías de displays, el tracking, temas de diseño de interfaces propiamente dicho, diseño de las interacciones y aspectos que tienen que ver con el realismo.

En este contexto a partir de una revisión de las distintas propuestas de interfaces no convencionales, poniendo especial énfasis

en los aspectos tecnológicos y en las cuestiones propias de las interacciones soportadas. Se revisarán y adecuarán los paradigmas de interacción existentes y potencialmente se definirán nuevos estilos para lograr interacciones efectivas

Adicionalmente esperamos poder diseñar interfaces que logren el objetivo de trascender el escritorio utilizando tecnología de bajo costo, que permita su diseño y experimentación en nuestras instalaciones.

A partir del relevamiento de distintas interfaces no convencionales, que trascienden el escritorio con mayor o menor éxito, se pretende:

- Analizar los aspectos tecnológicos de las mismas enfocándonos en los requerimientos de hardware y en los desafíos que implican su implementación.
- Detectar las interacciones soportadas por las distintas interfaces, revisando la validez de los distintos estilos de interacción y procediéndose a su adecuación en caso de ser necesario.

Basándonos en los resultados del relevamiento antes mencionado se buscarán las maneras de lograr interfaces de este tipo utilizando tecnología de bajo costo, dentro de nuestro alcance, para poder implementarlas y lograr de esta manera obtener un entorno en el cual validar las conclusiones del análisis de los paradigmas de interacción.

3. Objetivos

Como objetivos generales:

- Se analizarán interfaces no convencionales con el propósito de poder determinar los requerimientos tecnológicos y las interacciones soporta
- En ese contexto, se revisarán los paradigmas de interacción tradicionales y se adecuarán para hacer un uso eficiente de los recursos y satisfacer las expectativas y necesidades de los usuarios.

- Se diseñarán e implementarán interfaces que trasciendan el escritorio con tecnologías de bajo costo que permitan validar experimentalmente las conclusiones obtenidas.

Como objetivos particulares se estudiará la aplicación de estas interfaces en dos áreas donde la interacción juega un rol esencial.

Las aplicaciones de visualización constituyen, en este caso, un espacio de análisis y experimentación muy rico; pues las propias visualizaciones se convierten en las interfaces que posibilitan que el usuario, a través de las interacciones adecuadas, pueda explorar el espacio de información subyacente logrando el insight buscado de los datos.

Otro campo donde las interacciones tienen un rol preponderante es el de los juegos, en este campo tanto la visualización como las interacciones se integran naturalmente en la interfaz, abriendo todo un abanico de posibilidades y desafíos tendientes a lograr que la comunicación hombre-máquina se establezca de una manera natural y efectiva.

4. Formación de Recursos Humanos

En lo concerniente a la formación de recursos humanos se detallan las tesis en desarrollo y los cursos relacionados con el tema del proyecto dictados por los integrantes del grupo de investigación:

4.1 Tesis en Desarrollo

4.1.1 Tesis de Doctorado en Ciencias de la Computación

- Sergio Martig. Tema: *Interacción en Visualización de Información*. Dirección: Dra. Silvia Castro.
- Martín Larrea. Tema: *Visualización basada en Semántica*. Dirección: Dra. Silvia Castro.
- Sebastián Escarza. Tema: *Visualización de Ontologías*. Dirección: Dra. Silvia Castro.

- Dana Urribarri. Tema: *Escalabilidad Visual*. Dirección: Dra. Silvia Castro.
- Maximiliano Escudero. Tema: *Modelos de Terrenos para GIS Móvil*. Dirección: Dra. Silvia Castro.
- María Luján Ganuza. Tema: *Servicios Web en Visualización de Información*. Dirección: Dra. Silvia Castro – Lic. Sergio Martig

4.1.2 Tesis de Magister en Ciencias de la Computación

- José Schneider. Tema: *Realidad Espacial Aumentada*. Dirección: Dra. Silvia Castro – Lic. Sergio Martig

4.2 Cursos de pre y posgrado relacionados con el tema del proyecto dictados por integrantes del grupo de trabajo.

4.2.1 Cursos de Pregrado

- **Computación Gráfica** Materia optativa para los estudiantes de la Licenciatura en Ciencias de la Computación y obligatoria para los de Ingeniería en Sistemas de Computación. Universidad Nacional del Sur.
- **Comunicación Hombre-Máquina** Materia obligatoria para los alumnos del Profesorado en Computación. Universidad Nacional del Sur.
- **Interfaces Gráficas** Materia optativa para los estudiantes de la Licenciatura en Ciencias de la Computación y de la Ingeniería en Sistemas de Computación. Universidad Nacional del Sur.
- **Introducción a la Visualización** Materia optativa para los estudiantes de la Licenciatura en Ciencias de la Computación. Universidad Nacional del Sur.

4.2.2 Cursos de Posgrado

- **Sistemas de Modelamiento de Volúmenes** Materia del Posgrado en Ciencias de la Computación. UNS.

- **Tópicos avanzados en Curvas y Superficies** Materia del Posgrado en Ciencias de la Computación. UNS.
- **Computación Gráfica: Tópicos Avanzados.** Departamento de Informática y Estadística de la Facultad de Economía y Administración. Universidad Nacional del Comahue.
- **Modelamiento Multirresolución.** Departamento de Informática y Estadística de la Facultad de Economía y Administración. Universidad Nacional del Comahue.
- **Visualización** Materia del Posgrado en Ciencias de la Computación. UNS.
- **Visualización Científica** Materia del Posgrado en Ciencias de la Computación y del Magíster en Computación Científica. UNS.
- **Visualización de Información** Materia del Posgrado en Ciencias de la Computación. UNS.
- **Tópicos Avanzados en Visualización de Información** Materia del Posgrado en Ciencias de la Computación. UNS.
- **Interacción Humano-Computadora** Materia del Posgrado en Ciencias de la Computación y del Magíster en Computación Científica. UNS.
- **Modelado Geométrico Multirresolución de Superficies.** Materia del Posgrado en Ciencias de la Computación. UNS y UNLP.

5. Bibliografía

- 1 ARTag. <http://www.artag.net>
- 2 Baecker, R. M. and Buxton, W. A. S. , *Readings in Human-Computer Interaction*. San Mateo CA.: Morgan Kaufmann Publishers, 1995.
- 3 Baeza-Yates, R., Ribeiro-Neto, B., *Modern Information Retrieval*, Addison Wesley, 1999.
- 4 K.W. Brodlie, D.A. Duce, J.R. Gallop, J.P.R.B. Walton, J.D. Wood, *Distributed and Collaborative Visualization*, The Eurographics 2004
- 5 Card, S., Mackinlay, J., Shneiderman, B., *Readings in Information Visualization – Using Vision to Think*, Morgan Kaufmann, 1999.
- 6 Chi, Ed H. *A Framework for Information Visualization Spreadsheets*. Ph.D. Thesis. University of Minnesota, Computer Science Department. March, 1999.
- 7 Derthick, M., Roth, S., *Data Exploration across Temporal Contexts*, Proceedings of Intelligent User Interfaces (IUI'00), pp. 60-67, New Orleans, January 2000.
- 8 Dix, A., Finlay, J., Abowd, G., Beale, R., *Human-Computer Interaction*, Prentice Hall Europe, Second Edition, 1998.
- 9 Johnson, Ch. & Hansen, Ch., Eds., *The Visualization Handbook*, Elsevier Butterworth-Heinemann, 2005, ISBN: 0-12-387582-X.
- 10 Greenberg, S. & Flichett, Ch., Phidgets: Easy development of Physical Interfaces through Physical Widgets, UIST'01, 209-218, 2001.
- 11 Ishii, H. & Ullmer, B., *Tangible Bits: Towards seamless interfaces between people, bits and atoms*. Proc. ACM CHI'97, 234-241, 1997.
- 12 Martig, S., Castro, S., Fillottrani, P. & Estévez, E , *Un Modelo Unificado de Visualización*. Proceedings, pp. 881-892, 9º Congreso Argentino de Ciencias de la Computación. 6 al 10 de Octubre de 2003. La Plata. Argentina.
- 13 Nielsen, J., *Usability Engineering*, Morgan Kaufmann, 1993.
- 14 Raskin, J., *The Human Interface*, Addison Wesley, ACM Press, 2000.

15 Shneiderman, B., *Designing the User Interface*, Addison-Wesley Publishing Company, 1998.

16 Spence, R., *Information Visualization*, Addison-Wesley, ACM Press, 2001.

17 Stuart M. Charters, Nicolas S. Holliman, Malcolm Munro, *Visualisation on the Grid: A Web Service Approach*, 2005.

18 Tufte, E.R., *Envisioning Information*, Cheshire, CT Graphics Press, 1990.

19 Tufte, E.R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, CT Graphics Press, 1997.

20 Ware, C., *Information Visualization. Perception for Design*, Morgan Kaufmann Publishers, 2004, ISBN: 1-55860-819-2.

Además se pueden consultar las siguientes publicaciones periódicas:

- ACM SIGGRAPH Proceedings
- ACM Transactions on Graphics
- IEEE Transactions on Visualization and Computer Graphics
- IEEE Computers Graphics and Applications
- IEEE Visualization Proceedings
- IEEE Information Visualization Proceedings
- ACM Proceedings of the CHI (Computer Human Interface)

Recuperando el detalle y evaluando la calidad visual en GPU de modelos topográficos simplificados

María V. Cifuentes^{a,b}, Juan P. D'Amato^{a,c}, Lucas Lo Vercio^a, Marcelo J. Vénere^{a,d}

^aPLADEMA, Universidad Nacional del Centro, 7000 Tandil, Argentina,
(cifuentes,jpdamato}@exa.unicen.edu.ar)

^bComisión de Investigaciones Científicas de la Provincia de Buenos Aires

^cConsejo Nacional de Investigaciones Científicas

^dCNEA

Palabras Clave: Computer Graphics, Simplificación Topográfica.

Resumen. *La visualización interactiva de escenarios 3D es clave en la industria de entretenimientos, en multimedia y en otras aplicaciones de computer graphics. A la hora del rendering, el volumen y la complejidad de la información que define a cada ambiente resulta problemático. Por lo tanto, es necesario recurrir a técnicas que reduzcan los tiempos de rendering tales como la simplificación de modelos, el nivel de detalle adecuado para ciertas porciones del modelo, el uso de primitivas eficientes que aceleren el hardware gráfico, el balance entre CPU-GPU, entre otras. En este trabajo, presentamos una línea de investigación que propone un conjunto de técnicas las cuales han sido desarrolladas para ofrecer una alta performance del rendering de grandes terrenos en tiempo real.*

1 Introducción

Actualmente, para incrementar el realismo de las aplicaciones gráficas, librerías tales como OSG (<http://www.openscenegraph.org/projects/osg>) u OpenInventor (<http://oss.sgi.com/projects/inventor/index.html>), y motores de juegos como Torque o CryEngine incorporan diferentes técnicas de simplificación basadas en otorgar cierto nivel de detalle a cada parte del modelo, adaptando el número de polígonos de los objetos a su importancia dentro del escenario. Resultando modelos aproximados que son ampliamente empleados en aplicaciones de computación gráfica debido a que reducen la cantidad de información geométrica enviada al sistema gráfico. No obstante, estas estrategias tienen algunas deficiencias en la calidad visual resultante, al perderse ciertas características de las mallas tales como la apariencia, y exigen además un alto uso de CPU para su procesamiento.

Básicamente, nuestra propuesta de investigación intenta balancear la relación CPU-GPU migrando parte del procesamiento a la GPU para agilizar los tiempos de rendering consumidos en la representación de modelos topográficos simplificados.

2 Simplificación de Modelos Topográficos

La literatura brinda métodos [1, 2, 3] y métricas [4,5] para la simplificación poligonal de superficies. Algunos, muy difundidos [6, 7] usan técnicas iterativas basadas en

contracción de arcos para simplificar los modelos y mantener el error de aproximación usando métricas cuadráticas. Además, existe otra variedad de algoritmos de simplificación aplicables a casos característicos.

En particular, los algoritmos que reducen el volumen de información original de terrenos han constituido nuestra principal motivación. Básicamente, el proceso consiste en construir aproximaciones que conservan la similitud visual original del terreno aunque representado a través de unos pocos polígonos. Indicadores de error basados en la geometría del objeto, la ubicación de la cámara, la dirección de vista y máscaras binarias de visibilidad construidas mediante algoritmos de *ray-casting*, son algunas de las estrategias estudiadas e implementadas [8]. Actualmente, nos motiva la búsqueda de nuevos indicadores y/o criterios de simplificación basados en principios similares a los aplicados en el cálculo de modelos de iluminación conocidos. Por ejemplo, la simplificación guiada por la divergencia de los vectores normales asociados a cada triángulo resulta una alternativa. Por otra parte, se busca mejorar la performance aprovechando las capacidades de las unidades de procesamiento gráfico actuales.

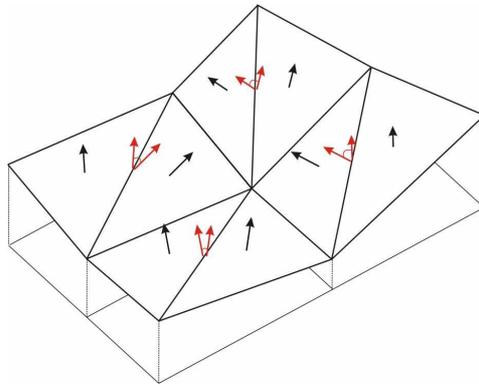


Figura 1- Normales y Divergencia correspondientes a una grilla regular de 4 elementos.

A partir del terreno original, el algoritmo de simplificación propuesto construye una jerarquía quadtree de mallas aproximadas de la que se pueden extraer aproximaciones de resolución variada. En una aplicación interactiva, la selección del nivel de detalle se realiza en tiempo de ejecución por lo que es necesario establecer un balance entre el número de polígonos con los que se representará el objeto y la cantidad de tiempo necesario para visualizarlo.

3 Criterios de similitud

Una evaluación de cuánto se parecen dos mallas, es medir el volumen encerrado entre la original y la simplificada. Para esto, se utiliza el volumen del prisma encerrado entre el centro del triángulo simplificado y el centro del triángulo original. En los casos de grillas regulares, el algoritmo es muy eficiente con un costo constante de cálculo por elemento, por acceder directamente a las posiciones de los triángulos originales. El siguiente pseudocódigo esquematiza el proceso:

para todo nodo terminal R_{aprox} **hacer**

para todo triángulo $\hat{t} \in R_{orig}$ **hacer**

determinar el centroide $p_c(x_c, y_c, z_c)$ de \hat{t}

identificar el triángulo $\hat{T} \in R_{aprox}$ que contiene al centroide p_c

calcular la componente z_{aprox} de p_c en $\hat{T} \in R_{aprox}$ (interpolación)

calcular $z_c - z_{aprox}$

acumular la diferencia de volumen para medir el error

fin para

fin para

Otra estrategia, aplicable a mallas de superficie no orientadas no regulares es la propuesta por [9] basada en una discretización del espacio mediante una grilla bidimensional y donde luego se computan intersecciones entre rayos provenientes de dicha grilla y los polígonos de las mallas comparadas. El volumen de diferencia se calcula como la suma de las distancias entre los triángulos intersecados.

Para el caso particular de mallas de altura, es posible realizar el mismo cálculo utilizando el buffer de profundidad, provisto por las GPU. Esto implica una mejora drástica de la performance, siendo el costo resultante el de recorrer los píxeles de dicho búffer. Esta estrategia, actualmente en evaluación, se encuentra limitada por la precisión de las GPU's y se requiere tener mucho cuidado en la medición de los datos.

4 Balance CPU-GPU

En cada frame, la malla aproximada extraída desde la jerarquía quadtree es enviada a la placa gráfica para su renderizado. OpenGL brinda primitivas especiales para renderizar tiras de cuadrantes, pero el número de vértices necesarios puede superar al usado por las tiras de triángulos. Debido a que la performance de las placas gráficas está supeditada a la cantidad de datos intercambiados entre CPU-GPU, se opta por triangular los cuadrantes y luego conectarlos en tiras de triángulos para reducir la cantidad de datos a transmitir, acelerando la etapa de rendering. Las tiras de triángulos mejoran notablemente los tiempos comparadas a las listas de triángulos, pero encontrar un conjunto adecuado de tiras de triángulos para el modelo aproximado [10] no es trivial y menos aún si es dinámico.

La idea es lograr un modelo aproximado con la calidad suficiente, pero utilizando la menor cantidad de polígonos, aprovechando además que las GPU trabajan eficientemente con texturas. Las GPU actuales permiten programar el pipeline de renderizado, a nivel de vértice y de textura, a través de un lenguaje gráfico denominado GLSL. Con estas capacidades, se propone una estrategia similar a un *bumpmapping* para reducir la cantidad de polígonos y al mismo tiempo mantener la apariencia de la malla.

En las mallas de superficie que analizamos, nos encontramos casos muy característicos, donde pequeñas variaciones de un vértice, ya sea en altura o en la normal, promueven la

generación en el árbol de muchos elementos, en general innecesarios. Estas pequeñas variaciones son los que dan un aspecto rugoso a la superficie, visible al aplicar el modelo de iluminación correspondiente. Es posible eliminar dichos vértices, si se permite un mayor error de volumen, pero con la consiguiente pérdida del aspecto aunque preservando las características principales como el lecho del río y los sectores más elevados (ver figura 2).

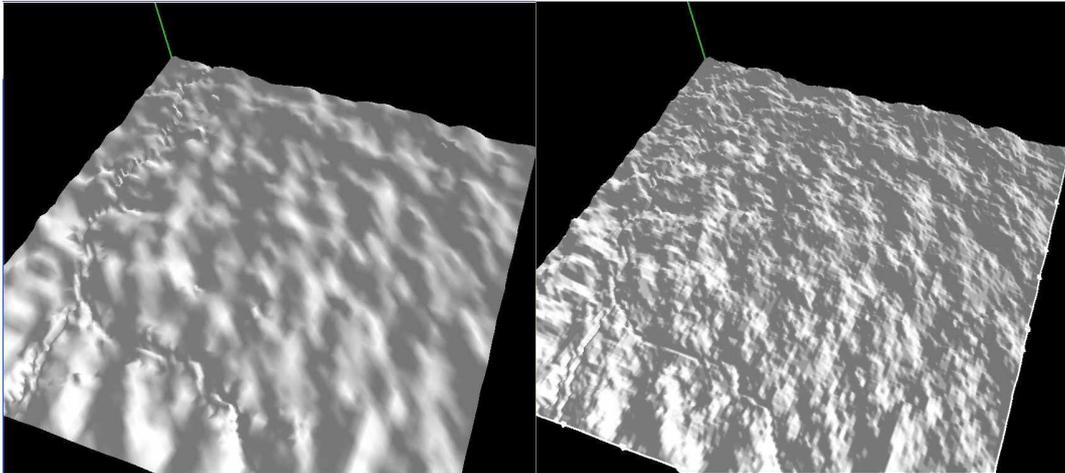


Figura 2- Terreno con baja y alta resolución. Se mantienen , izquierda y derecha respectivamente.

La propuesta para estos casos, es mantener el aspecto visual de la malla utilizando una aproximación de baja resolución combinada con la información asociada a los vectores normales del modelo original, permitir una cierta pérdida de volumen sin desmerecer la calidad visual de su apariencia. El proceso implica almacenar la información de las normales de la malla original codificada en una textura, sabiendo que esta información es inalterable se reduce la transferencia CPU-GPU. Al momento de la simplificación, se relaja la condición de volumen; es decir, se aplica un criterio de planaridad adecuado a la superficie que elimina las irregularidades de la malla (picos o serruchos). Finalmente al renderizar, en la etapa por píxel, el modelo de iluminación se calcula en base a las normales de la malla original. En la figura 3 se puede observar como aparecen los detalles sobre la superficie cuando se aplica el mapa de normales (ver figura 3). Esta estrategia brinda la facilidad adicional de no recalcular las normales en el modelo simplificado.

5 Conclusiones y Trabajos Futuros

La idea es continuar ahondado en la búsqueda de técnicas y estrategias de simplificación y algoritmos eficientes de rendering. Por un lado, construir tiras de triángulos que reduzcan la cantidad de vértices que se envían a la placa y que eviten la existencia de triángulos degenerados. Sumado a esto, se está estudiando un criterio unificado para medir la similitud de la malla resultante, basado en la medición de volumen y en el cambio visual.

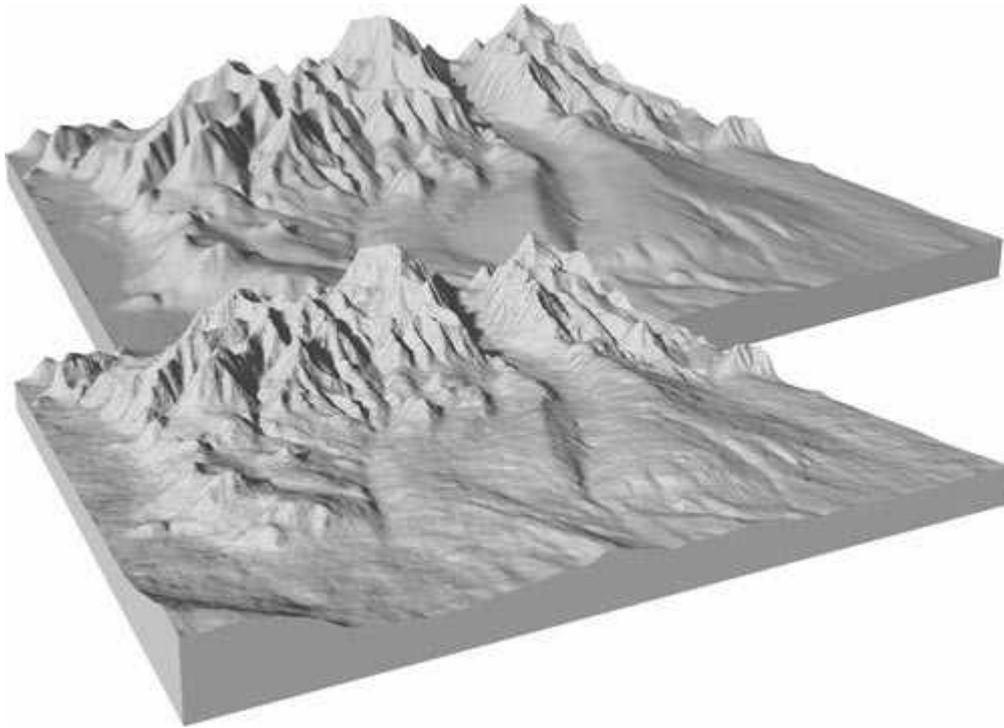


Figura 3 – Rendering de un terreno de baja resolución. Abajo se aplica el mapa de normales.

Referencias

- [1] D.P. Luebke. A developer's survey of polygonal simplification algorithms. *IEEE Computer Graphics and Applications*, 3(24):24–35, 2001.
- [2] P. Heckbert and M. Garland. Survey of polygonal surface simplification algorithms. Technical report, Multiresolution Surface Modeling Course Notes of SIGGRAPH'97, 1997.
- [3] Pai-Feng Lee, Bin-Shyan Jong, Point-based Simplification Algorithm, *WSEAS Transactions on Computer Research*, vol 1(3), pp. 61-66, 2008.
- [4] P. Castelló, M. Sbert, M. Chover, M. Feixas, Techniques for Computing Viewpoint Entropy of a 3D Scene, *ICCS 2006*, University of Reading, UK, May, 2006.
- [5] R. Alvarez, L. Tortosa, J. F. Vicent, A. Zamora, Error measurements and parameters choice in the GNG3D model for mesh simplification, *WSEAS Transactions on Information Science & Applications*, vol. 5 (5), pp. 579-588, 2008.
- [6] E. Puppo and R. Scopigno. Simplification, lod and multiresolution - principles and applications. *Tutorial Notes of EUROGRAPHICS'99*, 16(3), 1997.
- [7] M. Garland and P. Heckbert. Surface simplification using quadric error metrics. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pp. 209–216. ACM Press/Addison- Wesley Publishing Co., 1997.
- [8] M.V. Cifuentes, J.P. D'Amato, C. Garcia Bauza, M.J. Vénere, Simplificación de modelos topográficos, *IX Workshop de Investigadores en Ciencias de la Computación*, Trelew, Argentina, 2007.
- [9] Juan D'Amato, Mariana del Fresno, Marcelo Javier Vénere, Un indicador de calidad para evaluar superficies segmentadas", *ENIEF*, 2008.
- [10] M. Shafae, R., Pajarola, Dstrips: Dynamic Triangle Strips for Real-Time Mesh Simplification and Rendering. *Proceedings Pacific Graphics Conference 2003*.

Tratamiento de imágenes digitales y video. Visión y Reconstrucción 3D.

Claudia Russo, Federico Cristina, Sebastián H. Dapoto, Javier Vegas,
Verónica Artola, Marcelo Naiouf, Javier Giacomantone, Oscar Bria, María José Abásolo

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

{crusso, fcristina, sdapoto, jvegas, vartola, mnaiouf, jog, onb}@lidi.info.unlp.edu.ar, mjabasolo@uib.es

CONTEXTO

Esta línea de Investigación forma parte del Subproyecto “Tratamiento de imágenes digitales y video. Visión 3D”, dentro del Proyecto “Algoritmos Distribuidos y Paralelos. Aplicación a Sistemas Inteligentes y Tratamiento Masivo de Datos” del Instituto de Investigación en Informática LIDI, acreditado por la UNLP en el marco del Programa de Incentivos. Asimismo se integra a un proyecto financiado por la Agencia Española de Cooperación Internacional (AECI).

RESUMEN

La línea de investigación presentada se centra en el estudio y desarrollo de temas relacionados con software para tratamiento de imágenes y video, reconstrucción 3D y diseño y programación de interfaces gráficas en 3D, especialmente aplicados a problemas de tiempo real. Son estudiadas técnicas de análisis de tipologías edilicias en imágenes aéreas. Los temas estudiados abarcan aspectos clásicos referidos al tratamiento de imágenes tales como detección de bordes, segmentación de imágenes, reconocimiento de patrones, estimación de movimientos y flujo óptico, cálculos de trayectoria, velocidad y aceleración de objetos. En el área de visión 3D, el énfasis está puesto en el estudio, investigación y desarrollo de aplicaciones en campos como la arquitectura y la arqueología, donde resulta muy útil contar con métodos rápidos y flexibles para adquirir información 3D a partir de una escena real. El diseño de am-

bientes 3D es utilizado como interfaz gráfica de otras aplicaciones valiéndose del uso de Realidad Virtual.

Palabras Clave: Imagen, Video, Reconstrucción 3D, Procesamiento de Imágenes, Reconocimiento de Patrones, Visión, Interfaz gráfica, reconstrucción 3D.

1. INTRODUCCION

El procesamiento digital de señales (en particular el tratamiento de imágenes) adquiere un interés especial ya que constituye la base para muchas aplicaciones importantes [1]. Una señal es una variable de una o más dimensiones que toma valores de acuerdo a otra variable, como por ejemplo el tiempo o el espacio. Las señales pertenecientes al mundo real son continuas. Sin embargo, una computadora digital si bien puede trabajar con señales reales no puede hacerlo con señales continuas. Necesariamente las señales deben ser muestreadas y digitalizadas, y por lo tanto se convierten en una secuencia de números [2][3]. Una imagen es un caso particular de señal que puede ser digitalizada en forma de matriz.

El procesamiento y el análisis de imágenes digitales nace en el momento en que se dispone de recursos tecnológicos para captar y manipular, en forma de matrices de valores, gran cantidad de información espacial [2][3].

Muchas disciplinas entre las que se encuentran el reconocimiento de patrones en tiempo real, el tratamiento y transmisión de video en tiempo real y visión por computa-

dora, requieren un importante esfuerzo en la investigación de algoritmos paralelos aplicables en áreas tales como: robótica, industria manufacturera, ingeniería forestal y medicina. Esto motiva la investigación y el desarrollo en temas relacionados con software para tratamiento de señales, especialmente en problemas de tiempo real.

Cuando el objetivo se centra en extraer y clasificar objetos que aparecen en la imagen se pueden utilizar varias técnicas de análisis, entre ellas las de reconocimiento estadístico de patrones. En particular, en bioinformática y en imágenes médicas es necesario evaluar sistemas de clasificación binaria [22], para lo cual se utilizan técnicas que facilitan el diseño y la visualización comparativa de los resultados como las curvas ROC (Receiver Operating Characteristics) [23]. Estas técnicas representan en muchos casos una solución apropiada a problemas complejos del mundo real, como el tratamiento de imágenes funcionales obtenidas mediante resonancia magnética nuclear (fMRI), que permiten determinar la actividad del cerebro humano ante estímulos predefinidos en el paradigma del correspondiente estudio médico [23].

Puede decirse que el procesamiento de imágenes como tal comienza en los años 1950-1960 debido a la combinación de dos hechos: por una parte la aparición de las computadoras digitales y por otra los *Programas Espaciales*. El objetivo inicial era muy concreto: mejorar la calidad visual de dichas imágenes. En este contexto se ha avanzado en el estudio de técnicas de eliminación de ruido, “blurring”, entre otras [5]. De esta manera, aparecen los métodos de restauración de imágenes que utilizan filtros inversos. Con las técnicas de eliminación de ruido y restauración se desarrollan las denominadas técnicas de mejora de imágenes, aunque tanto la eliminación de ruido como la restauración pueden entenderse como técnicas de mejora.

Asimismo, ha habido un creciente interés en el estudio del problema de la restauración de súper-resolución de secuencias de video. Mientras que en el problema tradicional de resolución se dispone de una única imagen de entrada, en la tarea de la súper-resolución se trabaja a partir de múltiples imágenes submuestreadas y degradadas, y de esta manera se puede tomar ventaja de datos espacio-temporales adicionales disponibles en la secuencia de video [2][3]. Un problema común e importante que surge en las comunicaciones visuales es la necesidad de crear una secuencia de imágenes de video con resolución mejorada a partir de un stream de video fuente de menor resolución [5][6][7][25].

Un tema más, que se suele considerar parte del procesamiento, es la compresión de imágenes. Puede citarse que el comienzo de la compresión de imágenes se sitúa en 1950 cuando se solicita la patente del código predictivo. Sin embargo, surge la pregunta de por qué se puede considerar la compresión parte del procesamiento de imágenes. La razón es simple: por una parte la compresión era, al menos inicialmente, una forma de procesamiento orientada a que las imágenes ocupasen menos espacio para transmisión y almacenamiento. Han surgido otras aplicaciones como la eliminación de artificios en imágenes y video comprimidos, un campo en continua expansión [8]. Uno de los aspectos más sorprendentes cuando se estudia la percepción humana es la capacidad del observador para determinar la estructura 3-D (Visión 3D) de los objetos a partir de patrones bidimensionales de luz [9][10].

Hoy en día existe un interés por comprender la capacidad de los algoritmos para poder derivar mediciones tridimensionales de alta precisión. Un tema actual es el de Reconstrucción 3D a partir de imágenes, generalmente capturadas por adquisidores ubicados en distintas posiciones. Una aplicación de esta línea de investigación es la

reconstrucción de piezas como por ejemplo antropológicas, dentarias, etc. [11].

Los dispositivos escáner 3D permiten obtener modelos 3D de objetos, esculturas, edificios, yacimientos arqueológicos, zonas urbanas y paisajes naturales. Utilizan medios ópticos (como por ejemplo láser) proyectados en el objeto de interés. Actualmente existen en el mercado diversos equipos de escáner 3D pero el costo de los mismos es muy elevado y su portabilidad es en muchos casos reducida. La estereovisión o visión estereoscópica es un método de reconstrucción 3D a partir de fotografías que no requiere de ningún equipo especial sino que puede funcionar con cámaras de fotos estándar. Cuando se intenta evaluar una aproximación computacional para la percepción artificial de formas 3D es necesario tener en cuenta dos hechos. Por una parte, que existen numerosos atributos de la estructura 3D que potencialmente podrían estar representados en el sistema visual (curvatura, distancia relativa, orientación local, entre otras) cuyas dificultades computacionales no son las mismas, y por otra, que para la evaluación de las diferentes aproximaciones computacionales es necesario examinar la validez de las hipótesis subyacentes. Puesto que existen numerosas escenas que producen la misma imagen, todos los análisis computacionales de la percepción 3D deben restringirse a un conjunto de posibles interpretaciones suponiendo una serie de restricciones más o menos reales [12].

Ciertamente los modelos de los objetos contienen más información de la que normalmente los sensores pueden extraer de una imagen. Por ello se suele trabajar con un modelo parcial que se puede extraer tanto del modelo del objeto como de la imagen observada. En función de ese modelo parcial es posible realizar el reconocimiento [13][14].

Es de interés en el área de procesamiento de imágenes el tratamiento de video en

tiempo real para el análisis de información proveniente de la trayectoria de los objetos capturados [15][16][17][18][24].

El volumen de datos y el computo intensivo requerido motivan el estudio de la paralelización de las aplicaciones.

2. LINEAS DE INVESTIGACION y DESARROLLO

- Captura y procesamiento de señales.
- Análisis de señales en tiempo real.
- Tratamiento de imágenes. Técnicas para la adquisición, mejorado y segmentación.
- Análisis de formas. Parametrización.
- Extracción de características invariantes a rotación y escala.
- Técnicas de análisis de tipologías edilicias. Segmentación y clasificación.
- Visión estereoscópica.
- Reconstrucción de modelos 3D basado en la combinación de cámaras estereoscópicas y la utilización de luz estructurada.
- Técnicas de estimación de movimientos, cálculos de trayectoria, velocidad y aceleración de objetos.
- Interfaz gráfica 3D para el diseño de modelos de simulación. Realidad Virtual.
- Herramientas de Realidad Virtual. Plug-In y API para Java. Diseño y programación de Interfaces Gráficas en 3D.
- Compresión de imágenes y video. Algoritmos para tratamiento de imágenes y video comprimido en tiempo real.
- Migración de algoritmos a arquitecturas multiprocesador y soluciones paralelas.

3. RESULTADOS OBTENIDOS /ESPERADOS

- Desarrollar soluciones a problemas específicos de procesamiento de imágenes y video, en particular en tiempo real [26].
- Se han desarrollado prototipos específicos en el ámbito del entrenamiento deportivo, como el análisis de la trayectoria de un balón [24][26].
- Se encuentra en etapa de desarrollo la reconstrucción 3D del tiro capturado en un ambiente de simulación. Desarrollos realizados para la Empresa CN Sports.
- Se ha desarrollado un sistema de reconocimiento automático de texto Braille.
- Se desarrollo de un sistema, hardware y software, que permite la reconstrucción de modelos 3D basado en la combinación de cámaras estereoscópicas y la utilización de luz estructurada [27][19].
- Desarrollar una herramienta de Realidad Virtual. Plug-In y API para Java. Diseño y programación de Interfaces Gráficas en 3D. La herramienta permitirá diseñar ambientes 3D para ser utilizados como interfaz gráfica de otras aplicaciones [20][21][28].
- Implementación de un scanner 3D móvil, hardware y software, mediante el uso de técnicas de estimación de desplazamiento y herramientas de visualización 3D para dispositivos móviles.
- Desarrollar una solución que permita detectar construcciones en imágenes aéreas de baja calidad. Esta investigación se lleva a cabo junto a la Facultad de Arquitectura (FAU-UNLP).

4. FORMACION DE RECURSOS HUMANOS

En esta línea de I/D se espera concluir una tesis doctoral en 2009, se encuentra en desarrollo una tesina sobre una herramienta

gráfica 3D de realidad virtual y esta prevista la realización de tesinas y tesis de postgrado en procesamiento de imágenes. Se han concluido como resultado de este proyecto las siguientes tesinas: Reconocimiento de texto braille, Segmentación de Imágenes Médicas y Reconstrucción 3D de trayectorias en entrenamiento deportivo.

5. BIBLIOGRAFIA

[1] Murat Tekalp, Digital Video Processing, Prentice Hall, 1995.

[2] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing (2nd Edition) (Hardcover), Prentice Hall; 2nd edition 2002).

[3] John C. Russ, The Image Processing Handbook, Fourth Edition (Hardcover), CRC Press; 4th edition 2002

[4] Kenneth R. Castleman, Digital Image Processing (Hardcover), Prentice Hall; 1st edition, 1995.

[5] Molina, R., Katsaggelos, A. K., y Mateos, J., "Multichannel Image Restoration using Compound Gauss-Markov Random Fields ", ICASSP 2000, 2000.

[6] Molina, R., Nuñez, J., Cortijo, F. y Mateos, J., "Image Restoration in Astronomy. A Bayesian Approach", IEEE Signal Processing magazine, 2001.

[7] Borko Furht, Joshua Greenberg, Raymond Westwater , "Motion Estimation Algorithms for Video Compression (The International Series in Engineering and Computer Science) (Hardcover)", Kluwer International, 1996.

[8] Peter Symes, "Digital Video Compression (with CD-ROM)(Paperback)", McGraw Hill, 2004.

[9] Yi Ma, Stefano Soatto, Jana Kosecka, S. Shankar Sastry , "An Invitation to 3-D Vision (Hardcover)", Springer, 2003.

[10] Trucco, Alessandro Verri, "Introductory Techniques for 3-D Computer Vision (Paperback)", Prentice Hall; 1998.

- [11] Emanuele Trucco, Alessandro Verri, *Introductory Techniques for 3D Computer Vision*, Prentice Hall – 1998.
- [12] Arturo de la Escalera, *Visión por Computador, Fundamentos y Métodos*, Prentice Hall, 2001.
- [13] Ramesh Jain, Rangachar Kasturi, Brian G. Schunck, *Machine Vision*, Mac Graw-Hill Internacional Edition, 1995.
- [14] Akihito Yamada, Yoshiaki Shirai and Jun Miura. *Tracking Players and a Ball in Video Image Sequence for Retrieving Scenes in Soccer Games*, Dept. of Computer-Controlled Mechanical Systems, Osaka University. 2-1, Yamadaoka, Suita, Osaka 565-0871, Japan, 2000.
- [15] Alvaro Castromán & Ernesto Copello. *Fútbol de Robots Uruguayo para Torneos*. Tesis de Grado de la Carrera de Ingeniería en Computación. Facultad de Ingeniería - Universidad de la República – 2004.
- [16] Jinchang Ren, James Orwell and Graeme A. Jones. *Digital Imaging, Estimating the Position of a Football from Multiple Image Sequences.*, Research Center, Kingston University. Penrhyn Road, Kingston upon Thames, Surrey, KT1 2EE, UK, 2003.
- [17] Jinchang Ren, James Orwell, Graeme Jones, Ming Xu. *Digital Imaging , Real-time 3D Soccer Ball Tracking from Multiple Cameras*, Research Centre, Kingston University. Surrey, KT1 2EE, U. K., 2004.
- [18] Xiao-Feng Tong, Han-Qing Lu, Qing-Shan Liu. *An Effective and Fast Soccer Ball Detection and Tracking Method*, National Lab. of Pattern Recognition, Inst. of Automation, Chinese Academy of Sciences – 2003.
- [19] Ron Kimmel, M. Bronstein, A. Bronstein. *"Numerical Geometry of Images : Theory, Algorithms, and Applications"*. Springer.
- [20] Daniel Selman. *" Java 3D Programming"*. Hanning.(A guide to key concepts and effective techniques).
- [21] Andrew Davison, “Killer Game Programming in Java”. O’Reilly. (Java Gaming & Graphics Programming)
- [22] Kim H.Y., Giacomantone J. O., Cho, Z. H. *Robust Anisotropic Diffusion to Produce Enhanced Statistical Parametric Map*, *Computer Vision and Image Understanding*, v.99, p.435-452 (2005).
- [23] Kim H.Y., Giacomantone J. O., *A New Technique to Obtain Clear Statistical Parametric Map by Applying Anisotropic Diffusion to fMRI*, *IEEE, International Conference on Image Processing. Proceedings, Genova, Italy*,v.3, p.724-727 (2005).
- [24] Federico Cristina, Sebastián H. Dapoto, Claudia Russo, Armando De Giusti, María José Abasolo. *Mobile path and spin 3D tracking and reconstruction*. *IV International Conference on Articulated Motion and Deformable Objects 2006*. AMDO 2006, Springer, Volumen: LNCS 4069. ISBN: 978-3-540-36031-5, ISSN:1611-3349, p.120-131.
- [25] Federico Cristina, Sebastián H. Dapoto, Claudia Russo, Oscar Bria. *High Resolution Images from Low Resolution Video Sequences*. *Journal of Computer Science & Technology*. Vol. 5, No 1, April 2005. ISBN: 1666-6038, p. 30-36.
- [26] Federico Cristina, Sebastián H. Dapoto, Claudia Russo. *A Lightweight Method for Computing Ball Spin in Real Time*. *Journal of Computer Science & Technology*. Vol. 7, No 1, April 2007. ISBN: 1666-6038, p. 34-38.
- [27] Federico Cristina, Sebastián H. Dapoto, Javier Vegas, Verónica Artola, Claudia Russo, María José Abásolo, Armando De Giusti. *3D Scanner Development with Stereoscopic Cameras and Laser Illumination*. *IADIS Computer Graphics and Visualization CGV 2008*. ISBN: 978-972-8924-63-8, p. 249-253.
- [28] Federico Cristina, Sebastián H. Dapoto, Javier Vegas, Verónica Artola. *Herramienta de realidad virtual para diseño de interfaces gráficas 3D en Java*. *XIV Jornadas de Jóvenes Investigadores AUGM 2007*.

Reconocimiento Estadístico de Patrones Máquinas de Soporte Vectorial y Series Temporales

Javier Giacomantone, Tatiana Tarutina, Verónica Artola, Armando De Giusti

Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – UNLP

{ jog, vartola, degiusti }@lidi.info.unlp.edu.ar, ttarutina@gmail.com

CONTEXTO

Esta línea de Investigación forma parte del Subproyecto “Tratamiento de imágenes digitales y video. Visión 3D”, dentro del Proyecto acreditado “Algoritmos Distribuidos y Paralelos. Aplicación a Sistemas Inteligentes y Tratamiento Masivo de Datos” del Instituto de Investigación en Informática LIDI. Asimismo se integra con proyectos de cooperación bilateral con universidades del exterior y con un proyecto financiado por la Agencia Española de Cooperación Internacional (AECID). El planteo que se presenta se ha definido en la segunda mitad de 2008 y constituye una línea incipiente en el III-LIDI.

RESUMEN

Este trabajo describe una línea de I/D y los resultados esperados de la misma. El objetivo principal es estudiar, desarrollar y evaluar sistemas de reconocimiento automático de patrones en modo supervisado y no supervisado. En modo supervisado el objetivo principal es optimizar la generalización del clasificador. En particular son estudiados problemas caracterizados por medio de series temporales y clasificadores basados en máquinas de soporte vectorial (SVM). Los principales temas abordados son la selección de características, las técnicas de agrupamiento, el análisis de métricas y los métodos de optimización en SVM.

Palabras Clave: Reconocimiento de Patrones. Clasificación no lineal. Máquinas de Soporte Vectorial. Selección y extracción de características. Series temporales.

1. INTRODUCCION

Reconocimiento de patrones y en particular reconocimiento estadístico de patrones es un área de investigación interdisciplinaria tanto en la investigación básica de métodos fundamentales [1][2][3], como en sus aplicaciones [4][5]. El objetivo principal de un sistema de reconocimiento automático de patrones es descubrir la naturaleza subyacente de un fenómeno u objeto, describiendo y seleccionando las características fundamentales que permitan clasificarlos en una categoría determinada. Sistemas automáticos de reconocimiento de patrones permiten abordar problemas en informática, en ingeniería y en otras disciplinas científicas, por lo tanto el diseño de cada etapa requiere de criterios de análisis conjuntos para validar los resultados [6][7]. Las principales áreas de aplicación son, reconocimiento remoto, reconocimiento óptico de caracteres y escritura manuscrita, identificación de patrones en imágenes médicas, sistemas de clasificación en bioinformática, sistemas de identificación biométrica y clasificación de series temporales. Un modelo general de un sistema automático esta constituido por tres etapas, sensor, selector de características y clasificador. La primera etapa puede ser considerada a su vez como la que trata de obtener la representación más fiel del fenómeno estudiado, y un módulo que permite extraer las características del mismo. La línea de investigación propuesta está enfocada en la segunda y tercera etapa. Los métodos utilizados en reconocimiento de patrones se dividen en dos grandes categorías clasificación supervisada y clasificación no supervisada. El tipo de objetos o fenómenos considerados en esta línea de trabajo pueden ser descriptos por un conjunto de características numéri-

cas que definen patrones en un espacio n -dimensional. Por lo tanto el análisis de las distribuciones estadísticas de cada clase y los métodos de estimación de parámetros, permiten definir estrategias de diseño, evaluar y especificar los métodos de clasificación. La línea de investigación propuesta en este trabajo esta enfocada en el diseño de clasificadores basados en SVM y en la aplicación a problemas caracterizados por series temporales y la clasificación contextual de las mismas.

1.1 Máquinas de Soporte Vectorial

Las máquinas de soporte vectorial (SVM) [8][9] son herramientas fundamentales en sistemas de aprendizaje automático, permitiendo el tratamiento de problemas actuales en reconocimiento de patrones y minería de datos tales como, reconocimiento y caracterización de texto manuscrito, detección ultrasónica de fallas en materiales, clasificación de imágenes médicas [10], sistemas biométricos [11], clasificación en bioinformática [12][13] y en física de altas energías [14]. Las SVM implementan reglas de decisión complejas, por medio de una función no lineal que permite mapear los puntos de entrenamiento a un espacio de mayor dimensión. En el nuevo espacio de características las clases son separadas por un hiperplano, siendo este el que maximiza la distancia entre el mismo y los puntos de entrenamiento. La distancia se denomina margen y esos vectores son los vectores de soporte. Las SVM cumplen un rol muy importante en teoría de aprendizaje estadístico y cuando es necesario entrenar un clasificador no lineal en un espacio de características de considerable dimensión con un número limitado de muestras. Podemos diferenciar dos aspectos importantes que en general reciben la denominación de máquinas de soporte vectorial, el uso de SVM en clasificación SVC y el uso de las mismas en regresión SVR [15]. La línea de investigación propuesta estudia ambos aspectos y en particular en el caso de clasificación mediante SVM tiene como objetivo diseñar sistemas con alta capacidad de generaliza-

ción. Entre las tendencias actuales podemos mencionar las investigaciones sobre SVM paralelas y secuenciales (PSVM, SSVM) [16].

1.2 Series Temporales

Una serie temporal es una secuencia de puntos medidos a intervalos sucesivos, normalmente de tiempo, y en general a intervalos regulares. Las series temporales son el resultado de medidas de distintos fenómenos físicos en la naturaleza pero también son comunes en econometría, marketing, control industrial y como resultado de métodos de monitoreo y diagnóstico en medicina. Fundamentalmente su caracterización se da en el dominio espacial o en el dominio transformado de Fourier [17], Wavelets[18], Chirplet [19] y sus técnicas de análisis son fundamentalmente estadísticas y de procesamiento de señales. La clasificación de las series temporales obtenidas a partir de estudios funcionales del cerebro, son un ejemplo de abordaje multidisciplinario, donde uno de sus aspectos fundamentales es el de reconocimiento de patrones. La aplicación, adaptación y adecuada selección de kernels de SVM a series temporales es un tema de investigación actual [20][21]. La clasificación básica de las máquinas de soporte vectorial es binaria por lo tanto es importante el estudio de la extensión a multi-clasificación [22]. Las series pueden ser unidimensionales o multidimensionales con correlación tanto temporal como espacial. En el último caso se plantea un problema de reconocimiento de patrones complejo y de minería de datos (MDTSC multi-dimensional time series classification). Entre los temas actuales de investigación que involucran los conceptos anteriores podemos citar estudios en neurociencias por medio de resonancia magnética funcional (fMRI), electroencefalogramas (EEGs) y magneto-encefalogramas (MEGs) [23][24]. Los problemas anteriores requieren computo numérico intensivo demandando la especificación y desarrollo de sistemas paralelos para su implementación [25][26][27], debi-

do fundamentalmente a la complejidad y el volumen de datos procesados.

2. LINEAS DE INVESTIGACION y DESARROLLO

- Clasificación supervisada. Discriminadores lineales y no lineales.
- Métodos de estimación de parámetros para clasificadores Bayesianos.
- Clasificación no supervisada. Técnicas de agrupamiento (clustering).
- Selección y extracción de características.
- Métricas, pseudométricas y distancias ultramétricas en clasificación supervisada, no supervisada y selección de características.
- Criterios de evaluación de desempeño en sistemas de clasificación automática.
- Criterios y algoritmos para combinación de clasificadores.
- Maquinas de soporte vectorial. Kernels y algoritmos de optimización.
- Clasificación de series temporales y clasificación contextual.
- Caracterización y evaluación de la capacidad de generalización de los clasificadores propuestos.
- Paralelización y análisis de complejidad de los algoritmos propuestos

3. RESULTADOS OBTENIDOS /ESPERADOS

- ✓ Desarrollar modelos y optimizar algoritmos particulares de clasificación supervisada y no supervisada.
- ✓ Evaluación de los métodos de análisis de desempeño y su aplicación sobre los clasificadores y conjuntos de datos propuestos.
- ✓ Obtener mejoras y adecuar las técnicas de selección y extracción para el tratamiento de datos en espacios multidimensionales
- ✓ Dada la naturaleza interdisciplinaria de una línea de investigación como el reconocimiento de patrones, en particular en las áreas de aplicación, promover la

integración entre las distintas líneas de investigación.

- ✓ Evaluar las técnicas propuestas sobre datos simulados y reales.
- ✓ Dada la naturaleza específica de las aplicaciones que implican cómputo intensivo para resolver las soluciones numéricas propuestas, transferir estos resultados para su investigación y posible tratamiento mediante técnicas de procesamiento paralelo y distribuido
- ✓ Transferir los resultados obtenidos, nuevas técnicas, algoritmos y tratamiento de datos experimentales de nivel fundamental a las áreas de aplicación principales.

4. FORMACION DE RECURSOS HUMANOS

En esta línea de I/D existe cooperación entre distintos subproyectos de investigación en el III-LIDI, fundamentalmente por la utilidad de los métodos estudiados para resolver problemas de clasificación en tratamiento masivo de datos, como una etapa fundamental en un sistema de visión por computador y por ser particularmente viables para su cómputo paralelo. En el marco de esta línea de investigación hay un investigador realizando su doctorado y se espera la realización de tesinas y tesis desarrollando aspectos particulares en sistemas automáticos de reconocimiento de patrones.

5. BIBLIOGRAFIA

1. Fukunaga K. "Introduction to Statistical Pattern Recognition". Second Edition. Academic Press, 1990.
2. Devijer P. A., Kittler, J. "Pattern Recognition, A Statistical Approach". Prentice Hall, 1982.
3. Batagelj V, Bock H, Ferligoj A. "Data Science and Classification". Springer, 2006.
4. Devijer P, Kittler, J. "Pattern Recognition: theory and applications". Springer, 1986.

5. Anke Meyer-Baese. "Pattern Recognition for Medical Imaging". Academic Press, 2004.
6. Kim H.Y., Giacomantone J. O., Cho, Z. H. Robust Anisotropic Diffusion to Produce Enhanced Statistical Parametric Map, Computer Vision and Image Understanding, v.99, p.435-452 (2005).
7. Kim H.Y., Giacomantone J. O., A New Technique to Obtain Clear Statistical Parametric Map by Applying Anisotropic Diffusion to fMRI, IEEE, International Conference on Image Processing. Proceedings, Genova, Italy, v.3, p.724-727 (2005).
8. Cortes C, Vapnik V, Support vector networks. Machine Learning v.20, p.273-297 (1995).
9. Vapnik, V. The Nature of Statistical Learning Theory. N. Y. Springer (1995)
10. S. Li, T. Fevens, A. Krzyzak, S. Li. Automatic Clinical Image Segmentation Using Pathological Modelling, PCA and SVM, MLDN, LNAI 3587 pp.314-324, (2005).
11. Z. Lei, Y. Yang, Z. Wu. Ensembles of Support Vector Machine for Text-Independent Speaker Recognition, IJCSNS v.6 n.5A pp. 163-167, (2006).
12. Y. Li, J. Li. Predicting Subcellular Localization of Proteins Using Support Vector Machines with N-Terminal Amino Composition, ADMA 2005, LNAI 3584, pp. 618-625, (2005).
13. R. Boekhorst, I. Abnizova, L. Wernich. Discrimination of regulatory DNA by SVM on the basis of over- and under-represented motifs, ESANN pp. 481-486 (2008).
14. Vossen Anselm. Support Vector Machines in High Energy Physics, CERN, Geneva, Switzerland, pp.23-33 (2005).
15. Vapnik, V. Golowich S., Smola A. Support Vector Method for Function Approximation, Regression, Estimation and Signal Processing. In Advances in Neural Information Processing Systems, Vol 9, pag. 281-287. MIT Press, Cambridge, 1997.
16. L. Wang, M. Chang, J. Feng. Parallel and Sequential Support Vector Machines for Multi-label Classification, International Journal of Information Technology, v.11 n.9 pp. 11-18, (2005).
17. Grafakos Loukas. Classical and Modern Fourier Analysis, Prentice Hall. (2004).
18. D. B. Percival, A. T Walden. Wavelet Methods for Time Series Analysis, Cambridge University Press (2000).
19. J. R. Cui, et al. Time frequency analysis of visual evoked potentials using chirplet transform. IEE Electronic Letters v.41 p.n.4 pp.217-218 (2005).
20. S. Rüping. SVM kernels for time series analysis, G1-Workshop-Woche Lernen-Lehren-Wissen-Adaptivitet, pp.43-50 (2001).
21. K. Yang, C. Shahabi. A pca-based kernel for kernel pca on multivariate time series, IEEE Intern. Conf. on Data Mining (2005).
22. C. W. Hsu, C. J, Lin. A comparison of methods multi-class support vector machines, IEEE Trans. on Neural Networks v.13 pp. 415-425 (2002).
23. Javier Giacomantone, Armando De Giusti, ROC performance evaluation of RADSPM technique, XIV Congreso Argentino de Ciencias de la Computación (CACIC), Chiclecito (2008).
24. W. A. Chaovaitwongse, P. M. Pardalos. On the Time Series Support Vector Machine using Dynamic Time Warping Kernel for Brain Activity Classification, Cybernetics and Systems Analysis v.44 pp.125-138 (2008).
25. N. Goddard, G. Hood, J. Cohen, W. Eddy, C. Genovese, D. Noll, L. Nystrom. Online analysis of functional mri datasets on parallel platforms. J.Supercomput., 11(3):295-318, (1997).
26. A. De Giusti, M. Naiouf. L. De Giusti, F. Chichizola. Dynamic Load Balancing on Non-Homogeneous Clusters. Lectures Notes in Computer Science, v.4330 pp.65-73. Springer Verlag, (2006)
27. T. Eickermann, W. Frings, F. Hossfeld, S. Posse, G. Goebbels. Supercomputer-enhanced functional mri of the human brain. IEEE Concurrency, 8(1):11-13, (2000).

Extracción de Información a partir de Datos No Estructurados No Textuales

J. Fernandez, N. Miranda, R. Guerrero, F. Piccoli

LIDIC- Universidad Nacional de San Luis

Ejército de los Andes 950

Tel: 02652 420823, San Luis, Argentina

{jmfer, ncmiran, rag, mpiccoli}@unsl.edu.ar

1. Contexto

Esta propuesta de trabajo se lleva a cabo dentro de la línea de investigación “Informática Gráfica” del proyecto “Técnicas de Inteligencia Computacional aplicadas a Optimización, Minería de Datos y Coordinación de Agentes Inteligentes”. Dicho proyecto se desarrolla en el marco del Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC), de la Facultad de Ciencias Físicas, Matemáticas y Naturales de la Universidad Nacional de San Luis.

2. Resumen

La toma de decisiones y la resolución de problemas complejos del mundo real requiere una visión generalizada del problema, para la cual se debe aportar la mayor cantidad de información relevante disponible. La información estructurada ha sido la principal materia prima utilizada en los sistemas computacionales hasta la fecha, pero restringirse solo a ella conduce, muchas veces, a representar una visión parcial. Hoy comienza a reconocerse que gran parte de la información requerida para la toma de decisiones y la resolución de problemas de índole general proviene de información no estructurada, principalmente aquella almacenada en la forma de audio, imagen y video. En este contexto, el uso de información obtenida a

partir de datos no estructurados no textuales es de vital importancia. Para lograrlo es necesario replantear las etapas bien definidas de los procesos tradicionales, es decir se plantea la necesidad de revisar las etapas de adquisición, procesamiento y análisis de la información contenida en audio, imágenes y/o video. Esta propuesta de trabajo establece los lineamientos a seguir para incorporar información no estructurada en los procesos de resolución de problemas o toma de decisiones.

3. Introducción

Internet ha cambiado profundamente la forma en la que las personas se comunican, negocian y realizan el trabajo diario al tener acceso a infinitud de recursos e información. En los últimos años el uso de la información multimedial se fue convirtiendo en una combinación de factores al momento de la toma de decisiones y/o la resolución de problemas, mayormente de tipo empresarial. No obstante, el éxito de la Web ha originado uno de sus principales inconvenientes: la administración de la heterogeneidad de los datos en la que se encuentra disponible dicha información. Mas aun, la extracción de información para la toma de decisiones a partir de dichos datos, si bien es una tarea aparentemente simple para una persona, no lo es para una computadora, principalmente cuando el tipo de información es no estructurada (texto,

audio, imágenes y video).

Desde el punto de vista humano, el proceso de extracción de información es un proceso inherentemente global; a partir de información escasa, localizada y usualmente de mala calidad, nuestro cerebro puede reconocer los elementos básicos que la componen y sus interrelaciones, y elaborar una experiencia sensorial (cognitiva, visual, auditiva, etc.) coherente y, como resultado de ello, deducir la información necesaria.

Desde un punto de vista computacional, la mayoría de la información existente proviene de datos configurados en forma estructurada, lo cual permite que la misma sea fácilmente interpretada y directamente utilizada por una computadora una vez que la información es accedida desde estructuras clásicas de almacenamiento tales como matrices, registros de bases de datos, entre otros. Sin embargo, este tipo de organización de la información conduce a representar una visión parcial del problema y dejar fuera de consideración información que podrá ser de gran importancia para la resolución efectiva del mismo.

En Ciencias de la Computación existen diversas áreas que intentan abordar el tratamiento de información no estructurada no textual, como es el caso de las áreas de procesamiento de señales, visión por computadora, minería de imágenes. Para el reconocimiento de un objeto/elemento cualquiera (visual o auditivo), es necesario previamente el entendimiento del escenario total en el cual este se encuentra inmerso; es decir, implica la coordinación de muchas tareas diferentes entre las cuales se pueden mencionar: estimar las características del objeto/elemento (aparición, forma, orientación, etc), reconocer el objeto/elemento (individualización involucrando la detección de objetos/elementos ocultos), categorizar el/los objetos/elementos reconocidos así como también el escenario total.

En los comienzos, la mayoría de las investigaciones han intentado reconocer un objeto/elemento en forma aislada mediante información local asociada al mismo, asumiendo que toda la información relevante de un

objeto/elemento se encuentra contenida dentro de una pequeña ventana en un espacio visual y/o auditivo [9, 10, 15, 23, 26, 27, 29]. Los errores típicos cometidos por dichos sistemas es que actúan como cámaras oscuras identificando objetos/elementos en forma individual sin considerar el escenario en el cual este se encuentra inmerso. Es importante entonces, el poder aproximar el comportamiento humano observando fuera de la cámara oscura y considerando el escenario en su totalidad. En [9, 12, 14, 17, 21, 22] queda claramente bosquejado que la comprensión de un escenario es una tarea de reconocimiento global que va más allá de la simple identificación de los objetos/elementos interactuantes; debe además poder reconocer la interacción entre ellos.

Como ejemplos del uso de información no estructurada no textual se encuentran diversas aplicaciones en *reconocimiento de patrones musicales* [4, 16, 7, 8], *cartografía digital* [1, 2, 11, 13], *reconocimiento de rostros* [24, 28, 30], *conducción autónoma de robots* [9, 10, 18, 20, 25], *tours de realidad virtual* [3, 5, 6, 17, 26], *caminatas (walkthrough) en juegos de computadoras* [19, 23], entre otros.

4. Líneas de Investigación y Desarrollo

Trabajar con información no estructurada no textual implica trabajar con diferentes tipos de datos y en consecuencia con diferente complejidad intrínseca. Tal como se mencionó anteriormente, las fuentes de información no estructurada no textual pueden ser señales de audio, imágenes y/o videos; cada una de ellas en forma individual o en conjunto habilitarán diferentes líneas de investigación, algunas de ellas dependientes entre sí. Entre las líneas propuestas a seguir se encuentran:

- *Identificación y representación de los aspectos perceptuales.* Básicamente asociados a las imágenes y video, esta línea pretende extraer la mayor cantidad de información perceptual posible. Usualmente las investigaciones se han concentrado

en métodos de segmentación y clustering tradicionales. No obstante, es necesario incorporar nuevos criterios perceptuales adecuados (sin pérdida de información) que permitan la representación de las imágenes no solo considerando propiedades de colorimetría y luminancia, sino también en lo relativo a la identificación de los objetos que se encuentran en el escenario plasmado y su ubicación espacial relativa (profundidad 3D).

La correcta identificación y categorización de los objetos en forma semántica junto a la correcta reproducción de los lineamientos teóricos establecidos para la simulación del procesamiento realizado por el sistema visual humano, permitirá la reconstrucción del escenario completo capturado en una imagen o una colección de ellas.

- *Representación robusta de la información.* Identificar e individualizar una señal a pesar de las distorsiones naturales (compresión, codificación analógica, ruido, entre otros) y ataques maliciosos (adición de logo, distorsión geométrica, cortes en la señal, cambio en la colorimetría, entre otros) en forma eficiente conlleva a desarrollar un método robusto para la determinación de huellas digitales de señales (imagen, streamings de audio o video).

Para obtener una buena huella digital de una señal es necesario identificar como iguales dos variantes de un mismo elemento, lo cual implica el redefinir la manera en la cual se los ha venido identificando. Una alternativa es la inclusión del contenido perceptual, ya que dos elementos pueden ser identificados como “el mismo” por una persona y pueden no coincidir en su representación interna.

Idealmente la huella digital debe ser una invariante de la señal; aquellas características intrínsecas, no alteradas por su constante manipulación. Una huella digital de una imagen puede ser una des-

cripción global de la imagen o una descripción local de las características claves extraídas. En cambio, en un video, puede ser una descripción global del video, un conjunto de huellas digitales para todos los frames del video o de solo aquellos claves del video. El mismo concepto se aplica a las señales de audio.

La determinación de una huella digital permitiría por ejemplo identificar objetos de audio a partir de segmentos, la detección de duplicaciones, plagios, el rotulado automático (MP3 modernos), la consulta por ejemplos y filtrados en redes p2p, entre otros. En el caso de los videos, a través de la huella digital se podrá identificar objetos o comportamientos específicos durante las transmisiones de canales de TV, o en bases de datos de video compartidas (Google, YouTube), seguridad a través de circuitos cerrados de video, derechos de autor, entre otros.

- *Recuperación de Información no estructurada:* El concepto de búsqueda exacta es central para los repositorios de información o bases de datos “tradicionales”. Trasladar el mismo concepto a repositorios de señales, no es simple. Por la naturaleza de los datos, el objetivo será resolver computacionalmente la similitud perceptual entre objetos similares. Los desarrollos actuales utilizan medidas de distancia, a través de las cuales se pretende reflejar la similitud perceptual. Dos objetos iguales perceptualmente deberán recibir distancias pequeñas; mientras que dos objetos perceptualmente distintos deberán recibir distancias grandes. Para ello es necesario contar con una representación robusta de cada dato del repositorio y definir una efectiva función de distancia, la cual permita establecer una valoración sobre la similitud o disimilitud de dos objetos y relajar el concepto de búsqueda exacta por búsqueda por proximidad.

5. Resultados obtenidos / esperados

El principal aporte de esta propuesta sera la incorporacion de informacion no estructurada en los procesos de toma de decisiones y resolucion de problemas que, normalmente, queda fuera de consideracion en los enfoques clasicos.

Actualmente se esta trabajando en la obtencion de una huella digital robusta y eficiente para audio e imagenes as como tambien en la identificacion y recuperacion de informacion perceptual de profundidad. La eficiencia no solo se relaciona con la capacidad de representar un vocamente una señal, sino con el tiempo que implica su calculo. Para ello se estan desarrollando algoritmos de procesamiento de imagenes, los cuales aplican tecnicas de programacion de alta performance.

6. Formacion de Recursos Humanos

Como resultado de las investigaciones se cuenta con una tesis de maestr a concluida y dos tesis doctorales y una de maestr a en desarrollo; as como tambien varios trabajos de fin de carrera de la Licenciatura en Ciencias de la Computacion.

Ademas las investigaciones se encuadran en el marco de un proyecto dentro del Programa de Promocion de la Universidad Argentina para el Fortalecimiento de Redes Interuniversitarias III en los que participa nuestra universidad junto con las universidades Michoacana (Mexico) y de Zaragoza (España).

Referencias

[1] Armitage, R.P., Weaver, R. E., and Kent, M. "Remote sensing of semi-natural upland vegetation: the relationship between species composition and spectral response". In: *Vegetation Mapping: From Patch to Planet* (eds Millington, A. and Alexander, R.), John Wiley and sons (2000).

[2] Blaschke, T. and Hay, G., "Object-oriented image analysis and scale-space: Theory and methods for modeling and evaluating multi-scale landscape structure". *International Archives of Photogrammetry and Remote Sensing* 34: 22-29 (2001).

[3] A.L. Callahan and Dmitry and B. Goldgof and Ph. D and Ph. D and Thomas A. Sanocki and Melanie A. Sutton, "Function from visual analysis and physical interaction: a methodology for recognition of generic classes of objects", *Journal on Image and Vision Computing*, Vol 16, pp 745-763, 1998.

[4] A. Camarena-Ibarrola, E. Chavez, "On Musical Performances Identification, Entropy and String Matching", *MICAI 2006*.

[5] P. Carbonetto, N. de Freitas, and K. Barnard, "A statistical model for general contextual object recognition", in *Proc. ECCV*, 2004.

[6] (CUDA)-IEEE International Conference on Multimedia and Expo 2008 ? Pp 697:700 ? April 2008.

[7] Crockett, et al., "A Method for Characterizing and Identifying Audio Based on Auditory Scene Analysis", *AES Convention Paper 6416*, presented at the 118. sup. th Convention May 28-32, 2005, Barcelona, Spain. cited by other.

[8] Dixon, S.: Live tracking of musical performances using on-line time warping. *Proc of the 8th Int Conf on Digital Audio Effects (DAFx'05)* (2005)

[9] A. Ess, B. Leibe, and L. Van Gool. "Depth and appearance for mobile scene analysis". In *ICCV*, 2007.

[10] R. Fergus, P. Perona, and A. Zisserman. "Object class recognition by unsupervised scale-invariant learning". In *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[11] M.A. Fischler and R.C. Bolles. "Random sample consensus: A paradigm for modeling with applications to image analysis and automated cartography". *Communications of the ACM*, 24(6):381-395, June 1981.

- [12] G. Kim, C. Faloutsos, and M. Hebert, "Unsupervised Modeling and Recognition of Object Categories with Combination of Visual Contents and Geometric Similarity Links", ACM International Conference on Multimedia Information Retrieval (ACM MIR), October, 2008.
- [13] Hay, G.J., Marceau, D. J., Dube, P., and Bouchard, A. "A Multiscale Framework for Landscape Analysis: Object-Specific Analysis and Upscaling". *Landscape Ecology* 16 (6): 471-490 (2001).
- [14] Hoiem, A.A. Efros, and M. Hebert, "Closing the Loop on Scene Interpretation", In CVPR 2008.
- [15] A. Hoiem, A. Efros, and M. Hebert. "Geometric context from a single image". In ICCV, 2005.
- [16] Ibarrola, A.C., Chavez, E.. "A robust, entropy-based audio fingerprint". IEEE, July 2006.
- [17] B. Leibe, N. Cornelis, K. Cornelis, and L. Van Gool. "Dynamic 3d scene analysis from a moving vehicle". In CVPR, 2007.
- [18] Liu Yang, Rong Jin, Caroline Pantofaru, and Rahul Sukthankar, "Discriminative Cluster Refinement: Improving Object Category Recognition Given Limited Training Data", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June, 2007.
- [19] B. Moghaddam, "Principal Manifolds and Probabilistic Subspaces for Visual Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, Issue 6, June 2002, pp. 780-788.
- [20] Pech-Pacheco, Jose L., Alvarez-Borrego, Josue, Cristobal, Gabriel, Keil, Matthias S., "Automatic object identification irrespective of geometric changes", International Society for Optical Engineering (SPIE), Vol 42(2): 551-559 (2003).
- [21] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. "Objects in context". In ICCV, 2007.
- [22] A. Saxena, M. Sun, and A. Y. Ng. "Learning 3-d scene structure from a single still image". In ICCV 3dRR-07, 2007.
- [23] H. Schneiderman, "Learning a restricted bayesian network for object detection," in Proc. CVPR, 2004.
- [24] Torres M., "Is there any hope for face recognition?", Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2004, 21-23 April 2004, Lisboa, Portugal.
- [25] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward Objective Evaluation of Image Segmentation Algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 6, June, 2007, pp. 929-944.
- [26] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in Proc. ICCV, 2003.
- [27] P. Viola and M. J. Jones, "Robust real-time face detection," IJCV, vol. 57, no. 2, 2004.
- [28] Xiaoxing Li; Mori, G.; Hao Zhang, Expression-Invariant Face Recognition with Expression Classification, The 3rd Canadian Conference on Computer and Robot Vision, Volume , Issue , 07-09 June 2006 Page(s): 77 ? 77, Digital Object Identifier:10.1109/CCRV.2006.34. 2006.
- [29] W. Zhang, B. Yu, D. Samaras, and G. Zelinsky. "Object class recognition using multiple layer boosting with heterogeneous features". In IEEE Conference on Computer Vision and Pattern Recognition, 2005.
- [30] W.Y. Zhao, R. Chellappa, "Image-based Face Recognition: Issues and Methods, Image Recognition and Classification", Ed. B. Javidi, M. Dekker, 2002, pp. 375-402.

SFP Tool: una Herramienta para Medir Puntos Función

Pamela Ritter¹, Mabel Bertolami² Gabriela Oriana³

Departamento de Informática, Facultad de Ingeniería, UNPSJB, Argentina

Fax 0297 4550836

email: ¹ pamelaritter@gmail.com; ² mbertolami@gmx.net; ³ orianag@arnet.com.ar

Resumen. En este artículo es presentado el software SFP Tool que permite medir el tamaño funcional de un sistema software a partir de los escenarios generados en la Elicitación de Requerimientos. Específicamente, esta herramienta semiautomatizada facilita la aplicación del procedimiento Scenario Function Points. En comparación con la medición manual pueden observarse ventajas significativas, como son agilizar el proceso de medición debido a la generación automática de los formularios, disponer del LEL y Escenarios del sistema y las reglas del procedimiento SFP, evitar inconsistencias y errores de cálculo, y exportar los formularios a la planilla de cálculo.

1. Introducción

El tamaño del software es un factor clave en los modelos de estimación de costo y esfuerzo de un proyecto de software. El Análisis de Puntos Función (FPA) es una métrica ampliamente usada en la industria para cuantificar el tamaño del software a partir de los requerimientos. Para facilitar las estimaciones en las etapas previas a la definición de los requerimientos, fue desarrollado un enfoque que permite estimar el tamaño funcional de un sistema software a partir de los artefactos producidos en la Elicitación de Requerimientos. Específicamente el procedimiento Scenario Function Points (SFP) [2] determina los Puntos Función (FP) de los escenarios derivados del Léxico Extendido del Lenguaje (LEL) [4].

La ejecución del procedimiento SFP es soportada por un proceso que consta de varios pasos. En el marco del Proyecto de Investigación “Estimaciones de Tamaño en la Etapa Inicial de un Proyecto de Desarrollo de Software”¹, realizado en la Facultad de Ingeniería, Sede Comodoro Rivadavia de la Universidad Nacional de la Patagonia San Juan Bosco, fue desarrollado el software SFP Tool que permite semiautomatizar dicho proceso. Esta herramienta permitió integrar en un software los diferentes utilitarios usados para la medición manual. Su aplicación en la práctica demostró que, además de reducirse el esfuerzo como consecuencia de la automatización de algunos pasos del proceso, se evitan las inconsistencias entre formularios y los errores de cálculo, se agiliza el acceso a los documentos necesarios para la medición y se favorece la detección de errores u omisiones en los escenarios. Respecto a estos últimos, para asegurar la consistencia de las mediciones deberían ser inspeccionados previamente, tal como lo sugieren Doorn *et al.* [3]. La herramienta permite realizar revisiones de las mediciones, es de fácil uso y dispone de ayuda para el usuario.

Este documento está organizado de la siguiente manera: en la sección 2 se resumen los conceptos básicos del procedimiento SFP, en la sección 3 se describe el software SFP Tool, su funcionamiento y modo de uso y en la sección 4 se presentan las conclusiones.

2. Procedimiento SFP

Este procedimiento permite estimar el tamaño en FP de un sistema software a partir de los escenarios. Su diseño está basado en el Método IFPUG FPA [5] y en la estructura propuesta en el estándar ISO/IEC 14143-1 [6] para los métodos de medición del tamaño funcional.

¹ El objetivo principal de este proyecto de investigación es desarrollar un modelo de estimación del tamaño del software en la etapa inicial de un proyecto de desarrollo.

La definición de SFP establece asociaciones conceptuales entre los componentes lógicos de los escenarios y los propuestos por IFPUG. El modelo de SFP consta de tres tipos de Componentes Funcionales Básicos: recurso, Entrada Externa (EI) y Salida Externa (EO). Un recurso representa una entidad de soporte de información (sin distinguir entre datos “internos” y “externos” a la aplicación). Una EI tiene semántica similar pero no es equivalente a la EI de IFPUG, pues en esta propuesta las EIs mantienen un recurso. La EO tiene el propósito de presentar datos independientemente de la naturaleza de los mismos (éstos pueden ser recuperados desde un recurso o generados por el proceso) y es la combinación de EO y EQ de IFPUG. SFP incluye reglas de identificación y clasificación de los tipos de entidades, reglas de asignación numérica para valorar las entidades del modelo y una función para derivar el tamaño funcional desde los componentes individuales. La complejidad de EIs y EOs es valorada como Baja, Media o Alta según el número de Tipos de Recursos Referenciados (RTRs) y Tipos de Datos Elementales (DETs); para los recursos se considera el número de Tipos de Datos Elementales (DETs). La contribución en FP es determinada usando las tablas de IFPUG para EI, EQ e ILF respectivamente.

3. Software SFP Tool

En esta sección se presentan la descripción, el funcionamiento y los resultados de las mediciones del software SFP Tool.

3.1. Descripción

Esta herramienta fue implementada en lenguaje visual Delphi, usando dos componentes principales: *Formula One*© Visual Components, Inc. para el manejo de formularios y *RichView*©1997-2004 Sergey Tkachenko para el hipertexto. Usa como entrada el archivo .rtf generado por la herramienta Baseline Menthor Workbench (BMW) [1], el cual es usado para cargar el LEL y Escenarios (L&E) del proyecto software que se va a medir. Como se trata de una herramienta semiautomática, hay partes del proceso que son realizadas automáticamente y otras que requieren la intervención del usuario. Como resultado del proceso se genera un archivo de extensión .vts - también puede ser exportado a una planilla Microsoft Excel - en donde se presentan todos los formularios generados por la aplicación.

3.2. Nueva medición

El primer paso consiste en cargar el archivo .rtf del L&E. A continuación quedará habilitada la opción Medir. El programa generará automáticamente la Planilla de Episodios (Figura 1) donde estarán almacenados todos los escenarios con sus respectivos episodios, los que deberán ser clasificados utilizando las reglas definidas para este paso del proceso de medición.

Formulario 1		SFP	
		Planilla de episodios	
TÍTULO DE LA APLICACIÓN: Sistema de Notificaciones a Clientes de un Banco		FECHA: 27/11/2008	
AUTOR: Gabriela			
Escenario	ID	Episodio	Tipo
Confeccionar aviso de deuda vencida		El Empleado completa el aviso de deuda vencida con la fecha de emisión, plazo en días y los datos del informe de deuda del cliente.	
		El empleado lo entrega a Gerencia.	
		if Gerencia firmó el aviso de deuda vencida then El Empleado lo envía al Cliente por correo electrónico.	EI EO D
		El Empleado archiva una copia del Aviso de Deuda Vencida en la Carpeta de Crédito.	
confeccionar avisos de préstamos		if fecha de vencimiento de la cuota del balance de préstamos es igual a fecha de vencimiento de Informe de Deuda then El empleado obtiene los datos del Cliente y tipo de préstamo que está en mora.	

Figura 1. Ventana de la Planilla de Episodios

Los episodios son clasificados seleccionando una de las opciones disponibles (EI, EO, D) en la lista desplegable de la columna Tipo. La opción elegida será almacenada en dicha columna y automáticamente a las EIs y EOs se les asignará un número de ID correlativo.

TICS APLICADAS AL REGISTRO DE TRANSACCIONES ORIGINADAS EN PUESTOS DE TRABAJO MOVILES

Autores

Claudio Aciti – caciti@exa.unicen.edu.ar
Gustavo Illescas - illescas@exa.unicen.edu.ar
Gustavo Tripodi - gtripodi@exa.unicen.edu.ar
Oscar Goñi - oegoni@alumnos.exa.unicen.edu.ar
Marcelo Tosini – mtosini@exa.unicen.edu.ar
Jorge Doorn – jdoorn@exa.unicen.edu.ar

Facultad de Ciencias Exactas- Universidad Nacional del Centro de la Provincia de Buenos Aires.
Instituto de Investigación en Tecnología Informática Avanzada (INTIA) - Teléfono: +54 2293
439680. Dirección postal: Campus Universitario, Paraje Arroyo Seco, (7000) Tandil, ARGENTINA

INTRODUCCION

La telefonía celular, y más aún las redes corporativas de celulares, han pasado a formar parte del día a día, y no solo como medio de comunicación, sino también para otros tipos de servicios: acceso a Internet, TV móvil, pagos y transacciones, entre otros.

Estos servicios, permiten encontrar una solución al problema de registrar transacciones donde los puestos de trabajo están distribuidos en todo el país, o que por características del lugar no son aptos para usar una PC o que un colector de datos.

Usualmente, en estos lugares se completan planillas en papel, que luego son pasadas a formato digital por un operador con posibles inconvenientes como pérdida o demora en la entrega de las planillas, y/o demora o error en la carga de datos.

La solución propuesta es el uso de teléfonos celulares de baja gama, que por medio de una llamada o un mensaje de texto pueden registrar información directamente sobre una base de datos desde cualquier lugar, a cualquier hora, y sin necesitar un operador. El costo de implementar una solución de esta naturaleza es muy bajo o nulo y adicionalmente no habría necesidad de capacitar al personal para el uso.

El presente proyecto forma parte de la presentación al “Concurso de subsidios para proyectos de investigación aplicada en el área de tecnologías de la información y la comunicación” para la convocatoria SINERGIA TIC 2008 de la CIC.

El mismo fue realizado entre el Instituto de Investigación en Tecnología Informática Avanzada (INTIA) de la Facultad de Ciencias Exactas de la Universidad de Centro de la Provincia de Buenos Aires (UNCPBA) e INFOGES SRL, perteneciente al Polo Informático, (www.empresar-sys.com.ar) quien participó como adoptante de la solución. El proyecto fue aprobado y actualmente se encuentra en ejecución.

OBJETIVOS GENERAL Y PARCIALES.

Objetivo General

Proveer una solución, de bajo costo, al problema de registrar transacciones en organizaciones donde los puestos de trabajo se encuentran distribuidos en una amplia zona geográfica, o que por las características del lugar no son aptos para usar una PC o un colector de datos, y/o que son compartidos por otras empresas.

Objetivos Parciales

-Estudio de diferentes tecnologías y desarrollo de un prototipo de un servidor de voz.

- Diseño de un prototipo de un tablero de control para construir modelos a medida.
- Desarrollo de una página web desde la que se puedan hacer todas las transacciones necesarias y que se pueda generar los reportes necesarios.
- Proveer a la empresa adoptante de una solución transversal a una problemática común que mejore el servicio a sus clientes.

PLAN DE TRABAJO.

1. Descripción General

Telservidor es un servidor de voz que provee un servicio de atención al usuario. Este servidor responde llamadas telefónicas por parte del usuario y puede brindar y/o recibir información.

El servidor de voz está compuesto por:

- A) una PC con un modem y,
- B) una línea de teléfono conectado a este último (restricción técnica: 3 líneas como máximo).

El sistema provee un servicio de identificación de las llamadas y autenticación tanto de números de teléfono como de código de usuarios. Los usuarios se dividen en diferentes perfiles:

- A) administrador (puede administrar todas las características del sistema),
- B) usuario preferencial (acceso de escritura en determinado tipo de información)
- C) usuario común (solo permisos de lectura).

Los usuarios acceden al sistema llamando por teléfono a un número determinado. Una vez que el sistema atiende se genera una interacción entre el usuario y el sistema que culmina con la registración de una transacción en una base de datos.

2. Definición del servicio

Para registrar una transacción, el servicio se puede utilizar de tres formas:

A) Vía Teléfono: Un usuario, ya registrado por el administrador, llama al número de teléfono asociado a este servicio. El servidor de voz le atiende, reproduciendo un mensaje de bienvenida y le solicita que ingrese su identificador de usuario y su clave personal. El servidor valida esta información y si es correcta, le solicita que ingrese los datos correspondientes al servicio que está utilizando. El sistema reproduce los datos ingresados para que el usuario valide la información. Si el usuario confirma los datos ingresados, el sistema guarda la información y emite un mensaje de despedida. Caso contrario, solicita al usuario que vuelva a ingresar la información.

El ingreso de datos se hace por medio del teclado numérico, siguiendo las convenciones usuales (usando un # para terminar la entrada, o aceptar una determinada cantidad de caracteres).

B) Vía SMS: Un usuario envía un mensaje de texto a un número pre-establecido, indicando su nombre de usuario, su clave personal y los datos relacionados al servicio (en este caso, se debe seguir una convención especificada previamente).

C) Vía web: Un usuario accede a una página web y completa los datos que desea en un formulario. Debe ingresar nombre de usuario, clave y los datos relacionados al servicio.

D) Vía Mail: Un usuario envía un correo electrónico, con los datos correspondientes a una transacción, a una cuenta previamente establecida.

El sistema provee de un servidor de mail que tiene un plug-in que se activa cuando recibe un mail para una cuenta específica. En este caso el mail es parseado y los datos son registrados en la base de datos de las transacciones.

3. Herramientas complementarias

Al servidor de voz se le anexan dos herramientas complementarias al servidor de voz: un tablero de control y una página web.

A) Tablero de Control: El tablero de control es una herramienta gráfica que se usa para diseñar, mediante el uso de grafos de estados o diagramas de flujo, el curso de un servicio específico.

Los modelos, o cursos de acción, que se definen para los distintos servicios pueden reutilizarse, ya que es común que se repitan entre ellos (por ejemplo: autenticación de usuarios, identificación de números de telefónicos, validación de datos, etc.).

Esta herramienta también permite diseñar estructuras de bases de datos de forma dinámica.

B) Página Web: Por medio de una página web, el administrador podrá hacer los cambios que desee, por ejemplo administrar usuarios y perfiles, editar las transacciones, generar reportes , enviar mensajes, etc.

La página web además provee un formulario como alternativa de registro de transacción en caso de que no se haya podido hacer por teléfono celular.

4 Actividades

I Estudio del proyecto y Relevamiento de información

- Lectura de Bibliografía relacionada.
- Estudio de diferentes tecnologías para diseñar el servidor de voz.
- Estudio de diferentes tecnologías para diseñar el tablero de control.
- Estudio de diferentes tecnologías para diseñar la pagina web.

II Desarrollo de un prototipo de servidor de voz

- comunicación entre un teléfono y una PC (vía llamada telefónica y vía SMS).
- Registrar transacciones en una base de datos a partir de la comunicación lograda en el punto anterior.
- Investigar la cantidad de modems que pueden funcionar en una PC en paralelo.

III Desarrollo de un prototipo de un tablero para generar bases de datos a medida y el curso de la llamada:

- Diseñar una herramienta gráfica para generar bases de datos de forma dinámica.
- Diseñar una herramienta gráfica para generar el curso de una llamada de forma dinámica.
- Diseñar una herramienta gráfica para reutilizar los modelos definidos en el punto anterior.

IV Diseño de una pagina web con la siguiente funcionalidad.

- Validación de usuarios

- Acceso a base de datos con diferentes permisos
- Diseño de un formulario, como alternativa, para registrar una transacción.

V Testeo de las distintas etapas.

- Test de stress en el servidor de voz
- Test de funcionalidad en el tablero de control
- Test de stress en el acceso a la base de datos
- Test de funcionalidad en la página web.

RESULTADOS ESPERABLES A PARTIR DE LA EJECUCIÓN DEL PROYECTO

Transferencia de tecnología y conocimiento

La adoptante, INFOGES (www.empresar-sys.com.ar) es una empresa, que pertenece al Polo Tecnológico de la UNCPBA, dedicada al desarrollo de Soluciones y Sistemas, que apuesta al crecimiento de sus clientes utilizando metodologías de trabajo, especialmente diseñadas para satisfacer necesidades actuales y futuras. Crece bajo la filosofía donde lo tecnológico debe ser una inversión orientada a reducir costos e incrementar ingresos, para lo cual desarrolla servicios alineados con las estrategias de negocio de sus clientes. INFOGES tiene como clientes potenciales a transferir el producto a empresas de transporte de primer nivel como El Rápido SA, Urquiza SRL, Río Paraná, Nueva Chevallier SA y a las Municipales de Chivilcoy y Vicente Lopez.

Tanto la adoptante INFOGES como la parte beneficiaria cuentan con experiencia previa y vastos conocimientos en el tema propuesto. Además, ambas partes trabajan en conjunto habitualmente en la relación interdisciplinaria que los une.

Expectativas de Logro

Se tiene como expectativa de logro los siguientes casos:

Caso 1. Carga de Combustible. El sistema de carga de combustible, actualmente implantado por la adoptante en varias empresas de transporte, provee todas las herramientas para mantener actualizado el sistema con el registro de las transacciones en los expendedores de combustible propios. A estos expendedores los utilizan las empresas para abastecer a sus unidades y, en algunos casos, a unidades de otra compañía. El uso de telefonía móvil permitiría el registro online de dichas transacciones en las playas de carga de combustible no aptas para el uso de una computadora personal o de un recolector de datos, brindando así un mejor servicio, más rápido y eficiente.

Caso 2: Seguimiento de encomiendas On-Line:

El sistema de despacho y administración de encomiendas, actualmente implantado por la adoptante en las empresas más importantes de transporte de pasajeros y carga del país, con cobertura a nivel nacional en más de 500 puestos de trabajo, provee todas las herramientas para el seguimiento permanente de los cambios de estado y ubicación de una encomienda transportada hasta el momento de su entrega, devolución o decomiso. Esta información además está disponible al cliente a través de un servicio on-line vía WEB o por correo electrónico. Hoy en día el uso de la telefonía celular permitiría ampliar el espectro de clientes con posibilidad de consultar en tiempo real el estado de su envío con un medio de comunicación masivo y de común acceso a los clientes.

Caso 3. Toma de Inventarios:

A través de la toma de inventarios desde la telefonía móvil la adoptante va a completar el proceso de “Inventario Online” que ha desarrollado con antelación. Este proceso introduce una mejora sustancial en el control de stock en tiempos donde es impensado frenar la producción de una Organización para realizar dicho conteo. El Proceso a grandes rasgos incluye la toma del estado Sistémico del inventario, los movimientos ocurridos entre el primer estado y el registro de las cantidades físicas y la justificación de las diferencias. En el registro de las cantidades físicas es donde interviene la telefonía móvil, brindando mayor confiabilidad, rapidez y optimización del tiempo de los recursos humanos y tecnológicos.

Resultados indirectos

- Aprovechar, dentro de la solución, las redes corporativas dentro de una empresa e inter-empresariales.

Bibliografía

- [Asa.2004] Asakura, Y.; Okuyama, G.; Nakayama, Y.; Usui, K.; Nakamoto, Y.; Multi-application platform for mobile phones. Software Technologies for Future Embedded and Ubiquitous Systems, 2004. Proceedings. Second IEEE Workshop on. 11-12 May 2004 Page(s):139 – 143
- [Ast.2001] Astarita, V.; Florian, M.; The use of mobile phones in traffic management and control. Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE. 25-29 Aug. 2001 Page(s):10 - 15
- [Bal. 2006] Ballagas, R.; Borchers, J.; Rohs, M.; Sheridan, J.G.; The smart phone: a ubiquitous input device. Pervasive Computing, IEEE. Volume 5, Issue 1, Jan.-March 2006 Page(s):70 - 77
- [Bal.2000] Ballvé, Alberto. Tablero de Control. Ed. Macchi Buenos Aires 2000.
- [Gall.2004] Gallego, Domingo J. y Ongallo Carlos. Conocimiento y Gestión. Pearson Educación, Madrid, España. 2004.
- [Kap.2000] Robert S. Kaplan - David P. Norton. The Balanced Scorecard: Translating strategy into action. - Harvard Business School Press. (El Cuadro de Mando Integral 2da edición - 2000)
- [Olv.2000] Olve Nils-Göran, Roy Jan, Wetter Magnus. Implantando y gestionando el Cuadro de Mando Integral. Gestión 2000. Barcelona 2000.
- [Ong.2000] Ongallo C. Manual de comunicación. Guía para gestionar el Conocimiento en las Empresas y Organizaciones. Madrid, Dykinson. 2000.
- [Por.2005] Porter Michel E. Estrategia y ventaja competitiva. Ediciones Deusto, Planeta DeAgostini Profesional y Formación, S.L., Barcelona. Noviembre 2006. ISBN 950-857-035-0.
- [Wei.2007] Wei, C.Y.; Capturing Mobile Phone Usage: Research Methods for Mobile Studies. Professional Communication Conference, 2007. IPCC 2007. IEEE International. 1-3 Oct. 2007 Page(s):1 - 6
- [Wei.2005] Wei, C.; Kolko, B.E.; Studying mobile phone use in context: cultural, political, and economic dimensions of mobile phone use. Professional Communication Conference, 2005. IPCC 2005. Proceedings. International. 10-13 July 2005 Page(s):205 - 212

Ingeniería Inversa Aplicada a Software Numérico: Modelos Climáticos

Fernando G. Tinetti*, Pedro G. Cajaraville#, Juan C. Labraga##, Mónica A. López##, María G. Olguín#

#Departamento de Informática, Facultad de Ingeniería – (UNPSJB¹)

##Centro Nacional Patagónico (CENPAT-CONICET²)

III-LIDI, Facultad de Informática – (UNLP³)

fernando@info.unlp.edu.ar, gustavo.cajaraville@gmail.com, {labraga, monica}@cenpat.edu.ar, gabriela.olguin@gmail.com

CONTEXTO

Se presentan los avances realizados en el marco de la exposición efectuada con el mismo nombre en el WICC 2008 de General Pico, La Pampa. A su vez este trabajo forma parte del proyecto de investigación “Cómputo Optimizado y Paralelo en Clusters de PCs Aplicado a Modelos Numéricos del Clima”, que se está desarrollando en la Sede Puerto Madryn de la Universidad Nacional de la Patagonia San Juan Bosco. El objeto de estudio es el software heredado que implementa un modelo simulación de circulación general de la atmósfera.

RESUMEN

Los objetivos perseguidos fueron: recuperar el diseño del software que implementa al modelo; redocumentar las subrutinas incluidas en los niveles superiores del árbol de invocaciones y especificar los aspectos metodológicos que puedan ser utilizados para efectuar ingeniería inversa de otros modelos de simulación basados en el lenguaje Fortran. Se emplearon las técnicas de ingeniería inversa que se aplican para el análisis de software científico. Los principales resultados obtenidos fueron: a) Un conjunto de reglas para depuración de código quitando comentarios innecesarios y código anulado; b) La definición y aplicación de métricas para evaluar el volumen y la eficiencia de la depuración efectuada; c) La redocumentación del sistema mediante la generación de árboles de invocaciones partiendo de distintas subrutinas del sistema; d) La recopilación de información sobre el intercambio de datos entre subrutinas (uso de bloques common y pasaje de parámetros).

El modelo de referencia es en realidad el acoplamiento de un modelo atmosférico y un modelo oceánico. Estructuralmente el *modelo acoplado* es el *encargado* de invocar a rutinas de ambos modelos: atmosférico y oceánico, todo programado en Fortran 77. Se trata de un software heredado (legacy code) del que se contó con muy poca documentación sobre su diseño. Esta situación es bastante usual en el campo de los modelos numéricos, habitualmente implementados en programas de varios miles de líneas de código. Esto por supuesto dificulta las tareas de mantenimiento y cambios de funcionalidades, por lo que es muy conveniente la aplicación de ingeniería inversa para recuperar información del diseño del software.

Palabras Clave: ingeniería inversa, software heredado, modelado climático.

¹ Universidad Nacional de la Patagonia San Juan Bosco

² Centro Nacional Patagónico - Consejo Nacional de Investigaciones Científicas y Técnicas

³ Universidad Nacional de La Plata

* Investigador Adjunto sin Director, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

Profesor Titular, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco

1. INTRODUCCIÓN

La comprensión del sistema climático es un problema de gran interés científico mundial. Si bien es cierto que se han realizado avances considerables en el tema, aún son muchos los factores que continúan limitando la capacidad de detectar, atribuir y comprender el cambio climático actual y proyectar los cambios climáticos que podrían ocurrir en el futuro (*Intergovernmental Panel on Climate Change* [1]). En la actualidad, las herramientas más confiables para la investigación del clima, sus fluctuaciones y variaciones, son los Modelos de Circulación General de la Atmósfera (MCGA).

Un MCGA es una representación numérica espacial y temporal aproximada de los principales procesos físicos que ocurren en la atmósfera y de las interacciones con los otros componentes del medio ambiente. Esencialmente, está constituido por un sistema de ecuaciones diferenciales en derivadas parciales que expresan los principios de conservación de la cantidad de movimiento, la energía termodinámica y la masa del sistema. Estas ecuaciones expresan las leyes físicas que gobiernan la atmósfera. Actualmente, los MCGA están integrados con modelos numéricos de funcionamiento de los océanos, la criosfera, y representaciones simplificadas de la biosfera (*MCGA acoplados*), lo que permite simular las interacciones de la atmósfera con su entorno, en distintas escalas de tiempo.

Mediante el uso de computadoras con gran capacidad de procesamiento y la aplicación de diversos métodos de cálculo numérico, pueden obtenerse soluciones numéricas aproximadas del sistema de ecuaciones de un MCGA. De este modo, se obtiene la evolución temporal y espacial del sistema climático (en forma tridimensional), en función de las condiciones iniciales y de contorno elegidas y de los valores de ciertos parámetros climáticos (ejemplo: concentración de CO₂ atmosférico). La solución numérica de un MCGA en condiciones preestablecidas se denomina *experimento climático*. Integrando el modelo atmosférico con una representación de la criósfera y una representación de los océanos, se obtienen pronósticos de cambios climáticos y anomalías en el clima. Se utiliza por lo general un método espectral, esto es, representación de capas atmosféricas.

1.1 El Modelo de Circulación General de la Atmósfera

El modelo de circulación general de la atmósfera analizado es uno de los MCGA utilizado por la comunidad científica internacional y participó junto con otros modelos del *Program for Climate Model Diagnosis and Intercomparison* (PCMDI) [7] [3] [4] [5].

En el año 1999 fue cedido, luego de un período de capacitación, a un grupo de investigación de la Unidad de Investigación de Oceanografía Física y Meteorología (UIOM) del CENPAT (Centro Nacional Patagónico)-CONICET. Las aplicaciones de este modelo en la UIOM han permitido avances en el conocimiento de temas cruciales para el desarrollo económico y social de la Argentina: la predictibilidad de anomalías climáticas estacionales, el impacto del calentamiento global sobre el ciclo hidrológico en la Argentina, el calentamiento global y los cambios climáticos regionales y el impacto del cambio climático en los recursos hídricos de la cordillera de los Andes. El sistema informático de pronóstico de anomalías climáticas basado en este modelo desarrollado por el equipo de la UIOM, publica mensualmente en forma experimental información objetiva sobre la magnitud y la probabilidad de condiciones atípicas del clima.

El código original del modelo fue desarrollado para supercomputadoras (CRAY, Silicon Graphics, etc.) con varios procesadores bajo sistema operativo UNIX. Para su utilización en la Argentina fue necesario portar el modelo en computadores de rango intermedio (Sun Ultra10, Enterprise) con un

solo procesador. En una etapa posterior se implementó también en la supercomputadora Cray Origin 2000 “Clementina II”, instalada en las dependencias de la Secretaría de Ciencia, Tecnología e Innovación Productiva (SeCyT) para uso científico. En la actualidad la ejecución de este modelo se implementa en Computadoras Personales (PC) con procesadores Intel, bajo sistema operativo Linux y consta de un programa principal y más de 250 subrutinas.

1.2 Ingeniería Inversa

La ingeniería inversa del software es el proceso consistente en analizar un programa, en un esfuerzo por crear una representación del mismo con un nivel de abstracción más elevado que el código fuente. Es un proceso de recuperación del diseño. Las herramientas a utilizar extraen información acerca de los datos, arquitectura y diseño de procedimientos de un programa ya existente. A diferencia de lo usual en otras disciplinas, donde este proceso se aplica a los productos de la competencia, en ingeniería de software con frecuencia se debe aplicar a los propios trabajos de la organización, realizados hace muchos años. En el caso de este proyecto, se aplica a un trabajo hecho por otra organización, pero que se intenta optimizar con varios objetivos: desde la optimización de rendimiento numérico hasta el cambio/mejora de funcionalidades.

Según va aumentando la abstracción crece la complejidad del trabajo, así como la necesidad de comprensión de la aplicación. La ingeniería inversa debe ser capaz de abstraer, a partir del código fuente, información significativa del procesamiento que se realiza, las estructuras de datos que se usan en el programa y el interfaz con el usuario que se utiliza.

Los sistemas heredados tienen características propias que dificultan su comprensión y mantenimiento. En caso de aplicaciones científicas, de grandes dimensiones y gran complejidad algorítmica, ante la necesidad de adaptar el software a nuevos requerimientos, se intenta mantener, dentro de lo posible, la mayoría de los elementos de software. La ingeniería inversa en estos casos es un proceso iterativo que involucra combinaciones de actividades automáticas y manuales.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Se definieron dos líneas de investigación: en forma específica, recuperar el diseño del software que implementa al modelo acoplado y mejorar su documentación, y en forma genérica, avanzar en la especificación de metodologías que puedan ser utilizadas para efectuar ingeniería inversa de código heredado basado en el lenguaje Fortran.

Para la comprensión de la estructura de los programas de cómputo secuencial que implementan el modelo acoplado océano-atmósfera, se utilizó un método de refinamiento sucesivo, acorde a lo planteado por Chikofsky en “Reverse Engineering and Design Recovery: A Taxonomy” [2], buscando identificar, en principio, bloques de subrutinas y sus relaciones funcionales, para luego analizar con mayor detalle las subrutinas más significativas, identificando entradas, salidas y procesos involucrados, procediendo luego a su redocumentación. En definitiva, se intentó comprender:

- la estructura del software y los datos que se manejan: bloques de memoria compartida, variables pasadas como parámetro, etc.;
- el procesamiento que se realiza en el código: invocaciones entre subrutinas (en particular cuando son de modelos diferentes) y la dinámica de activaciones y pasaje de datos entre modelos.

El proceso realizado alternó tareas automáticas y manuales. Incluyó una primera etapa de recopilación de información y capacitación, luego depuración del código fuente disponible, una etapa iterativa para el análisis de la comunicación entre los modelos atmosférico y oceánico, y finalmente una etapa de redocumentación, para organizar y entregar los resultados obtenidos

Las herramientas utilizadas fueron seleccionadas de entre varias encontradas en Internet para entorno Linux (preferentemente de uso libre), y orientadas a ingeniería inversa, análisis de sistemas o procesamiento de texto, como ser: Understand for Fortran: para ingeniería inversa; gprof: profiler, para análisis dinámico; KProf: para graficar la salida del gprof y realizar comparaciones entre distintas corridas del profiler; SourceNav: para recorrer el código fuente en las tareas de redocumentación; Graphviz: paquete para generar grafos jerárquicos; Perl: lenguaje de programación orientado a procesamiento de texto.

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo principal planeado y cumplido de este proyecto fue recuperar el diseño del modelo climático acoplado océano-atmósfera en un nivel alto de abstracción, es decir:

- Diagramas de estructura de la comunicación entre los componentes de los modelos numéricos involucrados: acoplado, atmosférico y oceánico.
- Para cada modelo (acoplado, atmosférico y oceánico):
 - Árbol de invocación.
 - Listado de las rutinas utilizadas.
 - Listado de bloques *common* de Fortran 77.
 - Código re-documentado de las subrutinas.
- Implementación de reglas para automatizar el proceso de depuración del código fuente, que básicamente fueron:
 - Quitar líneas de comentario referidas a evolución histórica pero que no aportan información sobre diseño ni funcionalidad.
 - Quitar líneas de comentario que anulan código en desuso.

Si bien muchos de los resultados obtenidos en este proyecto estuvieron relacionados con el software específico en estudio, otros pueden aplicarse en forma genérica a sistemas heredados escritos en lenguaje Fortran 77:

- reglas generales aplicables a cualquier código Fortran 77 para depuración de código en desuso;
- definición de métricas para evaluar el volumen de código comentado en un sistema, y la eficiencia de su depuración;
- generación de árboles de invocaciones con representación gráfica de la estructura partiendo de distintas subrutinas del sistema;
- recopilación de información sobre el intercambio de datos entre subrutinas (uso de

bloques common y pasaje de parámetros), más la posibilidad de graficar o mostrar esos datos en una grilla.

4. FORMACIÓN DE RECURSOS HUMANOS

La consecución de estos objetivos enunciados anteriormente, permitió la consolidación del grupo de investigación como tal, avanzar en el conocimiento sobre modelos numéricos, ingeniería inversa de código científico heredado y cómputo paralelo.

5. BIBLIOGRAFÍA

[1] Intergovernmental Panel on Climate Change, Tercer Informe de Evaluación – Cambio Climático 2001 – Base Científica, disponible en <http://www.ipcc.ch/pub/un/ipccwgl1s.pdf>

[2] Elliot J. Chikofsky, James H. Cross II, Reverse Engineering and Design Recovery: A Taxonomy, 1990, IEEE Software, vol. 7, no. 1, www.lore.ua.ac.be/Research/Chikofsky1990-Taxonomy.pdf.

[3] G. J. Boer, K. Arpe, M. Blackburn, M. Déqué, W. L. Gates, T. L. Hart, H. le Treut, E. Roeckner, D. A. Sheinin, I. Simmonds, R. N. B. Smith, T. Tokioka, R. T. Wetherald, D. Williamson, “Some Results From an Intercomparison of the Climates Simulated by 14 Atmospheric General Circulation Models”, J. Geophys. Res., 97(D12), 12,771–12,786.

[4] G. J. Boer et al, “Intercomparison of climates simulated by 14 atmospheric general Circulation model”, WMO/TD-No 425, CAS/JSC Working Group on Numerical Experimentation.

[5] Global and Planetary Change, An Overview of results from the Coupled Model Intercomparison Project (CMIP), Feb. 2001, UCRL-JC 140274.

[6] <http://www.supercomputo.secyt.gov.ar/>

[7] Program for Climate Model Diagnosis and Intercomparison <http://www-pcmdi.llnl.gov/>

[8] F. G. Tinetti, P. G. Cajaraville, J. C. Labraga, M. A. López, "Cómputo Paralelo Aplicado a Modelos Numéricos del Clima", IX Workshop de Investigadores en Ciencias de la Computación, Universidad Nacional de La Patagonia San Juan Bosco (UNPSJB), Trelew, Chubut, Argentina, Mayo 3-4 de 2007.

[9] World Meteorological Organization, Numerical Method Used in Atmospheric Models, GARP Publication Series 17, 1979.

MEJORA EN LOS PROCESOS DE IDENTIFICACIÓN DE SERVICIOS, OBTENCIÓN DE REQUISITOS, INTEGRACIÓN DE INFORMACIÓN Y DESARROLLO DE SOFTWARE

Grupo de Investigación en Ingeniería de Software del Comahue (GIISCo)

<http://giisco.uncoma.edu.ar>

Departamento de Ciencias de la Computación

Universidad Nacional del Comahue

RESUMEN

Tomando como base conocimientos adquiridos en investigaciones anteriores, el presente proyecto aborda distintas temáticas en la mejora de procesos software, centrandolo en (1) mejoras a procesos específicos – ej. especificación de requisitos; (2) definición y validación de técnicas para evaluación de composiciones en dominios específicos (Sistemas de Información Geográficos, Requisitos en Desarrollos Globales); (3) evaluación de atributos de calidad específicos (integrabilidad y accesibilidad); (4) incorporación de orientación a aspectos, ingeniería cognitiva y semántica al proceso de desarrollo de software; y (5) generalización de prácticas a través de mejora de procesos de reuso.

Los resultados obtenidos a partir de las líneas citadas están siendo integrados en herramientas de software que faciliten las actividades de verificación de propiedades y búsqueda.

Los objetivos del proyecto se desprenden de varias actividades de cooperación, entre ellas el marco de un proyecto CyTED finalizado en diciembre 2008. Los resultados serán aplicados en el contexto de este proyecto así como de cooperaciones en curso (Agencia Italiana para la Protección y Servicios Técnicos del Medio Ambiente (APAT) y Universidad de Bari, Italia.

1. MOTIVACIÓN

Uno de los elementos presentes en cualquier técnica de modelado conceptual basado en servicios es la calidad de la especificación – a veces relacionada directamente con la calidad del proceso aplicado para obtener la especificación [1]. Una característica importante de los componentes es su especificidad: ¿Cuánto de específico es un

componente relacionado con una tarea particular? Está claro que cuanto más cercano sea el componente a la tarea, requerirá menos modificaciones - y será menor el costo de su implementación. Por otra parte, la cantidad de componentes se incrementa al hacerse éstos más específicos. Para abordar ese problema, generalmente se ofrece una familia de soluciones que capturan las variaciones de un diseño básico pero que trabajan en el contexto de un marco mayor y dependiente de un dominio. Por ejemplo, cuando las funciones de un sistema de información geográfico se implementan en componentes de software, aparecen como un conjunto de servicios a disposición de los desarrolladores. Esos servicios son parte de una arquitectura de software por niveles donde cada componente es implementado por una colección de objetos o módulos con una interacción bien definida. La información suministrada generalmente se encuentra en catálogos distribuidos, con datos no homogéneos, inconsistentes e incompletos; lo que complica el proceso de obtención de servicios confiables que atiendan requisitos determinados. Esos requisitos pueden además obtenerse de manera global cuando los equipos de trabajo se encuentran ubicados en lugares dispersos geográficamente. Por ello, es de extrema importancia que se potencie el uso de estándares a fin de asegurar la interoperabilidad y facilidad de uso.

Por lo general, la calidad deseada en las composiciones de componentes no se expresa directamente sino a través de expresiones vagas, por ejemplo “rendimiento aceptable”, “tamaño pequeño”, o “alta confiabilidad”. A pesar de ello, algunas investigaciones comienzan a definir taxonomías y ontologías que ayudan en la identificación, evaluación y uso de servicios. Sin embargo, todavía es evidente la necesidad de un entendimiento

común en lo que deba considerarse información estándar que facilite la búsqueda de composiciones software [2].

El complejo proceso de creación de software basado en componentes y servicios, abarca entonces cuestiones tan diversas como el tipo de documentación elegida, el proceso de obtención y modelado, la validez de los modelos creados, etc. Estos son parte de los nuevos retos con los que se enfrenta actualmente el proceso de desarrollo de software. En el caso de pequeñas y medianas empresas de la industria de software en Ibero América, abordar estos problemas es un desafío que también depende del grado de madurez de sus procesos software. Por ello, el enfoque debe ser integral – no sólo abarcando aspectos de modelado sino también de gestión y de crecimiento en los distintos aspectos que hacen a la madurez de los procesos.

2. CONTRIBUCIONES A LA FECHA

En [14] hemos enumerado una serie de contribuciones que profundizaron temas de evaluación de componentes software a través del uso de métricas, testing y formalización de coordinación. También hemos comenzado a explorar el modelado semántico y cognitivo como medio para mejorar propiedades específicas (ej. integrabilidad) y la orientación a aspectos para mejorar accesibilidad. La mejora de procesos en forma integral fue abordada desde la propuesta del proyecto CyTED CompetiSoft (Mejora de procesos para fomentar la competitividad de la pequeña y mediana industria del software en Iberoamérica. Proyecto CyTED 3789¹. Se validaron métodos específicos mediante experimentos, casos de estudio y experiencias piloto.

Durante el año 2008, hemos propuesto mejoras al proceso de desarrollo de software en general que han sido en algunos casos validadas con casos de estudio en empresas de carácter gubernamental [15][16][18]; hemos profundizado en la mejora de métodos para obtención de requisitos en desarrollos globales y validado mediante casos de estudio [9][10][11][12] así como en la priorización de requisitos [17]; hemos elaborado y validado, mediante casos de estudio, mejoras propuestas en el campo de la integración de información

geográfica [3][4] y en la accesibilidad de sitios web [7][8]; y hemos avanzado en las propuestas de mejora para la identificación de servicios [13][19] y para la sustitución de componentes [5][6].

Los casos de estudio han sido realizados en entidades académicas (mejoras al proceso de elicitación global), gubernamentales y empresas (mejoras al proceso software; integración de información) en el ámbito nacional e internacional. Actualmente, los resultados de los casos de estudio están siendo evaluados para su extensión, replicación o reformulación.

En el caso de propuestas (identificación de servicios, mejora de accesibilidad), nos encontramos elaborando casos de estudio para ser desarrollados y evaluados durante 2009.

3. IMPACTO Y TRANSFERENCIA

El mayor impacto del proyecto se centra en la formación de recursos humanos, consolidación de grupos de investigación e interacción entre grupos interdisciplinarios.

El contexto de desarrollo de este proyecto continuará en el año 2009 en conjunto con docentes de la Universidad de Castilla-La Mancha, España. En particular, el proyecto ha facilitado la realización de pasantías de investigación por parte de nuestros investigadores. Desde el punto de vista de la aplicabilidad del producto, notemos que el desarrollo de sistemas de información sigue siendo hoy un proceso costoso principalmente por la cantidad de software que nunca llega a utilizarse, que no se define correctamente, que no se actualiza, etc. El desarrollo basado en componentes y servicios es un avance, pero el costo asociado a ubicar componentes reutilizables y adaptarlos sigue siendo alto. Si asociamos a ello el uso de nuevas tecnologías (multimedia, ambientes móviles, etc.) la incidencia es notablemente mayor. Por otro lado, la aplicación de resultados al dominio de los sistemas de información geográficos permite profundizar y formalizar la cooperación iniciada con la Universidad de Bari, Italia, a través de la extensión de un sistema de integración de datos federados, desarrollado por la Agencia Italiana para la Protección y Servicios Técnicos del Medio Ambiente (APAT).

La mejora de procesos (en cualquiera de sus aspectos) ayuda a las organizaciones

¹ <http://alarcos.inf-cr.uclm.es/Competisoft/index.aspx>

desarrolladoras de software en el acceso a mercados que requieren mayor calidad del producto. También sirven para mejorar procesos en organizaciones que desean reformular sus procesos de negocios así como los servicios que brindan a clientes. En ese contexto, nuestras experiencias y resultados están influenciando las prácticas de desarrollo de software y de la organización en sí en ciertos ámbitos de gobierno, como un medio para el soporte de la gobernabilidad electrónica en la provincia de Neuquén.

4. INVESTIGADORES

GIISCo reúne aproximadamente a 12 (doce) investigadores, entre los que se cuentan docentes y alumnos de UNComa y asesores externos. La mayoría de los docentes-investigadores de GIISCo-UNComa han terminado o se encuentran próximos a terminar carreras de postgrado. En el año 2008 e inicio del 2009, 3 investigadores han finalizado sus doctorados con lo que se cuenta actualmente con 4 doctores, 1 doctorando y 2 maestrando entre los miembros del grupo.

Contacto: Dra. Alejandra Cechich
(acechich@uncoma.edu.ar;
acechich@gmail.com)

5. REFERENCIAS

- [1] Wallnau K., S. Hissam, and R. Seacord. *Building Systems from Commercial Components*. Addison-Wesley, 2002.
- [2] Cechich A., A. Réquile, J. Aguirre, and J. Luzuriaga. *Trends on COTS Component Identification*. 5th International Conference on COTS-Based Software Systems, ICCBSS 2006, 13-17 Febrero 2006, Orlando, USA. IEEE Computer Science Press.

6. ALGUNAS CONTRIBUCIONES AÑOS 2008/2009

- [3] A. Buccella, A. Cechich, and P. Fillotrani. *Ontology-Driven Geographic Information Integration: A Survey of Current Approaches*. Computers & Geosciences Special Issue on Geoscience Knowledge Representation in Cyberinfrastructure. B. Brodaric, P. Fox, and D. McGuinness Editors. Elsevier Science Publishers B. V. 2009.

- [4] A. Buccella, D. Gendarmi, F. Lanubile, G. Semeraro, A. Cechich and A. Colagrossi. *A Layered Ontology-Based Architecture for Integrating Geographic Information* 21th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems. June 18-20, Wroclaw (Poland). Studies in Computational Intelligence. Series Editor: Kacprzyk, Janusz. ISSN: 1860-949X, 2008.
- [5] A. Flores, M. Polo. *Testing-Based Component Assessment for Substitutability*. Proceedings of the Tenth International Conference on Enterprise Information Systems, Volume ISAS-2, Barcelona, España, Junio 12-16, 2008.
- [6] A. Flores, M. Polo. *Testing-Based Assessment Process for Upgrading Component Systems*. 24th IEEE International Conference on Software Maintenance (ICSM 2008), Sept. 28 – Oct. 4, 2008, Beijing, China
- [7] A. Martín, A. Cechich, and G. Rossi *Comparing Approaches to Web Accessibility Assessment*. Capítulo en Handbook on Research on Web Information Systems Quality. M. Moraga, C. Calero, and M. Piattini (Eds), IGI Global Publishing, USA, 2008.
- [8] A. Martín, A. Cechich. *Extendiendo MVC para Diseñar Interfaces de Usuario Accesibles*. XIV Congreso Argentino en Ciencias de la Computación. La Rioja, Argentina, Octubre 2008.
- [9] G. Aranda, A. Vizcaíno, A. Cechich, M. Piattini. *Strategies to Recommend Groupware Tools According to Virtual Team Characteristics*. The 7th IEEE International Conference on Cognitive Informatics, ICCI. IEEE Computer Science Press, 2008.
- [10] G. Aranda, A. Vizcaíno, A. Cechich, M. Piattini. *A Methodology for Reducing Geographical Dispersion Problems during Global Requirements Elicitation* WER 2008, Workshop en Ingeniería de Requisitos. Barcelona, España, Septiembre 2008. ISBN: 978-84-7653-144-0, pp. 117-127.
- [11] G. Aranda, A. Vizcaíno, A. Cechich, M. Piattini. *A Requirement Elicitation Methodology for Global Software Development Teams*. En "Encyclopedia of Information Science and Technology,

- Second Edition”, Editor: Mehdi Khosrow-Pour. Editorial IGI Global Publishing, USA. ISBN 9781605660264, 2008.
- [12] G. Aranda, A. Vizcaíno, A. Cechich, M. Piattini.. Evaluating Factors that Challenge Global Software Development ICSoft 2008, Sesión Especial: Global Software Development: Challenges and Advances. Porto, Portugal, Julio 2008. ISBN: 978-989-8111-52-4, pp. 355-363
- [13] G. Gatean, A. Cechich, A. Buccella. Un Esquema de Clasificación Facetado para Publicación de Catálogos de Componentes SIG. XIV Congreso Argentino en Ciencias de la Computación. La Rioja, Argentina, Octubre 2008.
- [14] GIISCo Research Group, UNComa, Identificación, Evaluación y Uso de Composiciones Software, WICC 2008: 10th Workshop de Investigadores en Ciencias de la Computación, La Pampa, Argentina, 5-6 Mayo 2008.
- [15] J. Luzuriaga, R. Martínez, A. Cechich. Setting SPI Practices in Latin America: An Exploratory Case Study in the Justice Area. International Conference on Theory and Practice of Electronic Governance, ICEGOV’. 1-4 Diciembre, El Cairo, Egipto, ACM Press, 2008.
- [16] Luzuriaga J., R. Martínez, A. Cechich.. Improving Resource Management: Lessons from a Case Study from a Middle-Range Governmental Organization. Capítulo en Software Process Improvement for Small and Medium Enterprises: Techniques and Case Studies. Hanna Oktaba and M. Piattini (Eds), IGI Global Publishing, USA, 2008.
- [17] N. Martínez, A. Cechich. Elicitación Basada en Psicología Cognitiva: Un Caso de Estudio. XIV Congreso Argentino en Ciencias de la Computación. La Rioja, Argentina, Octubre 2008.
- [18] R. Anaya, A. Cechich, and M. Henao A Model to Classify Knowledge Assets of a Process-Oriented Development. Capítulo en Software Process Improvement for Small and Medium Enterprises: Techniques and Case Studies. Hanna Oktaba and M. Piattini (Eds), IGI Global Publishing, USA, 2008.
- [19] V. Saldaño, A. Buccella, A. Cechich. Una Taxonomía de Servicios Geográficos para Facilitar la Identificación de Componentes. XIV Congreso Argentino en Ciencias de la Computación. La Rioja, Argentina, Octubre 2008.

HACIA UN MODELO DE TRAZAS DESACOPLADO DE LOS MODELOS DE REQUISITOS

Graciela Hadad¹, Andrea Vera¹, Jorge Doorn^{1,2}, Gladys Kaplan¹

¹Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de La Matanza

²Depto. de Computación y Sistemas, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Pcia. de Buenos Aires
gracielahadad@gmail.com, andreafabianavera@gmail.com, jdoorn@exa.unicen.edu.ar, gladyskaplan@gmail.com

CONTEXTO

La presente propuesta se corresponde con la línea de investigación “Gestión de Requisitos” en el proyecto de investigación “Consolidación de Requisitos de Software” del grupo de investigación en “Ingeniería de Requisitos” de la UNLaM.

RESUMEN

Como menciona Leite [Leite 97]: “el cambio es una propiedad intrínseca al software”, y por lo tanto merece una consideración especial en el proceso de desarrollo: la administración de dichos cambios. Desde la perspectiva de la ingeniería de requisitos, esta actividad se denomina: la gestión de requisitos. Ella involucra dos sub-actividades centrales: el control del versionado y la trazabilidad de los mismos. En este proyecto se aspira a crear un modelo de trazas que dé apoyo y contexto a los requisitos pero sin que perturbe a los modelos de requisitos, y especificar un proceso de administración de trazas que involucre facilidad de creación y navegación.

Palabras clave: *trazabilidad, requisitos, gestión de requisitos*

1. INTRODUCCIÓN

Los requisitos evolucionan como consecuencia de cambios que el Universo de Discurso (UdeD) impone al software. Numerosos estudios realizados en las décadas de los 80 y los 90 del siglo pasado han mostrado que el 50% o más de los requisitos van a cambiar antes que el sistema de software se ponga en operación [Kotonya 98]. Si bien estas estadísticas reflejan la situación de la industria del software en una época en la que poco se había avanzado en la Ingeniería de Requisitos, la casi totalidad de los investigadores en el área siguen atestiguando que los cambios en los requisitos de un sistema de software es una realidad vigente en la actualidad. Es por esta razón que se debe poner énfasis en especificaciones de requisitos que sean capaces de evolucionar a lo largo del proyecto de desarrollo y que permitan el seguimiento de los mismos desde cualquier punto del proceso de producción del software hasta sus orígenes.

La gestión de requisitos es la actividad que administra los cambios en los requisitos. Estos cambios son de dos tipos: cambios en requisitos existentes o aparición de nuevos requisitos. Los factores que provocan estos cambios son básicamente la evolución del UdeD, incluyendo cambios en las expectativas de los clientes y usuarios y el descubrimiento de defectos en la comprensión o definición de los requisitos en

subsiguientes etapas del proyecto.

Para que la gestión sea una actividad viable y exitosa, es indispensable la administración de las dependencias entre los requisitos y la administración de las vinculaciones entre el documento de requisitos y otros documentos, modelos y componentes del software [Kotonya 98] [Davis 99]. Pues un cambio o incorporación de un único requisito puede alterar un conjunto de requisitos en cualquiera de los Modelos de Procesos de Software, y dependiendo de la etapa en que se encuentre el proyecto es necesario la actualización de la documentación y otros artefactos del software, lo cual debe identificarse y evaluarse previo a su efectiva realización. Esto está intrínsecamente relacionado con un concepto clave: la trazabilidad de los requisitos. La trazabilidad de los requisitos se refiere a la habilidad de definir, capturar y seguir las pistas dejadas por los requisitos sobre otros elementos del ambiente de desarrollo del software y las pistas dejadas por dichos elementos sobre los requisitos [Pinheiro 96]. Por otro lado, esta actividad puede ser considerada desde la visión del ciclo de vida del software como parte de una actividad de mayor alcance: la Gestión de Configuración [Crnkovic 99].

Para identificar los requisitos afectados por los cambios, así como los documentos, modelos y componentes involucrados, se debe utilizar la información de trazabilidad. Es fundamental contar con alguna herramienta de soporte a la gestión pues generalmente se manejan grandes cantidades de información en distintos formatos, y es necesario conservar el rastro de cada cambio solicitado, en evaluación e implementado.

Otro problema que se presenta en la gestión de requisitos es el control del versionado [Sawyer 01] [CMMI 06], es decir, el mantenimiento de la historia de los cambios en los requisitos. El sistema de versionado [Conradi 98] es un soporte de servicios indispensable para garantizar una adecuada trazabilidad de los mismos, da una idea de la evolución de los requisitos, registrando todos los cambios realizados con su justificación.

La trazabilidad de los requisitos es una función de la gestión de requisitos, que se encarga de mantener vínculos entre requisitos dependientes, entre requisitos, diseño y código, y entre requisitos y fuentes de información que los originaron. Según [IEEE Std 830-1998], un requisito es rastreable si su origen es claro y si

facilita su referencia en la documentación de futuros desarrollos o mejoras. Kotonya & Sommerville ofrecen una definición más precisa [Kotonya 98]: un requisito es rastreable si se puede determinar quién lo sugirió, qué requisitos están relacionados con él y cómo se relaciona con otra información tal como: diseño del sistema, implementaciones y documentación del usuario. Muchos investigadores han estudiado el tema y realizado propuestas sobre la trazabilidad de los requisitos [Gotel 94] [Wieringa 95] [Palmer 96] [Jarke 98] [Pinheiro 04].

Se han propuesto distintas formas de trazabilidad de requisitos. Tanto Wieringa [Wieringa 95] como [IEEE Std 830-1998], mencionan dos tipos: Backward traceability (habilidad de rastreo de un requisito a sus fuentes) y Forward traceability (habilidad de rastreo de un requisito hacia los componentes de diseño y código). Mientras que Davis en [Davis 93] menciona cuatro tipos pues agrega el rastreo inverso en cada caso, es decir de una fuente a un requisito y de un componente a un requisito; CMMI [CMMI 06] la denomina rastreabilidad bidireccional. Por otro lado, Pinheiro en [Pinheiro 04] introduce otra clasificación: Inter-requirements traceability (habilidad de rastreo entre requisitos dependientes) y Extra-requirements traceability (habilidad de rastreo entre requisitos y otros artefactos). Como se observa, Wieringa y Davis hacen referencia a este último tipo de trazabilidad. Para realizar un cambio en un requisito es necesario poder rastrear tanto interna como externamente el requisito, con el fin de evaluar adecuadamente tanto los requisitos dependientes como los componentes que deberán adaptarse y, en ciertos casos, es importante analizar el origen del requisito.

La recolección y mantenimiento de la información de rastreo es de muy alto costo. Por lo tanto, se deben tener políticas que indiquen qué tipo de rastreos se realizarán y cómo se mantendrá dicha información. Como indica Kotonya & Sommerville [Kotonya 98], la información de rastreo que más habitualmente se mantiene en la práctica es la que corresponde a Inter-requirements traceability y a Forward Traceability desde el documento de requisitos al diseño.

A continuación se extrae la definición dada por Pinheiro & Goguen [Pinheiro 96] por su claridad y amplitud: “La trazabilidad de los requisitos se refiere a la habilidad de definir, capturar y seguir las pistas dejadas por los requisitos sobre otros elementos del ambiente de desarrollo del software y las pistas dejadas por dichos elementos sobre los requisitos”.

La trazabilidad no sólo se utiliza para administrar los cambios en los requisitos, sino que también es de fundamental ayuda para la verificación y validación de los requisitos y para el control del proceso de desarrollo [Palmer 96] [Davis 99], pues facilita detectar conflictos utilizando los vínculos establecidos entre los elementos rastreables, posibilita asegurar que decisiones tomadas

avanzado el desarrollo sean consistentes con decisiones tempranas, y permite verificar que todos los requisitos han sido implementados en el software, entre otros usos. A pesar de los múltiples propósitos que cubre la trazabilidad de los requisitos, muchas veces sólo es aplicada parcialmente. Esto es debido, por un lado, a su alto costo de producción y mantenimiento (gran diversidad de entidades rastreables) y, por otro lado, a la imperiosa necesidad de contar con herramientas automatizadas que permitan implementar adecuadamente la trazabilidad.

2. PERSPECTIVA DEL PROYECTO

En este proyecto se estudiarán mejores formas de proveer a la rastreabilidad de los requisitos de software, es decir, qué aspecto del negocio generó una determinada funcionalidad del software y cómo está resuelta en el software una determinada demanda del cliente o usuario.

Cuando se está en presencia de dos modelos sucesivos en el proceso de desarrollo de software, cualquiera sean estos, es natural pensar que las trazas de un modelo a otro pueden representarse por medio de una matriz de trazas o un mecanismo similar. Este tipo de soluciones ignora que la naturaleza de la traza es tal que hay información, muchas veces muy relevante, relacionada con cada vínculo entre modelos. Por otra parte, esa visión de dos modelos sucesivos es en casi todos los casos falsa ya que para la construcción de un modelo o producto en el proceso de desarrollo de software se debe recurrir a información de diversas fuentes, lo que hace necesario pensar en cubos y eventualmente hipercubos de trazas. Esto es realmente inviable en casos prácticos. El recurso al que se suele apelar consiste en utilizar vínculos explícitos para representar las trazas. Estos modelos basados en vínculos carecen de la expresividad necesaria para representar la complejidad de las trazas en el proceso de requisitos y peor aún contaminan los modelos en los que los vínculos están empotrados de tal manera que aquellos llegan a ser prácticamente ilegibles. Debe notarse que habitualmente los modelos ya tienen vínculos de naturaleza semántica por lo que se debería proceder además a diferenciar los vínculos por su propósito.

Visto con la perspectiva de las bases de datos se puede notar que la presencia de modelos relacionados por vínculos es, en alguna medida, la reintroducción de las ideas de las bases de datos jerárquicas y en red con punteros empotrados al dominio de la Ingeniería de Requisitos en particular y a la Ingeniería de Software en general.

El modelo de trazas que se aspira a diseñar dará apoyo a un proceso de requisitos específico [Leite 04] que se basa en modelos en lenguaje natural: Léxico Extendido del Lenguaje (glosario del universo de discurso), Escenarios Actuales (situaciones observables en el proceso del negocio), Escenarios Futuros (situaciones

esperadas en el negocio con la incorporación del nuevo software) y Especificación de Requisitos (descripciones explícitas de cada requisito del software).

Los escenarios futuros sirven de ancla para la pre y post trazabilidad, permitiendo el rastreo de los requisitos hacia sus orígenes (vinculando los escenarios futuros con los escenarios actuales y con el léxico extendido del lenguaje) y hacia el diseño y el código. Son un medio que facilita la gestión de los cambios en los requisitos a lo largo del ciclo de vida del software.

2.1. OBJETIVOS PROPUESTOS

Los objetivos propuestos para este proyecto son:

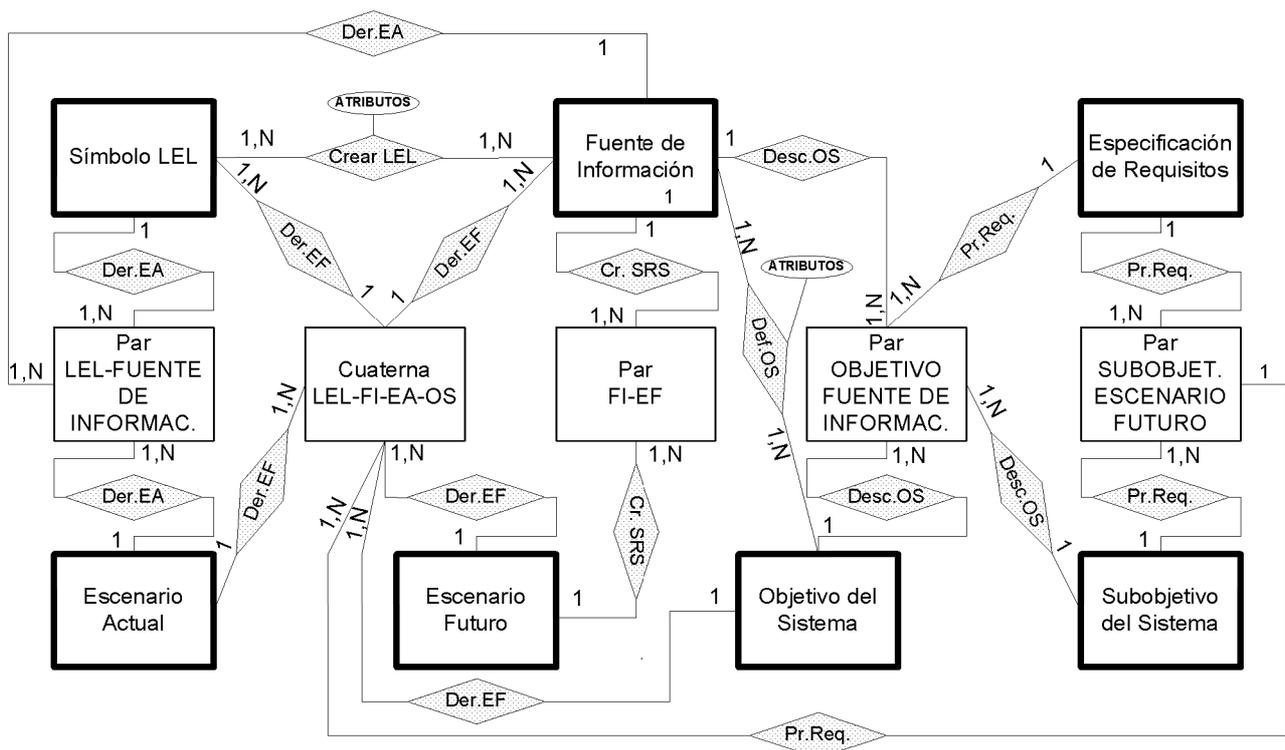
- Proveer un modelo de trazas que no interfiera con los modelos de requisitos.
- Definir un mecanismo de rastreo que brinde facilidad de navegación.
- Proveer un modelo de versionado.

Siendo los objetivos específicos los siguientes:

- Definir el modelo de trazas a ser utilizado (determinar qué componentes del modelo de trazas pueden ser capturados automáticamente).
- Definir el proceso de captura de trazas.
- Definir el modelo y el proceso de navegación con trazas.
- Definir el modelo de versionado para los modelos de requisitos.
- Detectar las relaciones existentes entre el modelo de trazas y el modelo de versionado.

3. ESTADO DE AVANCE

Se ha definido una primera versión del modelo de trazas, que se presenta en la Figura 1. Los componentes del proceso de requisitos se han modelado separando los datos visibles y manipulables por el usuario final de aquellos necesarios para el modelo de trazas. Estos últimos son los representados en la Figura 1 como entidades con marco grueso; no son los componentes propiamente dichos. Las entidades con marco fino corresponden a trazas de creación entre los distintos modelos de requisitos.



- | | |
|---|---|
| (Crear LEL) - Trazas de la creación del LEL | (Der.EF) – Trazas de la derivación de Escenarios Futuros |
| (Der.EA) – Trazas de la derivación de Escenarios Actuales | (Cr.SRS) – Trazas de la creación del SRS |
| (Def.OS) Trazas de la definición del Objetivo del Sistema | (Desc.OS) - Trazas de descomposición del objetivo del sistema |
| (Pr.Req.) - Trazas de la priorización de los requisitos | |

Figura 1 – Primera versión del modelo de trazas

Se han definido los atributos de las entidades del tipo Trazas. Se está estudiando los atributos a asignar a la entidad Fuentes de Información, la que por su variedad

es probable que se requiera construir una jerarquía con la misma.

Se planifica estudiar en detalle los procesos de construcción del LEL, derivación de escenarios actuales, entre otros, desde el punto de vista de las trazas para proveer mecanismos automáticos o semi automáticos de registro de las mismas.

4. FORMACION DE RECURSOS HUMANOS

El presente proyecto es parte de la tesis doctoral "Modelado del registro de trazas en la Ingeniería de Requisitos" que está desarrollando la Ing. Andrea Vera en la UNLP.

5. REFERENCIAS

- [CMMI 06] Software Engineering Institute, "Capability Maturity Model Integration", CMMI-DEV v1.2, CMU/SEI-2006-TR-008, Carnegie Mellon University, 2006, <http://www.sei.cmu.edu/cmmi/>.
- [Conradi 98] Conradi, R., Westfechtel, B., "Version Models for Software Configuration Management", ACM Computing Surveys, Vol.30, 1998, pp.232-282.
- [Crnkovic 99] Crnkovic, I., Funk, P., Larsson, M., "Processing Requirements by Software Configuration Management", 25th Euromicro Conference (EUROMICRO'99), IEEE Computer Society, Milán, Italia, Vol.2, Septiembre 1999, pp.2260-2265.
- [Davis 99] Davis, A., Leffingwell, D., "Making Requirements Management Work For You", Crosstalk, The Journal of Defense Software Engineering, Vol.12, N°4, Abril 1999.
- [Gotel 94] Gotel, O.C.Z., Finkelstein, A.C.W., "An analysis of the requirements traceability problem", ICRE'94, First IEEE International Conference on Requirements Engineering, IEEE Computer Society Press, Colorado Springs, Abril 1994, pp.94-101.
- [IEEE Std 830-1998] IEEE Recommended Practice for Software Requirements Specifications (ANSI), IEEE, Nueva York, 1998.
- [Jarke 98] Jarke, M., "Requirements tracing", Communications of the ACM, Vol.41, N°12, Diciembre 1998, pp.32-36.
- [Kotonya 98] Kotonya, G., Sommerville, I.: Requirements Engineering: Processes and Techniques. John Wiley & Sons, 1998.
- [Leite 97] Leite, J.C.S.P., "Software Evolution, The Requirements Engineering View", keynote address en anales de XXVI JAIIO - SoST'97 Simposio en Tecnología de Software, Buenos Aires, Argentina, Agosto 1997, pp.21-23.
- [Leite 04] Leite, J.C.S.P., Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., Ridao, M.N., "Defining System Context using Scenarios", en el libro "Perspectives on Software Requirements", Kluwer Academic Publishers, EEUU, ISBN: 1-4020-7625-8, capítulo 8, 2004, pp.169-199
- [Palmer 96] Palmer, J.D., "Traceability", en Software Engineering, editores M. Dorfman y R.H. Thayer, IEEE Computer Society Press, 1996, pp.266-276. Reimpreso en "Software Requirements Engineering", editores R.H. Thayer y M. Dorfman, IEEE Computer Society Press, 2º edición, Los Alamitos, CA, 1997, pp.364-374.
- [Pinheiro 04] Pinheiro, F.A.C., "Requirements Traceability", en el libro "Perspectives on Software Requirements", Kluwer Academic Publishers, Estados Unidos, ISBN: 1-4020-7625-8, capítulo 5, 2004, pp.91-113.
- [Pinheiro 96] Pinheiro, F.A.C., Goguen, J.A., "An object-oriented tool for tracing requirements", IEEE Software, Special issue of papers from ICRE'96, Vol.13, N°2, Marzo 1996, pp.52-64.
- [Sawyer 01] Sawyer, P., Kotonya, G., "Software Requirements", SWEBOK, Guide to the Software Engineering Body of Knowledge, editores P. Bourque y R. Dupuis, IEEE Computer Society, Los Alamitos, CA, capítulo 2, 2001, pp.31-56, IEEE Trial Version 1.00, http://www.swebok.org/stoneman/trial_1_00.html.
- [Wieringa 95] Wieringa, R.J., "An introduction to requirements traceability", Reporte Técnico IR-389, Faculty of Mathematics and Computer Science, University of Vrije, Amsterdam, Septiembre 1995.

PROCEDIMIENTOS DE LA EXPLOTACIÓN DE INFORMACIÓN PARA LA IDENTIFICACIÓN DE DATOS FALTANTES, CON RUIDO E INCONSISTENTES

Horacio Kuna, Ramón García-Martínez, Francisco Villatoro Machuca

Depto. de Informática, Facultad de Ciencias Exactas Químicas y Naturales. Universidad Nacional de Misiones.
Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. Universidad de Buenos Aires.
Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga.

hdkuna@unam.edu.ar, rgarciamar@fi.uba.ar

CONTEXTO

Esta línea de investigación se articula con los proyectos UBACyT 2008-2010-I012 “Aplicaciones de Explotación de Información Basada en Sistemas Inteligentes” y el “Programa de Investigación en Computación” de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones; y con las líneas del Doctorado en Ingeniería de Sistemas y Computación que funciona en el marco del convenio entre la Universidad Nacional del Nordeste y la Universidad de Málaga.

RESUMEN

La información se ha convertido, en la columna vertebral de las organizaciones, la aplicación de distintas técnicas, métodos y herramientas para garantizar mediante un proceso formal de Auditoría, la calidad y seguridad de la información es una tarea de significativa importancia.

En la actualidad no se encuentran procedimientos formales especialmente diseñados para aplicar técnicas de explotación de información en la Auditoría de Sistemas en general y a la búsqueda de datos con ruido, inconsistentes y faltantes, aplicándose en algunos casos metodologías diseñadas con otros objetivos como SEMMA o CRISP que no contemplan la especificidad de los objetivos que se persiguen, en otros casos no se aplica una metodología, esta situación provoca una disminución en la calidad del proceso de Auditoría de Sistemas.

Este proyecto busca establecer una taxonomía relacionada con la calidad de los datos, analizando los procesos de explotación de información que mejor aplican a la identificación de patrones de pistas de auditoría, se explorarán esas procesos analizando las ventajas y desventajas de cada una de ellos.

Palabras clave: procesos de explotación de información, auditoría de sistemas, pistas de auditoría.

1. INTRODUCCION

1.1 Auditoría de Sistemas

El actual estado de desarrollo de los sistemas de información hace que los mismos sean más complejos, integrados y relacionados. La administración efectiva de la Tecnología de la Información (TI) es un elemento crítico para la supervivencia y el éxito de las empresas, varias son las razones que producen esta criticidad, son por ejemplo, la dependencia que tienen las organizaciones de la información para su funcionamiento, el nivel de inversión que tienen en el área de TI, la potencialidad que tiene la TI para transformar las organizaciones, los riesgos y amenazas que en la actualidad tiene la información, la economía globalizada que exige un alto nivel de competitividad, entre otras. Existe una relación cada vez más fuerte entre los objetivos estratégicos de una empresa y la TI, se debe implementar un sistema adecuado de control interno que permita proteger todos los elementos relacionados con la TI, el personal, las instalaciones, la tecnología, los sistemas de aplicación y los datos.

Esto hace que sea cada vez más necesario en todas las organizaciones y no sólo en las grandes, el garantizar el cumplimiento de las normas y procedimientos establecidos para el manejo de las políticas relacionadas con la Tecnología de la Información. Existe una creciente necesidad de garantizar la seguridad y calidad de los servicios que se brindan dentro de una empresa en relación con la TI. Son muchos los riesgos que amenazan los recursos relacionados con la TI, por ejemplo, accesos indebidos a las bases de datos, falsificación de información para terceros, incumplimiento de leyes y regulaciones, fraudes, virus, destrucción de soportes documentales, acceso clandestinos a redes, entre otros.

La Auditoría de Sistemas es la evaluación sistemática de todos los aspectos relacionados con la Tecnología de la Información, uno de sus objetivos es proteger los activos que tienen las empresas y organizaciones, la información en este mundo globalizado es uno de los principales activos a resguardar. La detección de ruidos, inconsistencias o

incompletitud en los datos es fundamental en el proceso de auditoría ya que brindan pistas de posibles problemas necesarios de detectar y corregir, como por ejemplo, accesos no autorizados a las bases de datos, errores en los sistemas, etc. Utilizar métodos, técnicas y herramientas que asistan al auditor en la tarea de encontrar anomalías en las bases de datos es de suma importancia ya que hacen su trabajo más eficiente, eficaz y objetivo.

A nivel internacional existen diferentes normas que intentan estandarizar el proceso de la auditoría de sistemas, una de estos estándares es COBIT [COBIT, 2008] cuya misión es investigar, desarrollar, publicar y promover objetivos de control en tecnología de la información (TI) con autoridad, actualizados, de carácter internacional y aceptados generalmente para el uso cotidiano de gerentes de empresas y auditores. La *Information Systems Audit and Control Foundation* [COBIT, 2008] y los patrocinadores de COBIT, han diseñado este producto principalmente como una fuente de buenas prácticas para los auditores de sistemas. COBIT ha sido desarrollado como estándares para mejorar las prácticas de control y seguridad de las TI que provean un marco de referencia para la Administración, Usuarios y Auditores.

Existen distintos tipos de auditorías de sistema [Piattini 2003]: auditoría física, auditoría de la ofimática, auditoría de la dirección, auditoría de la explotación, auditoría del desarrollo, auditoría del mantenimiento, y auditoría de bases de datos, entre otras.

La norma *Statement on Auditing Standards 1009* [SAS, 2008] define a las Técnicas de Auditoría Asistida por Computadora (TAACs) como el conjunto de datos y programas que utiliza el auditor durante el desarrollo de su tarea, y explicita los más importantes pasos que el auditor de sistemas debe considerar cuando prepara la aplicación de las TAACs:

- Establecer los objetivos de auditoría de las TAACs.
- Determinar accesibilidad y disponibilidad de los sistemas de información, los programas/sistemas y datos de la organización.
- Definir los procedimientos a seguir (por ejemplo: una muestra estadística, recálculo, confirmación, entre otros).
- Definir los requerimientos de salida (*output*).
- Determinar los requerimientos de recursos.
- Documentar los costos y los beneficios esperados
- Obtener acceso a las facilidades de los sistemas de información de la organización, sus programas/sistemas y sus datos.
- Documentar las TAACs a utilizar incluyendo los objetivos, flujogramas de alto nivel y las instrucciones a ejecutar.

1.2 Explotación de Información

Se define la Explotación de Información (*Data Mining*) [Clark, 2000] como el proceso mediante el cual se extrae conocimiento comprensible y útil que previamente era desconocido desde bases de datos, en diversos formatos, en forma automática. Es decir, la Explotación de Información plantea dos desafíos, por un lado trabajar con grandes bases de datos y por el otro aplicar técnicas que conviertan en forma automática estos datos en conocimiento.

La Explotación de Información es un elemento fundamental de un proceso más amplio que tiene como objetivo el descubrimiento de conocimiento en grandes bases de datos [Fayyad *et al.* 1996; Britos *et al.*, 2005], en inglés "*Knowledge Discovery in Databases*" (KDD), este proceso, como lo muestra la figura 1, tiene una primer etapa de preparación de datos, luego el proceso de minería de datos, la obtención de patrones de comportamiento, y la evaluación e interpretación de los patrones descubiertos.

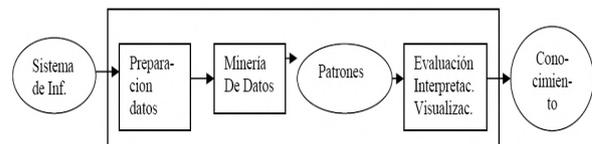


Fig.1. Proceso de KDD

1.3 Explotación de información y Auditoría de sistemas

El mayor desarrollo del uso de la Explotación de Información en actividades relacionadas con la auditoría de sistemas se relacionan con la detección de intrusos en redes de telecomunicaciones, también se encuentra en la literatura científica antecedentes relacionados con la detección de fraudes [Britos *et al.*, 2008b], análisis de logs de auditoría, no encontrándose antecedentes de la Explotación de Información en la búsqueda de datos faltantes, con ruido e inconsistentes en bases de datos.

Ante la necesidad existente de brindar al incipiente mercado una aproximación sistemática para la implementación de proyectos de Explotación de Información, diversas empresas [Britos *et al.*, 2008a] han especificado un proceso de modelado diseñado para guiar al usuario a través de una sucesión formal de pasos:

- SAS [2008] propone la utilización de la metodología SEMMA [SEMMA 2008] (Sample, Explore, Modify, Model, Assess).
- En el año 1999 un grupo de empresas europeas, NCR (Dinamarca), AG (Alemania), SPSS (Inglaterra) y OHRA (Holanda), desarrollaron una metodología de libre distribución CRISP-DM (Cross-Industry Standard Process for Data Mining) [CRISP-DM, 2008].
- La metodología P3TQ [Pyle, 2003] (Product, Place, Price, Time, Quantity), tiene dos modelos,

el Modelo de Explotación de Información y el Modelo de Negocio.

2. LINEAS DE INVESTIGACION y DESARROLLO

Existen procedimientos formales y globalmente establecidos relacionados con el uso genérico de las TAACs y para la implementación de un proceso de descubrimiento de conocimiento en grandes bases de datos, pero no existe un método para la aplicación específica de la Explotación de Información en la obtención de pistas de auditoría, otro problema es que existen trabajos relacionados con la comparación de las distintas técnicas de Explotación de Información aplicadas en general a la auditoría de sistemas pero no se encuentran antecedentes en lo relacionado al análisis de las distintas técnicas aplicadas a la búsqueda de datos faltantes, con ruido e inconsistentes. Ante esta situación aparecen dos realidades a la hora de implementar en el proceso de auditoría de sistemas la Explotación de Información, en algunos casos no se aplica ninguna metodología formal y en otros casos surge la necesidad de adaptar las metodologías existentes para implementar la Explotación de Información a la particularidad que implica utilizar esta tecnología en la Auditoría de Sistemas, siendo este proceso de adaptación empírico e informal.

Existen antecedentes de procedimientos para la implementación de un proceso de explotación de información, pero no procedimientos específicos de explotación de información para datos faltantes, con ruido e inconsistentes, en ese contexto en este proyecto se propone establecer procedimientos que identifiquen este tipo de datos. Se espera establecer una taxonomía relacionada con la calidad de los datos, analizando las técnicas de minería que mejor aplican, se explorarán esas técnicas analizando las ventajas y desventajas de cada una de ellas.

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto presentado comenzó a fines del año 2008 y tiene prevista como tareas para su primer año:

- Refinamiento investigación documental orientada a la identificación de trabajos previos vinculados a la explotación de información aplicados al proceso de auditoría de sistemas.
- Identificación de técnicas de explotación de información aplicadas la auditoría de sistemas.
- Identificación de casos de estudio aceptados por la comunidad internacional de aplicación de explotación de información en la auditoría de sistemas para su utilización en pruebas de concepto y validación del proyecto.

- Identificación y delimitación de problemas vinculados a la detección de datos faltantes con ruido e inconsistentes en bases de datos.
- Exploración de la aplicabilidad de técnicas de exploración de información a la resolución de los problemas planteados.

4. FORMACION DE RECURSOS HUMANOS

En el marco de este proyecto se esta desarrollando una tesis doctoral y se prevé el inicio de dos tesis de grado. Esta línea vincula al Grupo de Auditoría del Departamento de Informática de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones, al Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires y al Departamento de Lenguajes y Ciencias de la Computación de la Universidad de Málaga.

5. BIBLIOGRAFIA

- Britos, P.; Hossian, A.; García Martínez, R.; Sierra, E. 2005. *Minería de Datos Basada en Sistemas Inteligentes*. Nueva Librería,
- Britos, P.; Dieste, O.; García Martínez, R. 2008. *Requirements Elicitation in Data Mining for Business Intelligence Projects*. En: *Advances in Information Systems Research, Education and Practice*. Springer, p. 139–150.
- Britos, P.; Grosser, H.; Rodríguez, D.; García Martínez, R. 2008. *Detecting Unusual Changes of Users Consumption*. In *Artificial Intelligence and Practice II*. Springer. p. 297-306.
- Clark, P.; Boswell R. 2000. *Practical Machine Learning Tools and Techniques with Java Implementation*. Morgan Kaufmann Publisher.
- COBIT. 2008. *Control Objectives for Information and related Technology*. <http://www.isaca.org/cobit/>. Vigencia 16/04/08.
- CRISP-DM. 2008. <http://www.crisp-dm.org/>. Vigencia 15/09/08.
- Fayyad U.M.; Piatetsky Shapiro G.; Smyth P. 1996. *From Data Mining to Knowledge Discovery: An Overview*. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, p 1-34.
- Piattini, M.; Peso, E. 2003. *Auditoría Informática, un enfoque práctico*. Alfaomega-Rama,
- Pyle, D. 2003. *Business Modeling and Business intelligence*. Morgan Kaufmann Publishers,
- SAS. 2008. *Statement on Auditing Standards*. <http://www.aicpa.org/>. Vigencia 15/09/08.
- SEMMA. 2008. <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>. Vigencia 15/09/08.

SERVICIOS: EXPLORACIONES EN SOA y WEB.

López, G.¹; Jeder, I.¹; Echeverría, A.¹; Grossi, M.D.²; Servetto, A.²; Fierro, P. (PhD.)³

1. Laboratorio de Informática de Gestión - Facultad de Ingeniería. UBA.
Tel: 54-11-4343-0891. Extensión 141
e-mail: glopez@fi.uba.ar - Web: <http://www.fi.uba.ar/laboratorios/lig>
2. Laboratorio de Sistemas Operativos y Bases de Datos - Facultad de Ingeniería. UBA.
Tel: 54-11-4343-0891. Extensión 142
e-mail: aserve@gmail.com - <http://laboratorios.fi.uba.ar/lso/lasobd.htm>
3. Departamento de Estudios e Investigación de la Empresa - Facultad de Economía. Universidad de Salerno. Italia
Tel: + 390815519623
e-mail: fierrop@unisa.it – Web: <http://www.unisa.it>

CONTEXTO.

Esta línea de investigación forma parte del Proyecto de Investigación Acreditado UBACYT I422 “Arquitectura Orientada a Servicios – Composición Dinámica de Servicios basada en Agentes Inteligentes Autónomos”.

RESUMEN.

Service Oriented Architecture (SOA) es un paradigma para organizar y utilizar capacidades distribuidas que pueden estar bajo el control de diferentes dominios propietarios.

Cuando un servicio utiliza Internet como medio de comunicación así como estándares abiertos, basados en Internet, se está frente a un servicio WEB.

La tendencia observada indica que en todo lugar en el que se utilicen computadoras y teléfonos, los servicios móviles tienen el futuro asegurado y cada vez más demandante de capacidades variadas.

El presente trabajo pretende contribuir a la profundización del estudio de herramientas metodológicas para el desarrollo de aplicaciones distribuidas, orientadas a servicios y a negocios, destacando la complejidad creciente de las mismas, en particular, en lo que a corrección y confiabilidad se refiere.

PALABRAS CLAVE: Servicio, SOA, WEB, sistemas distribuidos.

INTRODUCCION.

De acuerdo con OASIS¹, Service Oriented Architecture (SOA) es un paradigma para organizar y utilizar capacidades distribuidas que pueden estar bajo el control de diferentes dominios propietarios.

En este documento, se adopta, entonces, tal concepto como definición de SOA.

Por otra parte, se entiende que ingeniería de software es la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento de software, así como el estudio de estos enfoques, en resumen, es la aplicación de la ingeniería al software.

¹ Organization for the Advancement of Structured Information Standards, <http://docs.oasis-open.org>

De acuerdo con la tendencia observada en todo lugar en el que se utilicen computadoras y teléfonos, no resulta descabellado suponer que para la población, los servicios públicos, servicios de seguridad, el comercio y la mayoría de las actividades de entretenimiento, relaciones sociales y ocio en general, los dispositivos móviles tienen el futuro asegurado y un mercado cada vez más demandante de capacidades variadas.

Servicios.

El término “orientado a servicios” podría interpretarse de varias maneras. Una de ellas es que representa un método de separación de incumbencias. Esto significa que la lógica requerida para resolver un problema complejo puede ser construida, llevada a cabo y administrada de una mejor manera si la misma se descompone en una colección de pequeñas piezas interrelacionadas. Cada una de estas piezas resuelve un parte específica del problema.

Cuando se aplica a la arquitectura, se utiliza un término técnico conocido como “Arquitectura Orientada a Servicios” - SOA. Esta expresión representa un modelo en el cual la lógica de automatización es descompuesta en pequeñas unidades de lógica. En conjunto, estas unidades conforman una pieza mayor de lógica de automatización de negocios. Un beneficio de esta división es que estas unidades individuales pueden ser distribuidas.

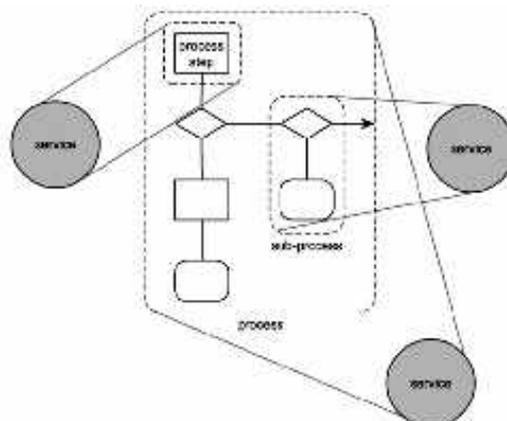
En este documento, el término negocio/s se define como cualquier actividad de interés del usuario, especialmente actividades que comparten múltiples usuarios. [1]

SOA promueve que las unidades individuales de lógica existan autónomamente, si bien comunicadas entre ellas. Estas unidades deben estar relacionadas de manera tal de poder evolucionar independientemente y al mismo tiempo, mantener cierta estandarización con las demás unidades. En el marco de SOA, estas unidades se conocen como servicios. En efecto, según OASIS, se entiende por servicio “un mecanismo para habilitar acceso a una o más capacidades, donde el acceso es provisto utilizando una interfaz prescrita y es ejecutado de manera consistente con restricciones y políticas especificadas por la descripción del servicio [2].

Para mantener su independencia, los servicios encapsulan la lógica dentro de un contexto. Este contexto puede ser una tarea de negocio, una entidad de negocio o alguna otra agrupación lógica.

Las incumbencias resueltas por un servicio pueden ser pequeñas o grandes. Por lo tanto, el tamaño y alcance de la lógica representada por los servicios puede variar. Además, la lógica de un servicio puede requerir de la lógica provista por otros servicios. En este caso, un servicio necesitará la colaboración de otros para resolver su problema.

Cuando se construye una solución automatizada basada en servicios, cada uno de éstos puede encapsular una tarea realizada por una rutina individual, un proceso compuesto de varias rutinas. Un servicio puede además encapsular la lógica completa del proceso. El siguiente diagrama ilustra lo dicho:



La interfaz de servicio es la descripción de las capacidades del servicio, esto es, de las operaciones que el servicio puede ofrecer. Por otra parte, la especificación describe explícitamente las interfaces que un cliente del mismo pudiera acceder. También, deben ser explícitas las interfaces que proveerá el ambiente de software en el se manifestará el servicio.

Para que un servicio sea de utilidad general, deberá contar con cualidades tales como bajo acoplamiento y transparencia de ubicación.

Servicios WEB.

Cuando un servicio utiliza Internet como medio de comunicación así como estándares abiertos, basados en Internet, se está frente a un servicio WEB.

En este documento se entiende, entonces, por servicio WEB, aquel servicio que queda identificado por una URI (Uniform Resource Identifier) y cumple con estas características mínimas: desde el punto de vista textual, presenta sus capacidades mediante códigos de programas escritos en lenguajes estándares en Internet así como protocolos que pueden implementarse mediante la observación de la interfaz y su propia descripción, basado en estándares abiertos, tales como XML.

Estos estándares permiten a un conjunto de sistemas de computación operar entre ellos sin importar sus plataformas, sistemas operativos, infraestructura de red o lenguaje de programación.

En definitiva, un servicio WEB es un software que puede ejecutarse a partir de una solicitud de cualquier computadora conectada a una red.

La mayoría de los servicios WEB implementan un diseño del tipo solicitud/respuesta (request/replay), en la mayoría de las arquitecturas distribuidas. Este tipo de interoperabilidad es la que promueve SOA.

La ventaja de los servicios WEB es que, en contraste con los sistemas distribuidos tradicionales, utilizan interfaces que se encuentran ampliamente aceptadas. Por ejemplo, para el envío de mensajes se basan en un tipo especial de XML conocido como SOAP.

Un mensaje SOAP puede ser enviado de una computadora a otra utilizando distintos protocolos: http, ftp, smtp, etc. Como ambas computadoras comprenden SOAP como un lenguaje común, la computadora receptora es capaz de procesar el mensaje y responder también utilizando SOAP.

Una de las mayores ventajas de utilizar servicios WEB en la implantación de sistemas distribuidos es la universalidad de las interfaces.

Dado que un servicio WEB puede enviar y recibir mensajes SOAP sobre protocolos de Internet, es posible interoperar con un servicio mediante la utilización de cualquier tipo de computadora.

Un servicio WEB corriendo en un mainframe con sistema operativo propietario podrá comunicarse con un UNIX, Windows o servidor Sun sin necesidad de modificar su interfaz.

Esta facilidad de interoperabilidad es una gran ventaja respecto de las soluciones propietarias.

Existen tres estándares distinguidos basados en XML, en relación con los servicios WEB:

SOAP – Simple Object Access Protocol: formato de mensajes.

WSDL – Web Services Description Language: posibilita la descripción de qué hace el servicio y cómo invocarlo.

UDDI – Universal Discover, Description and Integration: directorio de servicios WEB disponibles para su utilización.

Estos tres estándares se combinan para posibilitar a un servicio WEB, funcionar, describirse y ser encontrado en el ámbito de una red.

Servicios Internet- móviles.

De acuerdo con la tendencia observada en todo lugar en el que se utilicen computadoras y teléfonos, no resulta descabellado suponer que para la población, los servicios públicos, servicios de seguridad, el comercio y la mayoría de las actividades de entretenimiento, relaciones sociales y ocio en general,

los dispositivos móviles tienen el futuro asegurado y cada vez más demandante de capacidades variadas.

Las redes de dispositivos móviles y sus servicios asociados, entonces, serán más abundantes, complejas y extendidas en las diferentes sociedades humanas que las redes de computadoras convencionales (pcs de escritorio, workstations, servers, mini o macro computadoras).

Los servicios requeridos por los usuarios de tales redes móviles, deberán satisfacer cada vez más exigencias de QoS, corrección y confiabilidad de los sistemas de información, facilidad de acceso a los servicios, abaratamiento considerable de los costos operativos y de infraestructura para la población, alto rendimiento, bajo mantenimiento, facilidad de actualización de los diversos componentes involucrados.

En este sentido, se han estudiado [3] algunos aspectos de las aplicaciones de software que deberán mejorar significativamente de manera asegurar la aceptación del usuario.

Algunas de estas características son:

- Infraestructura no disponible,
- Servicio no disponible,
- Fallas durante la utilización del servicio,
- Incompatibilidades entre dispositivos del mismo tipo.

Respecto de la infraestructura no disponible, esto podría también ser un problema relacionado con cuestiones que quedan fuera de lo estrictamente técnico o tecnológico.

El servicio no disponible es una falla que puede estar relacionada con la corrección del sistema así como la estrategia de pruebas que se haya desarrollado durante la construcción del sistema.

La tercera característica también se relaciona con la corrección, confiabilidad y robustez del sistema desarrollado. La corrección de un software distribuido, en el que se manifiestan fenómenos de paralelismo o alta concurrencia, se convierte en un problema de no fácil solución y tampoco barato, dando como consecuencia, la posibilidad de liberación de software con alta tasa de errores o fallas no detectadas al momento de ofrecerlo a los usuarios.

Respecto de las incompatibilidades entre dispositivos del mismo tipo, la ausencia de acuerdos en la adopción de estándares o reglas mínimas de fabricación o desarrollo producen este tipo de fenómenos no deseados.

Comunicación entre servicios.

Dado que los servicios se manifiestan con mayor significación, en ambientes distribuidos, la comunicación entre servicios es vía mensajes. Luego que un servicio envía un mensaje, éste pierde el control de lo que sucede con el mensaje. Es por esto que los mensajes deben ser autónomos e incluir suficiente inteligencia de manera tal de poder manejar su lógica interna de procesamiento.

De acuerdo con la naturaleza de comunicación entre servicios, es fácil ver el tipo de problemas así como la complejidad de los mismos, en cuanto a corrección, confiabilidad, disponibilidad de los servicios.

El control de la concurrencia, las propiedades de integridad y seguridad (safety) así como la interacción pueden crear demandas conflictivas en un sistema distribuido [4].

Otro problema para profundizar es la complejidad computacional de los algoritmos que soportan la interacción de servicios.

Se puede afirmar que SOA se posiciona como una herramienta de ingeniería informática especialmente apta para la implementación de sistemas distribuidos y la integración de aplicaciones dentro y a través de las fronteras de las empresas [5].

RESULTADOS OBTENIDOS/ESPERADOS

Los protocolos de negocios, así como la composición de servicios de negocios constituyen destacables aspectos innovadores respecto de los servicios WEB. Se espera, entonces, avanzar en el estudio de la composición de servicios dado que es razonable pensar que será un aspecto principal en el desarrollo de los mismos.

Los protocolos de negocios definen cuáles secuencias de intercambio de mensajes, es decir cuál comunicación será soportada por el servicio.

Se espera poder profundizar el estudio de servicios en servicios en redes móviles.

Se continuará con el estudio de la evolución del software como un servicio manifestándose en una arquitectura de muy bajo acoplamiento, expresado mediante la utilización de estándares abiertos basados en XML y con vías de optimizar la comunicación y sincronización entre los mismos, sostenidos por Internet.

FORMACION DE RECURSOS HUMANOS

En el ámbito del Laboratorio de Gestión, se ha completado una tesis de grado en el tema y se están incorporando nuevos estudiantes avanzados para el estudio del tema así como la elaboración de nuevas tesis.

De esta manera, se puede afirmar que el grupo de interés por el tema está creciendo en el seno del Laboratorio de Gestión.

REFERENCIAS

[1] <http://docs.oasis-open.org/soa-rm/soa-ra/v1.0/soa-ra-pr-01.html>

[2] <http://docs.oasis-open.org/soa-rm/v1.0/>

[3] D.Cotroneo, C.Di Flora, S.Russo, "Improving Dependability of Service Oriented Architectures for Pervasive Computing", Proceedings of the Eighth IEEE International Workshop on Object-Oriented Real Time Dependable Systems, IEEE, 2003.

[4] Allison C., Bateman M, Nicoli, R., Ruddle A., "Adaptive QoS for Collaborative Service-Oriented Learning Environments", International Symposium on Cluster Computing and the Grid, IEEE, 2004.

[5] Alonso, G, Casati, F, "Web services and service - oriented architectures", Proceedings of the 21st International Conference on Data Engineering, ICDE 2005.

ESTEREOTIPOS DE CLASES ENTIDAD Y CLUSTERING DE OBJETOS EN SISTEMAS DE GESTION

Servetto, A.¹; López, G.²; Jeder, I.²; Echeverría, A.²; Grossi, M.D.¹

1. Laboratorio de Sistemas Operativos y Bases de Datos - Facultad de Ingeniería. UBA.
Tel: 54-11-4343-0891. Extensión 142
e-mail: aserve@gmail.com
2. Laboratorio de Informática de Gestión - Facultad de Ingeniería. UBA.
Tel: 54-11-4343-0891. Extensión 141
e-mail: glopez@fi.uba.ar

CONTEXTO

Esta línea de investigación forma parte de un proyecto del Laboratorio de Sistemas Operativos y Bases de Datos cuyo objetivo es el desarrollo de un Manejador de Bases de Objetos para Sistemas de Gestión, respetando los principios del software libre [1] y siguiendo los lineamientos del OMG [2].

RESUMEN

En las metodologías de desarrollo de software de gestión con objetos, los modelos de dominio se realizan en la etapa de elaboración o de análisis, y a medida que se avanza en el desarrollo las clases entidad de estos modelos se insertan en distintos diagramas de diseño correspondientes a la realización de casos de uso que responden a modelos de ejecución y no de persistencia, y el modelo integral de dominio se va refinando según esquemas de navegación que surgen de los diagramas de diseño [3]. El rendimiento de los manejadores de bases de objetos se puede optimizar si se consideran estereotipos de clases entidad propios de los sistemas de gestión que permitan clusterizar objetos en función de los patrones de acceso característicos.

Los propósitos de la línea de investigación y desarrollo que se presenta en este trabajo son establecer estereotipos de clases entidad para sistemas de gestión que faciliten el diseño de bases de objetos y estudiar y ensayar técnicas de clustering para manejadores de bases de objetos que permitan optimizar la recuperación.

PALABRAS CLAVE

bases, objetos, clusterización, estereotipos, clases, gestión

INTRODUCCION

Las clases de objetos en un sistema de gestión se pueden clasificar o estereotipar según la clase de datos que clasifiquen. En general pueden ser

- *De Datos Maestros*: objetos de un sistema de información que representan entidades de existencia real o ideal, por ejemplo productos o servicios, o valores de referencia para determinar características o atributos de otros datos (dominios de atributos definidos por extensión).
- *De Datos Transaccionales*: registros de hechos o eventos relacionados con datos maestros, por ejemplo de ventas de productos o de prestaciones de servicios.

Los datos transaccionales a su vez pueden diferenciarse según la posibilidad de que las transacciones que representan sean actualizables o no:

- Las *transacciones actualizables* son aquellas de las cuales se registra su previsión o programación en el tiempo, antes de que la transacción se produzca, y una vez producida se completan sus características (por ejemplo turnos para servicios que devienen en registros de

servicios).

- Las *transacciones no actualizables* son aquellas que se registran luego de producidas, como si fuera en una bitácora, y luego no se actualizan más (por ejemplo registros de facturación o de operaciones bancarias).

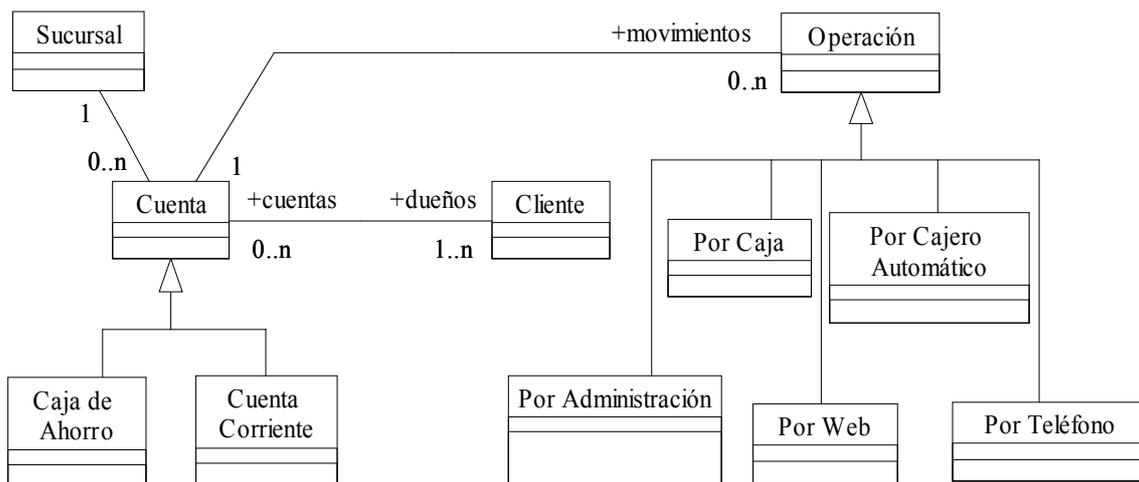
Las **operaciones con objetos** pueden ser de navegación o de consulta. Las operaciones de navegación implican accesos por identificador automático o del manejador, y las de consulta, expresadas en OQL (*Object Query Language*), por el índice de identificación de los objetos en el sistema de gestión o por otros criterios, que pueden sustentarse en otros índices.

En las bases de objetos los **identificadores de objetos** son automáticos y transparentes al usuario, por lo que no garantizan la unicidad de los objetos (es posible instanciar más de una vez un mismo objeto para el sistema de información pero con distintos identificadores): para garantizar la unicidad de los objetos se impone definir un criterio de identificación que sea propio del sistema de información. En el manejador que se está desarrollando se exige que para toda clase se precise qué atributo o conjunto de atributos identifica unívocamente a cada objeto, como es característico en el diseño conceptual de bases de datos.

LINEAS DE INVESTIGACION Y DESARROLLO

En aras de la flexibilidad y en atención a que en los sistemas de información los objetos transaccionales se suelen identificar naturalmente a partir del identificador del objeto maestro con el que se relaciona la transacción más un indicador de tiempo, en el manejador de bases de objetos de este proyecto se admiten identificadores mixtos y externos [4], es decir, compuestos por identificadores de objetos maestros más un discriminante propio, en el caso de los mixtos, o compuestos exclusivamente por identificadores de objetos maestros, en el caso de los externos.

Para la composición de los identificadores mixtos y externos, en el ODL (*Object Definition Language*) se usa el nombre de las clases de objetos maestros, que en los objetos se traduce al identificador automático o del sistema. Por ejemplo, para definir objetos transaccionales que representen operaciones de una cuenta bancaria se define una clase Operación, dependiente de una clase Cuenta:



La definición en XML (como estándar de representación) quedaría:

```
<clase nombre="Operación" tipo="TNA" instanciable="no">
  <atr nombre="momento" tipo="tiempo"/>
  <atr nombre="movimiento" tipo="(DE|CR)"/>
  <atr nombre="tipo" tipo="Tipo Operación"/>
  <atr nombre="monto" tipo="fracc"/>
</clase>
```

```

    <id tipo="mixto">
        <comp tipo="ext" pos="1" clase="Cuenta"/>
        <comp tipo="int" pos="2" atr="momento" orden="desc"/>
    </id>
</clase>

```

En el fragmento de XML precedente se define la clase Operación, de tipo Transaccional No Actualizable y no instanciable (abstracta). Los atributos de esta clase son momento, que registra fecha y hora de la operación; movimiento, que registra si es un DE(bit) o un CR(édito); el tipo (definido por el administrador por extensión en la clase Tipo de Operación) y el monto. El identificador de la clase en el sistema de información es mixto, y su primer componente, externo, es el identificador automático de la cuenta a la cual afecta la operación, y el segundo componente, interno, es la fecha y hora de la operación en orden descendente. Cada operación se asocia con una (multiplicidad mínima) y sólo una (multiplicidad máxima) cuenta, que sólo debe definirse en caso de que se pretenda navegar desde una operación hacia una cuenta.

La posibilidad de definir identificadores mixtos o externos representa una innovación en las bases de objetos, ya que al no requerirse explícitamente navegación desde una operación hacia la cuenta a la que pertenece, en una base de objetos tradicional no se caracterizaría a los objetos de Operación con un identificador de la cuenta y en el manejador de este trabajo sí. La introducción de este identificador posibilita una clusterización de las operaciones por la cuenta a la que afectan y en orden cronológico inverso, optimizándose así las recuperaciones de navegación.

La implementación de las asociaciones entre una clase de objetos maestros y una de objetos transaccionales con identificador mixto no se realiza mediante una colección de identificadores en los objetos maestros, que sería muy voluminosa, sino que, a la manera relacional, se maneja accediendo a los objetos transaccionales asociados por el identificador definido.

Como la clase Operación es transaccional no actualizable, los patrones de acceso característicos son para inserción o recuperación (no hay bajas ni actualización de objetos). Si la identificación de objetos en el sistema de información fuera por un número de operación independiente de la cuenta de pertenencia, una organización adecuada para la persistencia sería la secuencial indexada sin agrupar registros en bloques: todo nuevo registro se agrega al final del archivo, y el criterio de indexación de secuencias podría ser por umbrales horarios (por ejemplo se indexarían secuencias de operaciones por cada hora del día, por cada x cantidad de horas de cada día, o por cada día a partir de una hora determinada -por ejemplo la hora de clearing).

Pero en el caso de ejemplo, como la identificación depende de la cuenta a la que afecta, la agrupación o clusterización de operaciones se realiza según la cuenta, y las operaciones en cada cuenta se ordenan por cronología inversa; entonces la inserción de registros no puede realizarse siempre al final del archivo, sino al inicio de cada secuencia, siendo que cada secuencia es de operaciones en una cuenta en orden cronológico inverso. Por lo tanto la organización que se escoge es la secuencial indexada con registros organizados en bloques. Particularmente para esta combinación de clase o estereotipo de clase y definición de identificador se emplea árboles B# (árbol B+ con restricciones de contenido mínimo en nodos propias de los árboles B*) con los objetos organizados por el identificador del sistema de información. El fundamento de esta elección es que es común que las aplicaciones pidan iteradores sobre objetos por la clave de identificación del sistema de información, ya sea para listados o para búsquedas aproximadas, cuando se trata de objetos maestros, y para objetos transaccionales también, para realizar operaciones de totalización o síntesis (cortes de control); entonces se piensa que tener los objetos almacenados según el criterio de recuperación más probable importa una mejora del rendimiento general.

En cuanto a los identificadores automáticos o del manejador, los define automáticamente el manejador, y para los identificadores mixtos y externos los componentes externos son los

identificadores automáticos de las clases principales.

Para toda clase se crea por defecto un índice para el identificador automático con organización directa extensible y función de dispersión basada en los bits menos significativos del identificador de los objetos (resto de la división del identificador entre el tamaño de la tabla de dispersión).

En cuanto a la organización de los objetos secuencial indexada sin bloques, el índice se organiza con estructura B# y como referencias se usa el desplazamiento en bytes a cada objeto a partir del inicio del archivo.

Para otros índices también se usa árboles B#. Se admite la definición de índices que se clasifican según la multiplicidad de objetos referidos por cada clave, como **índices de identificación** (un objeto por clave) o **índices de clasificación** (muchos objetos por clave). Como referencias a objetos correspondientes a cada clave de recuperación (sea de identificación o de clasificación), se usa la clave de organización: el identificador del sistema de información.

En general, las **organizaciones de archivo** que se emplean para la persistencia de los objetos se detallan en el siguiente cuadro:

	Datos Maestros Actualizables	Datos Maestros No Actualizables (Dominios)	Transacciones Actualizables	Transacciones No Actualizables
Identificador Interno	B#	Sec. Ind. s/bloques	B#	Sec. Ind. s/bloques
Identificador Mixto o Externo	B#	B#	B#	B#

Para almacenar **jerarquías de objetos** se usa un único archivo, correspondiente a la clase raíz, sin importar que ésta sea abstracta o concreta. No se admite herencia múltiple, y si se desea identificar objetos de alguna subclase con un identificador propio se debe definir un índice de identificación, pero el identificador formal siempre es el de la clase raíz (no se admite identificadores en subclases). Los índices de subclases, tanto para identificadores automáticos como los definidos para recuperación se implementan como **índices selectivos**: indexan sólo los objetos de la subclase sobre la que están definidos.

RESULTADOS OBTENIDOS/ESPERADOS

En la implementación del manejador se usan técnicas relacionales, dada su probada eficacia, pero la mayor eficiencia del manejador en los casos planteados queda supeditada a la definición de la base, que si se efectúa sin usar identificadores mixtos o externos no se logra. En el caso de ejemplo, si las operaciones de identificar con un número de operación propio del sistema de información, la forma de implementar la asociación con la clase cuenta sería tener en cada objeto de cuenta una colección de identificadores automáticos de operación.

Todavía no se cuenta con resultados experimentales porque el manejador del proyecto se encuentra en una etapa de pruebas. Cuando el producto se encuentre estabilizado se estará en condiciones de comparar parámetros de rendimiento con productos similares.

Se está implementando paralelamente una versión del módulo de almacenamiento para el manejador de bases de objetos que implementa las asociaciones al estilo relacional, para comparar rendimientos con la propuesta de este trabajo.

FORMACION DE RECURSOS HUMANOS

En el proyecto de investigación y desarrollo de referencia participan docentes auxiliares y alumnos de grado de la carrera de Ingeniería en Informática en el marco de las materias Taller de Programación II (desarrollo de módulos experimentales del manejador de bases de objetos), Organización de Datos (implementación de técnicas de clusterización y almacenamiento de objetos), Trabajo Profesional, y Tesis; también participan alumnos de la carrera de Licenciatura en Análisis de Sistemas en el marco de la materia Aplicaciones Informáticas.

En el proyecto de referencia también participan investigadores del Laboratorio de Informática de Gestión.

BIBLIOGRAFIA

- [1] Malcolm Atkinson (University of Glasgow), François Bancilhon (Altaïr), David DeWitt (University of Wisconsin), Klaus Dittrich (University of Zurich), David Maier (Oregon Graduate Center) y Stanley Zdonik (Brown University), "The Object-Oriented Database System Manifiesto", Proc. of the First International Conference on Deductive and Object-Oriented Databases, Kyoto, Japón. 1990.
- [2] http://www.omg.org/technology/documents/corba_spec_catalog.htm#CCM
- [3] Jim Arlow, Ila Neustadt, "UML 2", Anaya Multimedia, 2006.
- [4] Carlo Batini, Stefano Ceri, Shamkant B. Navathe, "Diseño Conceptual de Bases de Datos: Un enfoque de entidades-interrelaciones", Addison-Wesley/Díaz de Santos, 1994.

Transformación de Procesos de Desarrollo de Software Tipo SPEM a Procesos Workflow. Una Propuesta de Caso de Estudio: SmallRUP

Fabio A. Zorzan¹, Daniel Riesco², Nora Szasz³

CONTEXTO

La línea de investigación presentada en este trabajo se desarrolla en el marco del proyecto “Ingeniería de Software: Automatización de Procesos de Desarrollo de Software”, presentado en la convocatoria 2009 para Proyectos y Programas de Investigación (PPI) de la secretaría de Ciencia y Técnica de la Universidad Nacional de Río Cuarto.

RESUMEN

Esta línea de investigación pretende aportar a la mejora de los procesos de desarrollo de software, en particular los especificados con el Software Process Engineering Metamodel versión 2 (SPEM). El caso de estudio propuesto consiste en transformar la especificación en SPEM de la metodología de desarrollo SmallRUP a una especificación de procesos que pueda ser utilizada como entrada de un motor workflow estándar. Para realizar esta transformación, se utilizará una transformación definida en el lenguaje Relations que forma parte de Query/Views/Transformations (QVT). Dicha transformación convierte una especificación en SPEM en una especificación de procesos en Business Process Modeling Notation (BPMN). La especificación BPMN, se transformará a su vez a una especificación en Business Process Execution Language (BPEL). Un vez obtenida la especificación de SmallRUP en BPEL, se la podrá utilizar como entrada a un motor workflow estándar, lo cual permitirá administrar la gestión de procesos de desarrollo que siguen la metodología SmallRUP con un motor workflow estándar.

Palabras clave— Workflow, SPEM, BPMN, QVT, Relations.

1 INTRODUCCIÓN

Un proceso de negocio es un conjunto de tareas lógicamente relacionadas, ejecutadas para obtener un resultado de negocio. Los procesos de negocio pueden ser controlados y administrados por un sistema basado en software. Los procesos de negocio automatizados de esta manera se denominan workflow. Esta automatización resulta en una importante potenciación de las virtudes de dicho proceso, obteniéndose mejoras en cuanto a rendimiento, eficiencia y productividad de la organización.

Dentro de la industria del desarrollo de software se encuentran los procesos de negocios tendientes a la construcción o generación de un producto (software) de calidad en un tiempo determinado [1]. En este marco, el proceso de negocio más importante involucra la metodología de desarrollo, utilizada para guiar la producción.

En [2] se presenta una transformación que convierte una especificación en SPEM versión 2 [3] (en adelante SPEM) en una especificación de procesos Workflow basado en el estándar BPMN [4],

¹ Fabio A. Zorzan pertenece al Departamento de Computación de la Universidad Nacional de Río Cuarto, Argentina; tel.: +54-358-4676529; e-mail: fzorzan@exa.unrc.edu.ar.

² Daniel Riesco pertenece al Departamento de Informática de la Universidad Nacional de San Luis, San Luis, Argentina; tel.: +54-2652-424027 int 251, e-mail: driesco@unsl.edu.ar

³ Nora Szasz, pertenece la Facultad de Ingeniería de la Universidad ORT Uruguay, tel.. +598-2-9021505; e-mail: nora.szasz@ort.edu.uy.

aceptado por la OMG. Esta traducción se obtuvo a través de una transformación definida mediante el lenguaje Relations que forma parte de QVT [5]. La transformación se definió entre el metamodelo SPEM y el metamodelo BPMN.

En este trabajo se presenta una línea de investigación que consiste en la elaboración de un caso de estudio para poner en práctica los resultados de las investigaciones antes mencionadas. La idea general es aportar a la optimización del proceso de producción de software transformando el proceso de desarrollo de software en un proceso de un workflow, con el objetivo de lograr la automatización de su gestión en todo o en parte. Para el caso de estudio se tomará la especificación de la metodología SmallRUP para el desarrollo de software [6]. La misma se especificará en SPEM, y luego se sucesivas transformaciones se convertirá en una especificación en Business Process Execution Language (BPEL) [7], un lenguaje estándar para la implementación de procesos workflow, la cual puede ser utilizada como entrada a un motor workflow estándar, y de esta manera, administrar la gestión de procesos de desarrollo que siguen la metodología SmallRUP con un motor workflow estándar

1.1 ELEMENTOS BÁSICOS

1.1.1 SPEM

Los procesos en el desarrollo de software pueden ser vistos como productos, ya que están constantemente cambiando y evolucionando. También deben ser administrados y configurados para adaptarlos a las organizaciones y a las nuevas necesidades del entorno. Las diferentes técnicas y procesos definieron sus propios estándares y terminologías, agregando de esta forma la necesidad de un estándar unificado en esta área.

Para especificar las actividades propuestas por un proceso de desarrollo particular y de esta forma proveer una solución a la necesidad antes planteada, la OMG definió un metamodelo para la Ingeniería de Procesos de Software (SPEM). SPEM ha evolucionado desde su definición original y a la fecha está disponible la versión 2.0 que hace una precisa separación entre la definición de los procesos y la ejecución de los mismos, la cual no estaba claramente definida en versiones anteriores.

Para la definición de nuevos lenguajes, la OMG define una arquitectura basada en cuatro niveles de abstracción que van a permitir distinguir entre los distintos niveles conceptuales que intervienen en el modelado de un sistema. A esos niveles se les denomina M0, M1, M2 y M3. SPEM está dentro del nivel M2 y describe un metamodelo genérico para la descripción de procesos software concretos que está basado en MOF [8] y extiende al metamodelo UML.

1.1.2 WORKFLOW

Un workflow se define como la automatización total o parcial de un proceso de negocio, durante la cual documentos, información o tareas son intercambiadas entre los participantes conforme a un conjunto de reglas procedimentales preestablecidas [9]. Un workflow comprende un número de pasos lógicos, conocidos como actividades. Una actividad puede involucrar la interacción manual o automática con el usuario. Un motor workflow es un sistema de software que controla la ejecución de las actividades definidas en el workflow. La WfMC [10] ha definido un Modelo de Referencia Workflow (Workflow Reference Model). Este modelo define 5 interfaces para la interoperabilidad de diferentes productos con un motor workflow.

En nuestra investigación interesa la interfaz 1, que especifica el formato de intercambio común para soportar la transferencia de definiciones de procesos entre productos diferentes. Los lenguajes de definición de procesos XPDL[11] (definido por la WfMC) o BPEL [7] (adoptado por OASIS)

implementan la interfase 1. Existen varios motores workflow que implementan estos lenguajes: OFBiz Workflow Engine [12] o Open Business Engine [13] soportan XPD L, mientras que WebSphere Process Server [14] y BPEL Process Manager [15] implementan BPEL.

A la hora de modelar un proceso de negocio es importante poder utilizar una herramienta independiente de la implementación para poder utilizar la especificación del proceso de negocio con diferentes plataformas. BPMN es una herramienta de estas características que es muy utilizada.

1.1.3 BPMN

La OMG junto con la Business Process Modeling Initiative (BPMI) han desarrollado una notación, denominada BPMN, para el modelado de procesos de negocio. BPMN define una notación para la definición de procesos de negocio, lo que es una plataforma independiente con respecto a definiciones específicas de procesos de negocio (como por ejemplo XPD L o BPEL). Esta notación define una representación abstracta para la especificación de procesos de negocio que se ejecutan dentro de una empresa. Partiendo de un modelo BPMN se puede obtener la definición de un proceso de negocio en un lenguaje específico mediante una transformación. En [4] está definida la correspondencia de BPMN a BPEL. Los elementos de la notación están especificados en el metamodelo BPMN [16]. Este metamodelo está definido en el nivel M2 de la OMG y está basado en MOF.

1.1.4 QVT

El planteamiento QVT se basa principalmente en: la definición de un lenguaje para las consultas (Queries) sobre los modelos MOF, la búsqueda de un estándar para generar vistas (Views) que revelen aspectos específicos de los sistemas modelados, y finalmente, la definición de un lenguaje para la descripción de transformaciones (Transformations) de modelos MOF.

En este trabajo se utiliza el componente de QVT que tiene como objetivo definir transformaciones. Estas transformaciones describen relaciones entre un metamodelo fuente F y un metamodelo objetivo O, ambos especificados en MOF. La transformación definida se utiliza para obtener un modelo objetivo que es una instancia del metamodelo O a partir de un modelo fuente que es una instancia del metamodelo F.

La especificación de QVT que se utiliza tiene una naturaleza híbrida declarativa/imperativa y se denomina lenguaje Relations. Este lenguaje permite realizar “pattern matching” de objetos complejos y definir “templates” de creación de objetos. En la actualidad hay herramientas que implementan este lenguaje, como por ejemplo MOMENT [17] y MediniQVT [18].

2 LINEAS DE INVESTIGACION Y DESARROLLO

2.1 CASO DE ESTUDIO PROPUESTO

Como se mencionó en la introducción, el caso de estudio utilizará el Proceso Unificado Rational (Rational Unified Process - RUP), y en particular una instanciación para pequeños proyectos definida en [6], denominada SmallRUP. La elección del RUP se debió a sus principales características: impone buenas prácticas en el desarrollo de software moderno para una amplia gama de proyectos y organizaciones, está embebido en técnicas orientadas a objetos, usa UML como notación principal, permite a las organizaciones del software ajustar el proceso a su necesidad específica y cubre diferentes dominios particulares. La propuesta consiste en aplicar sucesivas transformaciones a la especificación en SPEM de SmallRUP, como se muestra en la figura 1. Estas transformaciones tienen por objetivo obtener una especificación de procesos workflow que sirva de entrada a un motor workflow estándar. La primera transformación a aplicar se define en Relations de QVT. Esta transformación convierte la

especificación en SPEM de SmallRUP en una especificación de procesos BPMN. Una vez obtenida la especificación BPMN, se aplica la segunda transformación. Esta transformación consistirá en convertir la especificación BPMN de SmallRUP en una especificación en el lenguaje de implementación workflow BPEL.

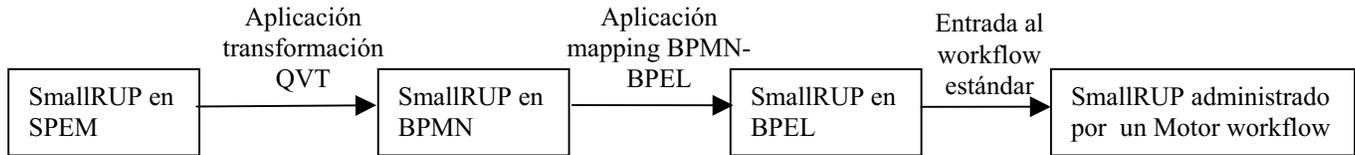


Figura 1: Secuencia de transformaciones.

2.1.1 ESTADO DE AVANCE

Hasta el momento se ha avanzado principalmente en la primera parte de este caso de estudio, esto es, en la transformación que convierte la especificación SPEM en una especificación BPMN.

Para llevar a cabo esta transformación se utilizó la herramienta MediniQVT. Teniendo en cuenta que MediniQVT utiliza a Eclipse Modeling Framework (EMF) [19] para representar los modelos/metamodelos involucrados en las transformaciones, se especificaron los metamodelos SPEM y BPMN en EMF, y luego también el modelo SmallRUP basado en el metamodelo SPEM. Una vez definida la especificación de SmallRUP en SPEM, se le aplicó la transformación en Relations, obteniéndose la especificación SmallRUP basada en el metamodelo BPMN en formato EMF. Actualmente se está verificando la especificación BPMN para comprobar si cumple con los requerimientos de la metodología de desarrollo que debe soportar.

3 RESULTADOS OBTENIDOS/ESPERADOS

Esta propuesta tiene como objetivo hacer una contribución a la mejora de los procesos de desarrollo de software, viendo a éste como un proceso de negocio particular. La elaboración y culminación exitosa de este caso de estudio nos mostraría que es posible la utilización de un motor workflow estándar para la administración, automática o semi-automática, de la gestión de procesos de desarrollo de software especificados bajo el estándar SPEM.

El beneficio de esta transformación se reflejará también en el dinamismo de los cambios en los procesos de desarrollo de software: cualquier cambio en la especificación del proceso podrá ser propagado a la especificación workflow de dicho proceso, y adaptado rápidamente la especificación de entrada al workflow.

Los principales objetivos de esta línea de investigación son:

- Formalizar la automatización del desarrollo de software utilizando la tecnología de flujo de trabajo con teorías como lenguajes formales de especificación y otros.
- Encarar trabajos conjuntos con universidades nacionales y con centros internacionales de excelencia como se viene realizando desde el año 2006, cuyos resultados fueron publicados en conferencias nacionales e internacionales.
- Servir como marco para dar un fuerte respaldo a la elaboración de trabajos finales de grado y tesis de posgrado.

4 FORMACION DE RECURSOS HUMANOS

Los estudios realizados en esta línea de investigación sirven como marco para la elaboración de una tesis de Maestría que está actualmente en curso. Los temas abordados en esta línea de investigación brindan un fuerte aporte al proceso de perfeccionamiento continuo de los autores de este trabajo, que se desempeñan como docentes de carreras de computación en Universidades Nacionales como del exterior.

5 BIBLIOGRAFIA

- [1] N. Debnath, D. Riesco, G. Montejano, et al, “Supporting the SPEM with a UML Extended Workflow Metamodel”, ACS/IEEE International Conference on Computer Systems and Applications (AICCSA'06). Conference to be held in Dubai/Sharjah during March 8-11, 2006, www.ieee.org .
- [2] N. Debnath, F. A. Zorzan, G. Montejano and D. Riesco, “Transformation of BPMN Subprocesses Based in SPEM Using QVT”, 2007 IEEE INTERNATIONAL CONFERENCE on ELECTRO/INFORMATION TECHNOLOGY, May 17-20, 2007, Marriott O Hare, Chicago, IL, USA. <http://www.eit-conference.org/eit2007/>
- [3] Object Management Group, “ Software & Systems Process Engineering Metamodel Specification, v2”; Proposed Available Specification ptc/2007-08-07 of the Object Management Group, Inc; <http://www.omg.org/docs/formal/07-11-01.pdf> , último acceso Marzo 2008.
- [4] Object Management Group “Business Process Modeling Notation (BPMN) Specification”. Final Adopted Specification dtc/06-02-01, http://www.bpmn.org/Documents/OMG_Final_Adopted_BPMN_1-0_Spec_06-02-01.pdf, último acceso Octubre 2008.
- [5] Object Management Group, “Meta Object Facility (MOF) 2.0 Query/View/Transformation Specification” Final Adopted Specification ptc/05-11-01, <http://www.omg.org/docs/ptc/05-11-01.pdf>, último acceso Febrero 2009.
- [6] Gary Pollice “Using the RUP for small projects: Expanding upon Extreme Programming”, A Rational Software White Paper–04/08/15, <ftp://ftp.software.ibm.com/software/rational/web/whitepapers/2003/tp183.pdf>, último acceso Diciembre 2008.
- [7] BEA, IBM, Microsoft, SAP and Siebel, “Business Process Execution Language for Web Services Version 1.1”, S. Thatte, et al., May 2003, <ftp://www6.software.ibm.com/software/developer/library/ws-bpel.pdf>, último acceso, Febrero 2009.
- [8] Object Management Group “Meta Object Facility (MOF) Core Specification” OMG Available Specification. Version 2.0. formal/06-01-01, <http://www.omg.org/docs/formal/06-01-01.pdf>, último acceso, Noviembre 2008.
- [9] Rob Allen, Open Image Systems Inc., United Kingdom Chair, WfMC External Relations Committee; “The Workflow Handbook 2001”; Workflow Management Coalition; October 2001.
- [10] Workflow Management Coalition; “The Workflow Reference Modelo”. The Workflow Management Coalition Specification; WfMC-TC-1003 Version 1.1 Issue; Enero de 1995.
- [11] Workflow Management Coalition, Workflow Standard – Workflow Process Definition Interface -XML Process Definition Language, Workflow Management Coalition , WfMC-TC-1025, 2002, http://www.wfmc.org/standards/docs/TC-025_10_xpdl_102502.pdf, último acceso Diciembre 2008.
- [12] OFBiz Workflow Engine, <http://incubator.apache.org/ofbiz/docs/workflow.html>, último acceso Febrero 2009.
- [13] Open Business Engine, <http://obe.sourceforge.net/> , último acceso Enero 2008.
- [14] IBM, “WebSphere Process Server”, <http://www-306.ibm.com/software/integration/wps/>, último acceso, Diciembre 2008.
- [15] Oracle, “BPEL Process Manager” <http://www.oracle.com/technology/products/ias/bpel/index.html>, último acceso, Febrero 2007.
- [16] Object Management Group, BPMN Documents “BPMNModel UML Documentation”. Draft Specification, <http://www.bpmn.org/Documents/BPMNMetaModel.zip> , último acceso, Octubre 2006.
- [17] Pascual Queralt, Luis Hoyos, Artur Boronat, José Á. Carsí e Isidro Ramos; “Un motor de transformación de modelos con soporte para el lenguaje qvt relations”, Desarrollo de Software Dirigido por Modelos - DSDM'06 (Junto a JISBD'06). October 2006. Sitges, Spain. – 2006.
- [18] ikv++: medini QVT. <http://www.ikv.de/>, último acceso, Febrero 2009.
- [19] “Eclipse Modeling Framework”, URL: <http://www.eclipse.org/modeling/emf/>. último acceso, Marzo 2009.

Un Cálculo de Patrones Arquitectónicos

Alejandro Sánchez*
asanchez@unsl.edu.ar

Luis Soares Barbosa†
lsb@di.uminho.pt

Daniel Riesco*
driesco@unsl.edu.ar

CONTEXTO

La presente línea de investigación se encuentra enmarcada en el proyecto “Ingeniería de Software, Conceptos, Métodos y Herramientas en un contexto de Evolución de la Ingeniería del Software” acreditado por incentivos, de la Universidad Nacional de San Luis. La línea se desarrolla a partir de la colaboración surgida por el proyecto LerNet del programa ALFA con la Universidad de Minho (Portugal).

RESUMEN

En esta línea de investigación proponemos desarrollar un framework semántico y un cálculo que permita describir, transformar y razonar sobre patrones arquitectónicos de software. El trabajo apunta a responder a las necesidades que emergen a partir del éxito de la orientación a servicios, en donde se necesita tratar con arquitecturas dinámicamente reconfigurables y auto-adaptables. Se basa en un enfoque desarrollado en la Universidad de Minho para el cálculo de componentes y se aplican avances del áreas de modelos y lenguajes de coordinación.

Palabras clave: Patrones Arquitectónicos de Software, Arquitectura de Software, Cálculo, Coalgebra

1 INTRODUCCION

La continua evolución hacia sistemas de computación extremadamente grandes, altamente dinámicos y heterogéneos requiere enfoques innovadores para dominar su complejidad. Los servicios no solo están transformando la Web, de estar centrada en documentos a ser una infraestructura viva, sino que también desafían nuestro entendimiento de como desarrollar aplicaciones, e incluso, de la naturaleza misma del software (visto como un servicio a ser contratado en lugar de un producto a adquirir). El impacto de este cambio, tanto en la economía mundial como en la vida cotidiana, esta tan solo comenzando a apreciarse.

Complejos sistemas de software se construyen conectando servicios que interactúan intercambiando datos, llevando a cabo computaciones, y modificando su ambiente. Los servicios son entidades dinámicas, que se ejecutan en

diferentes plataformas, a menudo pertenecen a distintas organizaciones, interactúan a través de interfaces públicas, y típicamente permanecen mínimamente acoplados, si no completamente ignorantes unos de los otros.

El correcto diseño de tales sistemas es difícil por que su complejidad va más allá del alcance actual de los métodos formales. Dificultades adicionales emergen con servicios provistos por terceros, frecuentemente subespecificados o que no cumplen con sus especificaciones. Es más, siendo el cambio la norma en lugar de la excepción, la reconfiguración dinámica y la auto-adaptabilidad, es decir, la capacidad de un sistema de ajustarse a sí mismo durante su ejecución en respuesta a su percepción del contexto, se volverán un importante tema en el futuro cercano.

1.1 Antecedentes

El problema se puede enmarcar en dos áreas de investigación dentro de la Ingeniería del Software: *arquitecturas de software*, y *lenguajes y modelos de coordinación*.

La primera emerge como un disciplina [26, 19, 2, 18, 12] en la Ingeniería de Software, de la necesidad de considerar de manera explícita en el desarrollo de sistemas, crecientemente grandes y complejos, los efectos, problemas y oportunidades de la estructura, organización y comportamiento emergente de un sistema completo. En una definición amplia, la arquitectura de un sistema describe su organización fundamental, la cual ilumina las decisiones de diseño de alto nivel:

- ¿cómo esta compuesto y de que partes interactuantes?
- ¿cuales son las interacciones y los patrones de comunicación presentes?
- ¿cuales son las propiedades claves de partes en las que el sistema completo descansa y/o impone?

En su papel de modelo, la arquitectura actúa como una abstracción de un sistema y suprime detalles de elementos que no afectan como utilizan, son utilizados por, se relacionan o interactúan con otros elementos. Entonces, se enfoca en los elementos e interfaces estructurales por los cuales un sistema es compuesto, sus comportamientos separados y combinados especificados como colaboraciones entre dichos elementos, y finalmente la composición de estos elementos

*Universidad Nacional de San Luis, Facultad de Ciencias Físico-Matemáticas y Naturales, Departamento de Informática, Tel: +54 (0) 2652 424027 / Fax: +54 (0) 2652 430059, San Luis, Argentina

†Universidade do Minho, Escola de Engenharia, Departamento de Informática, Centro de Ciências e Tecnologias de Computação, Tel: +351 253604463 / Fax: +351 253604471, Braga, Portugal

estructurales y comportamentales en sistemas más grandes. De acuerdo con la norma ANSI/IEEE Std 1471-2000, la cual es parte de un procesos de estandarización en marcha, la arquitectura describe la organización fundamental del sistema, materializado en sus componentes, sus relaciones entre ellos y el ambiente, y los principios gobernando su diseño y evolución.

Por el otro lado, el paradigma de coordinación [21, 5], el cual promueve una estricta separación entre la computación efectiva y su control, apareció como una solución al problema de manejar la interacción entre actividades concurrentes en un sistema. Emergió de la necesidad de explotar el pleno potencial de masivos sistemas en paralelo, lo cual requiere modelos capaces de tratar, de una manera explícita, la concurrencia de la cooperación entre un gran número de componentes poco acoplados, autónomos y heterogéneos. Los modelos de coordinación [20, 25, 6] hacen una clara distinción entre estos componentes y sus interacciones, y se enfocan en el comportamiento *emergente* combinado de estos. Tradicionalmente, los lenguajes y modelos de coordinación han evolucionado alrededor de la noción de un espacio de datos compartido — una abstracción de memoria para el intercambio de datos, accesible a todos los procesos cooperando para alcanzar una meta. El primer lenguaje de coordinación que introduce tal noción fue Linda [1]; varios modelos relacionados evolucionaron más tarde sobre nociones similares [21, 13]. El modelo subyacente era *dirigido por datos*, en el sentido que los procesos podían examinar la naturaleza de los datos intercambiados y actuar de acuerdo a estos. Una familia alternativa de modelos, *dirigida por eventos* o *dirigida por control*, se ajusta mejor a sistemas cuyos componentes interactúan entre ellos pasándose y recibiendo eventos, la presencia de los cuales dispara actividad. Un modelo pionero en esta familia es MANIFOLD [8], el cual implementa el modelo IWIM [4]. Al contrario que con la familia dirigida por datos, en donde los coordinadores manejaban directamente los valores de datos, en estos modelos, los procesos se ven como cajas negras que se comunican con su ambiente a través de interfaces claramente definidas (frecuentemente llamadas puertos de entrada y salida). Durante los últimos 15 años, la aparición masiva de sistemas concurrentes y heterogéneos, y el crecimiento de la complejidad de los protocolos de interacción, ha llevado a la coordinación a un lugar central en el desarrollo de software. Este avance contribuyó a ampliar su alcance de aplicación y significó el desarrollo de un número de modelos, lenguajes y semánticas específicos.

Los modelos de coordinación y las descripciones arquitectónicas nacieron dentro de contextos, preocupaciones y dominios de aplicación diferentes. Sin embargo, su foco es similar y recientes tendencias en la industria de software enfatizan la relevancia de los principios básicos subyacentes. Recordar, por ejemplo, los desafíos vinculados al cambio desde el paradigma *programming-in-the-large* dos décadas atrás, al reciente paradigma de *programming-in-the-world*, donde no solo se tiene que dominar la complejidad de construir y desplegar grandes aplicaciones en tiempo y dentro del presupuesto, sino que también se tiene que administrar

una estructura abierta de componentes autónomos, posiblemente distribuidos y altamente heterogéneos. O también considere el cambio de entender al *software como producto* a *software como servicio* [17], enfatizando su estructura abierta, dinámica, reconfigurable y evolutiva. Términos tales como *orquestración* y *coreografía*, y los intensos esfuerzos de investigación asociados (ver [15, 3, 14, 29], entre otros), acentúan la relevancia de los principales temas de investigación en arquitecturas y coordinación para la Ingeniería de Software. En cierto modo, una definición temprana de coordinación que enfatiza su meta de encontrar soluciones al problema de manejar la interacción entre programas concurrentes [5], puede ser tomada como un desafío clave a este dominio de la Ingeniería.

Tanto la *arquitectura de software* como los *modelos de coordinación* tratan la interacción de componentes, abstrayendo los detalles de la computación y enfocándose en la naturaleza y forma de las interacciones. En consecuencia, la sincronización, comunicación, reconfiguración, inicio y fin de las computaciones son temas que les concierne.

También se debe remarcar que, a pesar del notable progreso en la representación y uso de las arquitecturas de software, la especificación del diseño arquitectónico se mantiene mayormente informal en la actualidad. Típicamente, dichos diseños se apoyan en notaciones con semántica pobre, y frecuentemente limitada a expresar solo las propiedades estructurales más básicas. Por el otro lado, los lenguajes y modelos de coordinación recientes — tales como REO [6, 7] y ORC [22, 24] — presentan un mayor grado de formalidad, lo cual acentúa el caso para una *vista dirigida por coordinación* de la arquitectura de sistemas.

1.2 Objetivos

Ha llegado el momento de profundizar en el desarrollo de una *teoría de patrones arquitectónicos*, comprendiendo una semántica y un cálculo, construidos en base a la experiencia ganada de la investigación en coordinación. Es más, esta teoría parece fundamental a los efectos de proveer bases sólidas para el diseño orientado a servicios. En este contexto, el trabajo apunta a tratar las siguientes preguntas transversales a los fundamentos de la arquitectura de software:

- ¿Cómo especificar, transformar y razonar sobre patrones arquitectónicos, y calcular el comportamiento emergente del sistema?
- ¿Cómo aplicar tal cálculo de patrones arquitectónicos para diseñar, analizar y transformar redes evolutivas de componentes dinámicamente reconfigurables y arquitecturas auto-adaptables?

1.3 Relevancia

Estos temas, y en particular, el foco en arquitecturas dinámicas y auto-adaptables, son relevantes a un amplio rango de sistemas. Este rango va desde comercio electrónico a sistemas móviles embebidos, operados con un mínimo de supervisión humana, en el contexto en el cual

la distinción entre 'desarrollo', 'despliegue' y 'mantenimiento' tiende a ser difusa. A pesar de ser una realidad tecnológica, la reconfiguración en tiempo de ejecución es difícil de modelar, analizar y predecir. A su vez, las menos comunes arquitecturas, capaces de monitorear y adaptarse a sí mismas a fallas (por ejemplo, pérdida de conexiones o fallas en servicios), a recursos variables (como disponibilidad de ancho de banda) y a cambios no predecibles en el contexto, crecerán en relevancia en el futuro próximo.

Estos son grandes desafíos de investigación para los cuales la teoría de patrones arquitectónicos a desarrollarse en este trabajo puede proveer, aunque sea de manera parcial, respuestas relevantes. En realidad, al ser un área emergente, la computación orientada a servicios requiere de claras definiciones de modelos y métodos para la especificación, orquestación y desarrollo de servicios, así como una técnica específica para su análisis y transformación.

2 LINEAS DE INVESTIGACION y DESARROLLO

Este trabajo apunta a desarrollar una nueva y rigurosa disciplina de patrones arquitectónicos. Esta consistirá de un cálculo para tales patrones y sus transformaciones. Irá mas allá de las conocidas nociones *ad hoc* de estilos arquitectónicos (tal como se presentan en las referencias clásicas del tema), y buscará formulaciones independientes de cualquier tecnología. Describimos el enfoque a seguir a través de los siguientes puntos:

- El cálculo adoptará un punto de vista exógeno para la coordinación y se basará en la existencia de un conjunto pequeño de combinadores genéricos. Tal como en los lenguajes de coordinación modernos, el punto de vista exógeno implica un desacoplamiento estricto entre servicios, para soportar un mínimo de dependencias entre componentes. En contraste con los enfoques menos estructurados de tales lenguajes, el conjunto de combinadores genéricos permitirán el diseño de conectores de una manera estructurada. En consecuencia, preveemos dos versiones diferentes, pero relacionadas, para este cálculo:
 - Sobre lenguajes de coordinación modernos (tales como Reo y Orc): se dirige el trabajo a una amplia comunidad de usuarios y testers que estos lenguajes tienen.
 - Sobre el llamado enfoque Minho para el cálculo de componentes: se aplica este enfoque puramente composicional equipado con las nociones de bisimulación y refinamiento [23]. Diferentes trabajos siguiendo este enfoque han mostrado su aplicación para la especificación de conectores [11], la componentización de una especificación [16], el cálculo de invariantes [9], y la especificación de componentes parciales (componentes que fallan o mueren, exhibiendo un comportamiento más efímero que lo esper-

ado) y la extensión de dichos componentes con un mecanismo de ciclos *try-again* [10].

- Proponemos *dialgebras* [27] como un modelo básico para el cálculo de patrones arquitectónicos. Los servicios y sus configuraciones aparecen naturalmente equipadas con operaciones 'constructoras' y 'destructoras' (estas últimas también llamadas observadores), las cuales requieren nociones de equivalencia semántica y refinamiento tanto estructurales (algebraicas) como comportamentales (coalgebraicas). Las dialgebras generalizan álgebras y coalgebras, pero su estudio esta todavía en su infancia con respecto a los métodos algebraicos/coalgebraicos.
- El cálculo debe extenderse para tratar con propiedades arquitectónicas no funcionales (relacionadas a calidad de servicio).
- En el nivel experimental, este trabajo apunta a caracterizar y clasificar un numero de patrones arquitectónicos, estudiando su integración, uso y relevancia. En lugar de ser postulados, su identificación y clasificación procederá de la inspección de casos reales. Para esto, extenderemos herramientas previamente desarrolladas en la Universidad de Minho [28], con el objetivo de extraer especificaciones de coordinación de código ejecutable. También aprovecharemos un enorme repositorio de datos arquitectónicos mantenidos por una empresa Holandesa, especializada en análisis de código, la cual ya ha accedido en esta forma de colaboración.

3 RESULTADOS ESPERADOS/OBTENIDOS

Se espera que el trabajo en esta línea ofrezca resultados en el nivel teórico y de aplicación. En el nivel teórico se espera obtener un framework semántico y un cálculo de patrones arquitectónicos, aplicables a arquitecturas dinámicamente reconfigurables y auto-adaptables. En el nivel de aplicación se prevee extender herramientas para soportar la extracción de datos arquitectónicos del código ejecutable de sistemas reales.

4 FORMACION DE RECURSOS HUMANOS

La línea de investigación es llevada adelante por investigadores de la Universidad de Minho y la Universidad de San Luis. La misma será el contexto en el que se desarrollarán tesis de maestría y doctorado

BIBLIOGRAFIA

- [1] S. Ahuja, N. Carriero, and D. Gelernter. Linda and friends. *IEEE Computer*, 19(8):26–34, 1986.

- [2] R. Allen and D. Garlan. A formal basis for architectural connection. *ACM TOSEM*, 6(3):213–249, 1997.
- [3] L. F. Andrade and J. L. Fiadeiro. Composition contracts for service interaction. *Journal of Universal Computer Science*, 10(4):751–761, 2004.
- [4] F. Arbab. The iwim model for coordination of concurrent activities. In P. Ciancarini and C. Hankin, editors, *Proc. Coordination Languages and Models, First Inter. Conf., COORDINATION '96, Cesena, Italy, April 15-17*, volume 1061, pages 34–56. Springer Lect. Notes Comp. Sci. (1061), 1996.
- [5] F. Arbab. What do you mean, coordination. In *Bulletin of the Dutch Association for Theoretical Computer Science (NVTI)*, 1998.
- [6] F. Arbab. Abstract behaviour types: a foundation model for components and their composition. In F. S. de Boer, M. Bonsangue, S. Graf, and W.-P. de Roever, editors, *Proc. First International Symposium on Formal Methods for Components and Objects (FMCO'02)*, pages 33–70. Springer Lect. Notes Comp. Sci. (2852), 2003.
- [7] F. Arbab. Reo: a channel-based coordination model for component composition. *Mathematical Structures in Comp. Sci.*, 14(3):329–366, 2004.
- [8] F. Arbab, I. Herman, and P. Spilling. An overview of manifold and its implementation. *Concurrency - Practice and Experience*, 5(1):23–70, 1993.
- [9] L. S. Barbosa and J. N. Oliveira. Transposing partial components: an exercise on coalgebraic refinement. *Theor. Comp. Sci.*, 365(1-2):2–22, 2006.
- [10] L. S. Barbosa, J. N. Oliveira, and A. M. Silva. Calculating invariants as coreflexive bisimulations. In J. Meseguer and G. Rosu, editors, *Algebraic Methodology and Software Technology, 12th International Conference, AMAST 2008, Urbana, IL, USA, July 28-31, 2008, Proceedings*, pages 83–99. Springer Lect. Notes Comp. Sci. (5140), 2008.
- [11] M. Barbosa and L. Barbosa. Specifying software connectors. In K. Araki and Z. Liu, editors, *Proc. First International Colloquium on Theoretical Aspects of Computing (ICTAC'04), Guiyang, China*, pages 53–68. Springer Lect. Notes Comp. Sci. (3407), 2004.
- [12] L. Bass, P. Clements, and R. Kazman. *Software Architecture in Practice (2nd ed.)*. Addison-Wesley, 2003.
- [13] R. Bjornson, N. Carriero, and D. Gelernter. From weaving threads to untangling the web: A view of coordination from linda's perspective. In D. Garlan and D. L. Metayer, editors, *Proc. of Second Inter. Conf. on Coordination Languages and Models, COORDINATION '97, Berlin, Germany*, pages 1–17. Springer Lect. Notes Comp. Sci. (1282), 1997.
- [14] A. Brogi, C. Canal, E. Pimentel, and A. Vallecillo. Formalizing web services choreographies. In *Proc. First Inter. Workshop on Web Services and Formal Methods*, volume 105, pages 73–94, Pisa, Italy, 2004.
- [15] N. Busi, R. Gorrieri, C. Guidi, R. Luchi, and G. Zavattaro. Choreography and orchestration: A synergic approach for systems design. In B. Benatallah, F. Casati, and P. Traverso, editors, *Proc. ICSOC 2005 Thrid Inter. Conf. on Service-Oriented Computing*, pages 228–240, 2005.
- [16] A. Cruz, L. Barbosa, and J. Oliveira. From algebras to objects: Generation and composition. *Journal of Universal Computer Science*, 11(10):1580–1612, 2005.
- [17] J. L. Fiadeiro. Software services: scientific challenge or industrial hype? In K. Araki and Z. Liu, editors, *Proc. First International Colloquium on Theoretical Aspects of Computing (ICTAC'04), Guiyang, China*, pages 1–13. Springer Lect. Notes Comp. Sci. (3407), 2004.
- [18] D. Garlan. Formal modeling and analysis of software architecture: Components, connectors and events. In M. Bernardo and P. Inverardi, editors, *Third International Summer School on Formal Methods for the Design of Computer, Communication and Software Systems: Software Architectures (SFM 2003)*. Springer Lect. Notes Comp. Sci., Tutorial, (2004), Bertinoro, Italy, September 2003.
- [19] D. Garlan and M. Shaw. An introduction to software architecture. In V. Ambriola and G. Tortora, editors, *Advances in Software Engineering and Knowledge Engineering (volume I)*. World Scientific Publishing Co., 1993.
- [20] D. Gelernter and N. Carrier. Coordination languages and their significance. *Communication of the ACM*, 2(35):97–107, February 1992.
- [21] D. L. M. Jean-Marc Andreoli, Chris Hankin. *Coordination Programming: Mechanisms, Models, and Semantics*. Imperial College Press, 1996.
- [22] D. Kitchin, W. R. Cook, and J. Misra. A language for task orchestration and its semantic properties. In C. Baier and H. Hermanns, editors, *Proc. 17th Inter. Conf. Concurrency Theory, CONCUR 2006, Bonn, Germany, August 27-30*, pages 477–491. Springer Lect. Notes Comp. Sci. (4137), 2006.
- [23] S. Meng and L. S. Barbosa. Components as coalgebras: The refinement dimension. *Theor. Comp. Sci.*, 351:276–294, 2005.
- [24] J. Misra and W. R. Cook. Computation orchestration: A basis for wide-area computing. *Software and System Modeling*, 6(1):83–110, 2007.

- [25] G. Papadopoulos and F. Arbab. Coordination models and languages. In *Advances in Computers — The Engineering of Large Systems*, volume 46, pages 329–400. 1998.
- [26] D. E. Perry and A. L. Wolf. Foundations for the study of software architecture. *ACM SIGSOFT Software Engineering Notes*, 17(4):40–52, 1992.
- [27] E. Poll and J. Zwanenburg. From algebras and coalgebras to dialgebras. In H. Reichel, editor, *Coalgebraic Methods in Computer Science (CMCS'2001)*, number 44 in ENTCS. Elsevier, 2001.
- [28] N. F. Rodrigues and L. S. Barbosa. Coordinspector: a tool for extracting coordination data from legacy code. In *SCAM '08: Proc. of the Eighth IEEE Inter. Working Conference on Source Code Analysis and Manipulation*. IEEE Computer Society, 2008. to appear.
- [29] Q. Zongyan, Z. Xiangpeng, C. Chao, and Y. Hongli. Towards the theoretical foundation of choreography. In P. Patel-Schneider and P. Shenoy, editors, *Proceedings of the 16th Int Conf. on World Wide Web*, pages 973–982. ACM, 2007.

Transformaciones Genéricas para la Implementación de Web Services en Diferentes Plataformas

Ariel Arsaute, Marcela Daniele, Mariana Frutos, Paola Martellotto, Fabio Zorzan

Departamento de Computación

Facultad de Ciencias Exactas, Físico-Químicas y Naturales - Universidad Nacional de Río Cuarto

CONTEXTO

La línea de investigación presentada en este trabajo se desarrolla en el marco del proyecto “Definición de Transformaciones Genéricas para la implementación de Web Services en diferentes plataformas”, presentado en el Programa de Apoyo a Grupos de Reciente Formación en la Convocatoria 2008 del Ministerio de Ciencia y Tecnología de la provincia de Córdoba.

RESUMEN

Esta línea de investigación pretende aportar al desarrollo de aplicaciones basadas en Web Services, y propone la definición de diferentes transformaciones que reciban como entrada un modelo genérico para Web Services, y retornen modelos que representan a diferentes plataformas/lenguajes de implementación para Web Services. Dichas transformaciones se enmarcan dentro de la metodología de desarrollo de Arquitecturas Dirigidas por Modelos (Model Driven Architecture-MDA) el cual sostiene que si el desarrollo está guiado por los modelos de software, se obtienen beneficios importantes en aspectos fundamentales como la productividad, portabilidad, interoperabilidad y mantenimiento. La definición de la transformación se realiza por medio del lenguaje Relations que forma parte de Query/Views/Transformations (QVT) o con Reglas de Transformación de Grafos, para los cuales existen herramientas de software que permiten, tanto la definición, como la ejecución de transformaciones.

Palabras Clave: Web Services, QVT, Reglas de Transformación de Grafos, MDA

1. INTRODUCCIÓN

Desde los inicios de la era de la información la reutilización ha supuesto una práctica habitual para el desarrollo de aplicaciones de software. En este momento, la reutilización de software continúa siendo un aspecto esencial en los sistemas de información, a través del uso de componentes de software.

La tendencia actual en el desarrollo de componentes es reutilizarlos eficientemente en diferentes proyectos, tanto como sea posible. Esta reutilización está limitada a un lenguaje de programación o a una plataforma en particular.

Con la aparición de los Web Services (Servicios Web) [1], esta limitación fue superada, proporcionando un modelo diferente para el desarrollo de aplicaciones, ofreciendo la capacidad de acceder a servicios heterogéneos de forma unificada e interoperable a través de Internet. Los Web Services son servicios autónomos e independientes que se ofrecen mediante la web. Principalmente, permiten que las aplicaciones sean más modulares y desacopladas, facilitando su reutilización en distintas plataformas o lenguajes de programación.

En la actualidad, la construcción de software se enfrenta a continuos cambios en las tecnologías de implementación, lo que implica realizar esfuerzos importantes tanto en el diseño para integrar las diferentes tecnologías que incorpora, como en el mantenimiento para adaptar la aplicación a los

cambios en los requisitos y en las tecnologías de implementación. Es difícil satisfacer requisitos de escalabilidad, seguridad y eficiencia. La idea clave que subyace a las arquitecturas dirigidas por modelos MDA [2] es que, si el desarrollo está guiado por los modelos de software, se obtendrán beneficios importantes en aspectos fundamentales como son la productividad, la portabilidad, la interoperabilidad y el mantenimiento.

La necesidad de compartir información entre sistemas implementados en distintos lenguajes de programación, las ventajas que provee la técnica MDA y los importantes beneficios aportados por la incorporación de Web Services a las aplicaciones, conduce a pensar en la necesidad de contar con una solución genérica que aproveche las fortalezas de estas tecnologías. Esta línea de investigación tiene como objetivo proveer una solución genérica y eficiente para el tipo de problemas planteados, permitiendo transformaciones, ya sea empleando QVT [3] o Reglas de Transformación de Grafos [4], para transformar el metamodelo de Web Services a diferentes metamodelos de lenguajes de implementación.

1.1. WEB SERVICES

Los Web Services proveen esencialmente un medio estándar de comunicación entre diferentes aplicaciones de software. Su uso en Internet se ha extendido rápidamente debido a la creciente necesidad de comunicación e interoperabilidad entre aplicaciones. Se definen como aplicaciones auto-contenidas y auto-descritas, que pueden ser publicadas, localizadas e invocadas a través de la Web. Una vez desarrolladas, otras aplicaciones (y otros Web Services) pueden descubrirlas e invocar el servicio dado [1].

Los Web Services se basan en estándares abiertos, lo que permite a las aplicaciones comunicarse más libre e independientemente de las plataformas en que se están ejecutando. Las tecnologías subyacentes que permiten su implementación son: Web Service Description Language (WSDL) [5] [6], Simple Object Access Protocol (SOAP) [7], Hiper Text Transport Protocol (HTTP) y Universal Description, Discovery and Integration (UDDI) [8]. El uso de estos protocolos estándares es necesario para lograr la interoperabilidad en ambientes heterogéneos, con independencia del Sistema Operativo, el lenguaje de programación, etc.

1.2. MDA

En la actualidad, la construcción de software se enfrenta a continuos cambios en las tecnologías de implementación, lo que implica realizar esfuerzos importantes tanto en el diseño, para integrar las diferentes tecnologías que incorpora, como en el mantenimiento, para adaptar la aplicación a los cambios en los requisitos y en las tecnologías de implementación. Para conseguir los beneficios de MDA se plantea el siguiente proceso de desarrollo: de los requisitos se obtiene un modelo independiente de la plataforma (Platform Independent Model-PIM), luego este modelo es transformado con la ayuda de herramientas en uno o más modelos específicos de la plataforma (Platform Specific Model-PSM), y finalmente cada PSM es transformado en código. Por lo tanto, MDA incorpora la idea de transformación de modelos (PIM a PSM, PSM a código) y se necesitan herramientas para automatizar esta tarea. Estas herramientas constituyen uno de los elementos básicos de MDA.

1.3. TRANSFORMACIONES DE MODELOS

La transformación de modelos es una de las bases en las que se fundamenta la metodología MDA, en particular la transformación automática de modelos. Para lograr esta automaticidad en la transformación de modelos se pueden utilizar diferentes herramientas, para esta propuesta se plantea la

utilización de dos enfoques para la definición de las transformaciones antes planteadas, el primero es el uso del lenguaje Relations que forma parte de QVT [3], y el segundo el uso de técnicas de transformación de grafos [4]. A continuación se describe brevemente cada uno de estos enfoques.

1.4. QVT

El planteamiento QVT se basa principalmente en la definición de un lenguaje para las consultas (Queries) sobre los modelos Meta Object Facility (MOF) [9], la búsqueda de un estándar para generar vistas (Views) que revelen aspectos específicos de los sistemas modelados, y finalmente, la definición de un lenguaje para la descripción de transformaciones (Transformations) de modelos MOF.

En este trabajo se utiliza el componente de QVT que tiene como objetivo definir transformaciones. Estas transformaciones describen relaciones entre un metamodelo fuente F y un metamodelo objetivo O, ambos metamodelos deben estar especificados en MOF. Luego esta transformación definida se utiliza para obtener un modelo objetivo, el cual es una instancia del metamodelo O, a partir de un modelo fuente que es una instancia del metamodelo F. Una característica muy importante de estas transformaciones es que pueden ser bidireccionales (multidimensionales también).

1.4.1. EL LENGUAJE RELATIONS

La especificación de QVT que se utiliza tiene una naturaleza híbrida declarativa/imperativa. El lenguaje Relations es una especificación declarativa de relaciones entre metamodelos MOF. Este lenguaje permite realizar “pattern matching” de objetos complejos y definir “templates” de creación de objetos.

1.5. LAS REGLAS DE TRANSFORMACIÓN DE GRAFOS

Para la construcción y evolución de modelos es necesario asegurar la consistencia entre los mismos, a través de la transformación de modelos. Para llevar a cabo esta transformación de modelos se propone usar las reglas de transformación de grafos y aplicarlas a la transformación de modelos en el campo de MDA. La transformación de grafos es una técnica gráfica, formal, declarativa y de alto nivel muy usada para transformación y simulación de modelos, chequeo de consistencia entre modelos o vistas y optimización en diversos contextos. Los artefactos producidos para conceptualizar un sistema son llamados modelos, y los diagramas son usados para visualizar su estructura compleja de una forma más intuitiva y natural. A través de los grafos pueden ser definidas las estructuras de esos modelos, entonces la transformación de grafos puede ser explorada para especificar cómo deben ser construidos y cómo deben evolucionar. Esta técnica formal para la manipulación de grafos está basada en reglas. Los modelos gráficos se interpretan como grafos y se usan las reglas de transformación de grafos para construir diagramas correctos.

2. LINEAS DE INVESTIGACION Y DESARROLLO

Una de las alternativas para implementar la transformación propuesta es usar QVT. A continuación se presenta esta alternativa.

2.1. TRANSFORMACIÓN DE MODELOS WEB SERVICES BASADA EN QVT

El esquema general de la transformación, de Web Services genéricos a la definición de estos en un lenguaje de implementación, puede ser visto en tres niveles: metamodelo, definición/modelo y ejecución, como lo muestra la Figura 1.

En el nivel *metamodelo* se encuentran los metamodelos involucrados en la transformación, estos son: el metamodelo de Web Services y el metamodelo de algún lenguaje de implementación, por ejemplo Java. Entre estos modelos se define la transformación, en este caso, mediante el lenguaje de transformación Relations de QVT. Pasando al nivel de *modelo/definición* se encuentran los modelos específicos que definen un Web Service. Por ejemplo, a partir de un modelo Web Service para reservas de hoteles de una agencia de viajes, y por aplicación de la transformación a nivel metamodelo, se obtiene el modelo Java que implementa dicho Web Service. Por último, en el nivel de *ejecución*, se encuentran los sistemas desarrollados en Java que implementan el Web Service de la agencia de viajes.

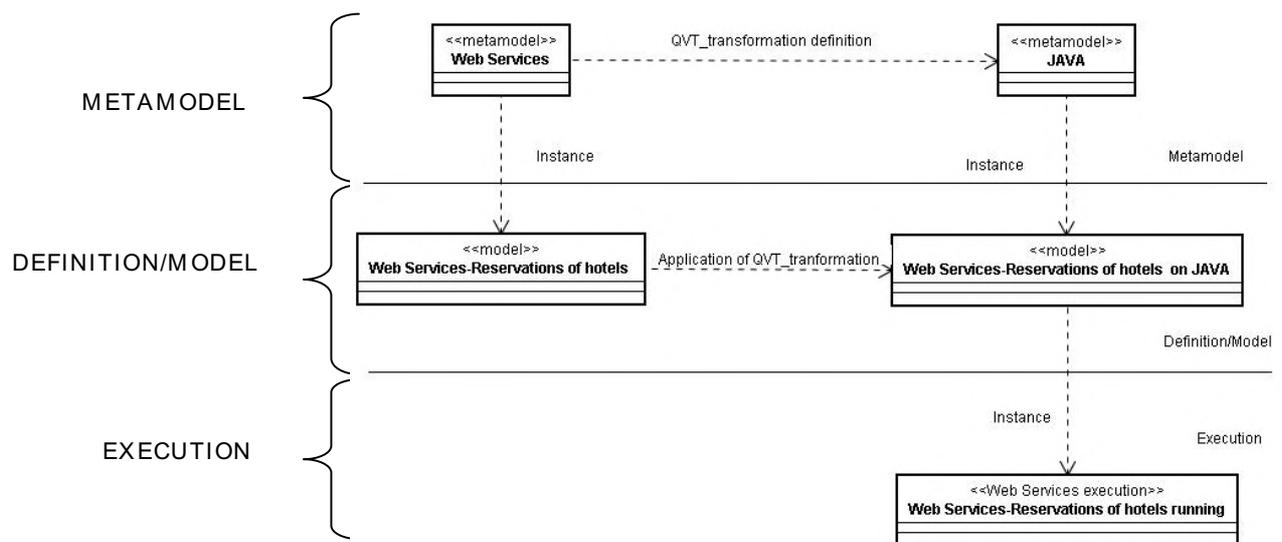


Figura 1: Vista general de la transformación.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados de la investigación aportan a la mejora del desarrollo de aplicaciones que utilizan Web Services, esto se logra con la utilización de herramientas (transformaciones, metamodelos) para el desarrollo de este tipo de aplicaciones en el marco de MDA.

El beneficio de estas transformaciones queda reflejado teniendo en cuenta los constantes cambios en los requerimientos de un sistema. Esto se debe a que cualquier cambio en la especificación de un Web Service de una aplicación, podrá ser propagado a la especificación en el lenguaje de implementación de la aplicación, y así, adaptar rápidamente los sistemas (escritos en diferentes lenguajes) a los cambios en los Web Services que implementan.

En síntesis, los principales objetivos planteados en esta investigación son:

- Aplicar, adaptar y mejorar las técnicas utilizadas en arquitecturas orientadas a modelos, tanto para el desarrollo de software como aquellas orientadas a mejorar la calidad del software.
- Definir y formalizar el metamodelo para Web Services.
- Definir y formalizar metamodelos para lenguajes de programación específicos.
- Especificar las transformaciones entre metamodelos, que tengan como origen el metamodelo Web Services y como destino un metamodelo de un lenguaje/plataforma específico.
- Desarrollar un caso de estudio, implementando los módulos necesarios para soportar la interoperabilidad entre sistemas a través de Web Services.

4. FORMACION DE RECURSOS HUMANOS

Los estudios realizados en esta línea de investigación sirven como marco para la elaboración de trabajos finales de grado y tesis de posgrado. En este punto, una estudiante de la carrera Licenciatura en Ciencias de la Computación ha iniciado su trabajo final y consiste en la implementación de los Web Services necesarios para compartir información de alumnos entre dos sistemas utilizados de la UNRC, denominados SIAL (Sistema Integral de Alumnos) y SIAT (Sistema Informático de Apoyo a la Teleformación). El primero permite la gestión de alumnos de grado y es desarrollado por el Centro de Cómputos de la UNRC, y el segundo permite la gestión de alumnos de cursos y carreras a distancia y es desarrollado por la Secretaría de Extensión de la UNRC. La aplicación de las transformaciones especificadas posibilitará el desarrollo de este trabajo final utilizando la metodología MDA, alcanzando las ventajas de su aplicación al desarrollo de software, como son la portabilidad, interoperabilidad, productividad y mantenimiento.

Los temas abordados en esta línea de investigación brindan un fuerte aporte al proceso de perfeccionamiento continuo de los autores de este trabajo, que se desempeñan como docentes de las carreras de computación que se dictan en la Universidad Nacional de Río Cuarto y participan en asignaturas relacionadas a dichos temas.

Además, este proyecto aporta a la construcción de estructuras genéricas orientadas a la reutilización de modelos en diferentes plataformas de implementación. Al mismo tiempo, se proveen herramientas para la construcción de software en el marco de MDA, que podrán ser usadas por otros desarrolladores de aplicaciones que involucran el uso de Web Services.

5. BIBLIOGRAFIA

- [1] World Wide Web Consortium. Web Service Architecture. <http://www.w3.org/TR/ws-arch/>.
- [2] OMG, Model Driven Architecture (MDA) Guide. Miller, J., Mukerji, J. (eds) 2003. document number ormsc/2001-07-01. Obt. en Mayo de 2007 de <http://www.omg.org/docs/omg/03-06-01.pdf>
- [3] Object Management Group, "Meta Object Facility (MOF) 2.0 Query/View/Transformation Specification" Final Adopted Specification ptc/05-11-01, <http://www.omg.org/docs/ptc/05-11-01.pdf>, último acceso Febrero 2008.
- [4] Baresi L., Heckel R. Tutorial Introduction to Graph Transformation: A Software Engineering Perspective. First International Conference on Graph Transformation. Spain. v. 2505, pp. 402-429. (ICGT 2002). 2002.
- [5] World Wide Web Consortium. Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language. <http://www.w3.org/TR/wsdl20>. Último acceso Mayo 2007.
- [6] World Wide Web Consortium. Web Services Description Language (WSDL) Version 2.0 Part 0: Primer. <http://www.w3.org/TR/wsdl20-primer>. Último acceso Mayo 2007.
- [7] World Wide Web Consortium. SOAP Version 1.2 Part 1: Messaging Framework. <http://www.w3.org/TR/soap12-part1/>. Último acceso Mayo 2007.
- [8] OASIS. UDDI Version 3.0.2. http://uddi.org/pubs/uddi_v3.htm. Último acceso Mayo 2007.
- [9] Object Management Group "Meta Object Facility (MOF) Core Specification" OMG Available Specification. Vers 2.0. formal/06-01-01, <http://www.omg.org/docs/formal/06-01-01.pdf>. (Nov. 2006).

Un marco para la Medición de Modelos Workflow

Mario Peralta, Daniel Riesco, Carlos Salgado, Germán Montejano

Departamento de Informática – Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis

Ejército de los Andes 950, C.P. 5700, San Luis, Argentina

{mperalta, driesco, csalgado, gmonte}@unsl.edu.ar – web: <http://www.sel.unsl.edu.ar>

Félix García Rubio

Grupo Alarcos

Departamento de Tecnologías y Sistemas de Información

Centro Mixto de Investigación y Desarrollo de Software UCLM-INDRA

Universidad de Castilla-La Mancha

Paseo de la Universidad, 4, 13071 Ciudad Real, España

Felix.Garcia@uclm.es – web: <http://alarcos.inf-cr.uclm.es>

CONTEXTO

El presente trabajo se enmarca en los Proyectos de Investigación: Ingeniería de Software, Conceptos, Métodos y Herramientas en un Contexto de “Ingeniería de Software en Evolución” - Facultad de Ciencias Físico Matemáticas y Naturales, Universidad Nacional de San Luis, y Proyecto MECENAS (Junta de Comunidades de Castilla-La Mancha, Consejería de Educación y Ciencia, PBI06-0024), Grupo ALARCOS, Universidad de Castilla La Mancha, España.

RESUMEN

En todo proceso de negocio es muy importante su mejora continua, lo que lleva a las organizaciones a buscar herramientas que den soporte a la realización de estas mejoras. La Gestión de Procesos de Negocio proporciona este soporte mediante los Sistemas de Gestión de Procesos de Negocio (BPMS). Una de las tecnologías más significativas para dar soporte a la automatización de los procesos de negocio son los Sistemas de Gestión de Workflow (SGW). Para favorecer y dar flexibilidad a los SGW, de manera que puedan adaptarse a los cambios constantes del negocio, es fundamental tener un lenguaje de modelado

que permita una fácil adaptación de los modelos. De la misma manera es fundamental tener herramientas que permitan medir la calidad de dichos modelos. Acorde a esto, nuestra investigación se centra en la definición de un marco de trabajo para el modelado y la medición de procesos workflow que ayude a la mejora y mantenimiento de los modelos y de los procesos que ellos representan.

Palabras clave: *Workflow – Sistema de Gestión Workflow – Proceso de Negocio – Lenguaje de Modelado Workflow – Métricas*

1. INTRODUCCION

Analizando el ciclo de vida de los procesos de negocio [1], es de gran importancia llevar a cabo una mejora continua de los mismos y ello ha llevado a las organizaciones a buscar herramientas que proporcionen el soporte necesario para poder realizar dichas mejoras. Hoy en día, la Gestión de Procesos de Negocio (BPM), definida como “la capacidad de descubrir, diseñar, desarrollar, ejecutar e interactuar con la operación, optimización y análisis de procesos a nivel de diseño de negocio” ([2]), proporciona este soporte mediante los Sistemas de Gestión de Procesos de

Negocio (BPMS). En este sentido, Una de las tecnologías más significativas para dar soporte a la gestión de procesos de negocio son los Sistemas de Gestión de Flujos de Trabajo (Workflow Management Systems, WFMS) que soportan la automatización de los procesos de negocio y que la Workflow Management Coalition define como: “un sistema que define, crea y dirige la ejecución de flujos de trabajo a través del uso de software que funciona en uno o más motores de flujo de trabajo, siendo capaz de interpretar la definición de proceso, de interactuar con los participantes del flujo de trabajo e invocar el uso de herramientas y aplicaciones de las IT” [3].

En el campo de los procesos workflow, se puede observar un importante trabajo en cuanto al modelado de dichos procesos, surgiendo así diversas líneas de investigación. Entre estos trabajos podemos destacar propuestas en las que se utilizan los Patrones Workflows para realizar la evaluación y/o comparación de distintos lenguajes de modelado. Por ejemplo, en [4] se utilizan los patrones workflow, junto con un conjunto de patrones de comunicación para analizar BPEL4WS. En el trabajo se presenta una representación posible de cada patrón, respecto del flujo de control, en BPEL4WS. Además, se lo compara con otros lenguajes de modelado workflow como XLANG y WSFL, y Staffware PLC's Staffware e IBM's MQSeries Workflow, cuya evaluación es presentada en [5]. En [6] se examina cómo dos lenguajes de modelado: *Diagramas de proceso de Negocio* (BPMN) de BPMI y los *Diagramas de Actividad* de UML de la OMG pueden representar gráficamente los patrones workflow respecto al flujo de control. En [7], [8] y [9] los autores presentan una evaluación de los diagramas de Actividad contrastados con los patrones workflow desde las tres perspectivas: del flujo de control, datos y recursos. En [10], se presenta una evaluación de BPMN en función de su expresividad respecto de los Patrones Workflow.

Desde otro punto de vista, en el ámbito de la medición, se pueden observar diversos trabajos en cuanto a la medición de los procesos de negocio como en [11], [12]. Sin embargo, y a pesar de la importancia de medir la calidad de los procesos workflow, es muy poco el trabajo detectado en cuanto a la medición de calidad de estos procesos y de los modelos que los representan. Entre los trabajos en este campo se puede destacar las propuestas realizadas en [13], [14], [15], donde se propone una métrica para la medición de la complejidad del flujo de control basada en la complejidad ciclomática de McCabe. En [15] se introduce el concepto de complejidad workflow y se propone una definición de esta complejidad y una clasificación de los procesos workflows de acuerdo a su complejidad en *Ordenados, Estructurados y Aleatorios*. Otros trabajos que se pueden mencionar en este campo es la propuesta introducida en [16, 17]. En estos trabajos los autores introducen métricas para evaluar la cohesión interna de las actividades en un proceso workflow y el acoplamiento entre las actividades de un proceso Workflow. Estas propuestas presentan alternativas para la medición de algunos aspectos del modelado de procesos workflows. Pero no cubren todos los aspectos relevantes a los mismos.

Bajo estas consideraciones, en el ámbito de nuestro trabajo de investigación consistente en la definición de un marco de trabajo para el modelado y la medición de procesos workflow que ayude a la mejora y mantenimiento de los modelos y de los procesos que ellos representan, proponemos un conjunto de medidas para modelos de procesos workflow que creemos ayudarán a la evaluación, comparación y mejora de los modelos workflows y, en consecuencia, de los procesos que ellos representan.

2. LINEAS DE INVESTIGACION y DESARROLLO

Todo sistema de gestión workflow debe garantizar que la organización realiza las

tareas correctas, en el momento y de la forma adecuada. Por ello se considera fundamental tener una buena representación del proceso en la que se incluyan todos los aspectos de interés para dicho proceso. Para poder lograr dicha representación es necesario disponer de un lenguaje de modelado de procesos adecuado que, además, permita adaptar esa representación a los cambios continuos que los negocios de hoy experimentan.

En la actualidad existen numerosos lenguajes de modelado que pueden ser considerados apropiados según las necesidades del negocio, por lo que a veces los diseñadores deben decidir, sobre un conjunto de herramientas que dan soporte a distintos lenguajes de modelado, cuál o cuáles se adecuan más a sus necesidades. Por ello creemos que es necesario tener criterios que ayuden a tomar esta decisión.

Otro de los aspectos fundamentales en todo proceso workflow, es tener herramientas que permitan medir sus elementos más relevantes con el fin de detectar las áreas y aspectos a mejorar del proceso y, de esta manera, promover su mejora continua. Para ello creemos que es necesario proveer un marco que permita realizar dicha medición.

De acuerdo a esta necesidad, siguiendo la metodología propuesta en [18], hemos definido un conjunto de medidas iniciales y elementales que creemos servirán como indicadores de la complejidad estructural de los modelos de procesos workflow.

Nuestra propuesta pretende proveer un marco que permita incluir todos los aspectos más relevantes del modelado workflow. Y para ello consideramos que es necesario definir un conjunto más amplio de medidas que permitan la medición individual de cada uno de los aspectos relevantes para el modelado como así también la complejidad estructural global del proceso completo y de los modelos que lo representan.

Además, debido a que no se han detectado trabajos que propongan la definición de un marco de medición y comparación de la mantenibilidad de los modelos workflow,

nuestra investigación se centra en la evaluación de lenguajes de modelado workflow desde el punto de vista de la mantenibilidad de los modelos definidos con dichos lenguajes. El objetivo es proporcionar a los diseñadores de procesos workflow los criterios necesarios para la selección del lenguaje de modelado más adecuado que facilite la evolución de los modelos desarrollados.

3. RESULTADOS OBTENIDOS/ESPERADOS

Como mencionamos en el apartado anterior, y siguiendo los objetivos de nuestra investigación, hemos definido un conjunto de medidas para la evolución de modelos workflow. Dichas medidas son una adaptación de las medidas propuestas en [19] para la medición de procesos software. Acorde a los lineamientos establecidos en [18] para la definición de medidas, hemos realizado un primer experimento para validar empíricamente las medidas propuestas, comparando dos lenguajes de modelado workflow como lo son YAWL y los Diagramas de Actividad de UML 2.0. A través de este experimento observamos una correlación de las medidas con el tiempo de entendibilidad de los modelos YAWL en 4 de las 9 medidas validadas mientras que para Diagramas de Actividad solo para 2. Respecto de los tiempos de modificabilidad no se observó correlación en ninguno de los lenguajes analizados.

En la continuidad del trabajo se pretende ampliar el número de medidas, como así también replicar el experimento de manera de poder dar mayor consistencia a los resultados obtenidos hasta el momento.

4. FORMACION DE RECURSOS HUMANOS

Basados en la temática planteada, se están desarrollando tesis de Maestría y Doctorado por parte de algunos integrantes de los Proyecto. En el marco de la Maestría en

Ingeniería de Software que se dicta en la UNSL, dentro del marco del Proyecto, se han dictado charlas destinadas a los maestrandos acerca de la temática del Modelado Workflow y sobre el trabajo que se está desarrollando al respecto.

5. BIBLIOGRAFIA

- [1] D. Georgakopoulos and A. Tsalgaidou, "Technology and Tools for Comprehensive Business Process Lifecycle Management,," *Workflow Management Systems and Interoperability. Springer V. p.*, pp. 324-365, 1998.
- [2] H. Smith, "The Emergence of Business Process Management," *CSC's Research Services*, 2002.
- [3] P. Lawrence, *Workflow Handbook 1997.*. New York: Workflow Management Coalition, 1997.
- [4] P. Wohed, W. M. P. van der Aalst, M. Dumas, and A. H. M. ter Hofstede, "Pattern Based Analysis of BPEL4WS," Queensland University of Technology, Brisbane FIT-TR-2002-04, QUT, 2002.
- [5] W. M. P. van der Aalst, A. H. M. ter Hofstede, B. Kiepuszewski, and A. P. Barros, "Workflow Patterns," Queensland University of Technology, Brisbane QUT Technical report. FIT-TR-2002-02, 2002.
- [6] S. A. White, "Process Modeling Notations and Workflow Patterns," in *Workflow Handbook 2004*, L. Fischer, Ed.: Published in association with the Workflow Management Coalition (WfMC), 2004.
- [7] P. Wohed, W. M. P. van der Aalst, M. Dumas, A. H.M. ter Hofstede, and N. Russell, "Pattern-based Analysis of UML Activity Diagrams," 2004.
- [8] P. Wohed, W. M. P. v. d. Aalst, M. Dumas, A. H.M. ter Hofstede, and N. Russell, "Pattern-based Analysis of the Control-flow Perspective of UML Activity Diagrams," 2005.
- [9] N. Russell, W. M. P. van der Aalst, A. H. M. ter Hofstede, and P. Wohed, "On the Suitability of UML 2.0 Activity Diagrams for Business Process Modelling," 2006.
- [10] P. Wohed, W. M. P. van der Aalst, M. Dumas, A. H. M. ter Hofstede, and N. Russell, "On the Suitability of BPMN for Business Process Modelling," presented at 4th International Conference on Business Process Management (BPM 2006), LNCS., Vienna, Austria, 2006.
- [11] E. Rolon, F. Ruiz, F. Ó. Garcia Rubio, and M. Piattini, "Aplicación de Métricas Software en la Evaluación de Modelos de Procesos de Negocio."
- [12] E. Rolon, F. Ó. Garcia Rubio, F. Ruiz, and M. Piattini, "Validating a Set of Measures for Business Process Models Usability and Maintainability."
- [13] J. Cardoso, "How to Measure the Control-flow Complexity of Web Processes and Workflows," in *Workflow Handbook 2005*, 2005.
- [14] J. Cardoso, "Control-flow Complexity Measurement of Processes and Weyuker's Properties," 2005.
- [15] J. Cardoso, "Approaches to Compute Workflows Complexity," presented at Dagstuhl Seminar, The Role of Business Processes in Service Oriented Architectures, Dagstuhl, germany, 2006.
- [16] H. A. Reijers, "A Cohesion Metric for the Definition of Activities in a Workflow Process,," *Eighth CAiSE/IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design 2003*, pp. 116-125, 2003.
- [17] H. A. Reijers and I. T. P. Vanderfeesten, "Cohesion and

- Coupling Metrics for Workflow Process Design," *BPM 2004, LNCS 3080*, pp. 290-305, 2004.
- [18] M. Serrano, M. Piattini, C. Calero, M. Genero, and D. Miranda, "Un método para la definición de métricas de software.," presented at 1er Workshop en Métodos de Investigación y Fundamentos filosóficos en Ingeniería del Software y Sistemas de Información (MIFISIS'2002),, 2002.
- [19] F. Ó. García Rubio, "FMESP: Marco de Trabajo Integrado para el Modelado y la Medición de los Procesos Software," in *Departamento de Informática*. Ciudad Real. España: U.C.L.M. Universidad de Castilla La Mancha. España, 2004, pp. 491.

Calidad en los Modelos de Procesos de Negocio

C. Salgado, M. Peralta, D. Riesco, G. Montejano
Departamento de Informática
Universidad Nacional de San Luis
San Luis, Capital, Argentina
Ejército de los Andes 950 – San Luis – San Luis – Argentina
C.P.: 5700
Tel.: 54-02652-424027 – Int. 251
[csalgado, mperalta, driesco, gmonte]@unsl.edu.ar – web: <http://www.sel.unsl.edu.ar>

CONTEXTO

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software, Conceptos, Métodos y Herramientas en un Contexto de “Ingeniería de Software en Evolución” – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis.

RESUMEN

La Gestión de Procesos de Negocio combina una visión centrada en procesos y de integración de funcionalidades destinadas a mejorar la efectividad de la organización. Ella proporciona los medios necesarios para una efectiva implementación de los procesos, y provee las funcionalidades necesarias para que los gerentes de la empresa puedan controlar y modificar los flujos de trabajo en dichos procesos. Una herramienta de gran ayuda para mejorar dicho control son los modelos de procesos, ya que brindan una descripción fácil de entender de su estructura y complejidad. Considerando la importancia que tienen los modelos en el éxito o fracaso del proceso de negocio, nuestra propuesta se centra en la evaluación del nivel conceptual de estos modelos. Bajo estas consideraciones, el uso de métricas puede ser un aspecto clave para obtener modelos de más calidad que puedan servir como soporte para mejorar el mantenimiento, actualización, adaptación, de los procesos de negocio. En este sentido proponemos una especificación formal de un conjunto de

métricas, propuestas para modelos conceptuales de procesos de negocio representados con BPMN, utilizando OCL, basándonos en el metamodelo del OMG para definir de forma uniforme cada una de las métricas en los diferentes modelos del paquete BPDM.

Palabras clave: Procesos de Negocio, Métricas, BPDM, Especificación de Métricas, Metamodelos OMG.

1. INTRODUCCION

Dada la compleja naturaleza de los procesos de negocio se han llevado a cabo estudios sobre diversos aspectos tales como la utilidad [1], evaluación de la calidad [2] o la medición [3], siendo aún más frecuentes los estudios referentes a la utilización de las diferentes herramientas y lenguajes para llevar a cabo el modelado de procesos de negocio [4] [5] [6]. Este interés se ha visto fomentado principalmente por la diversidad de notaciones y lenguajes existentes para el modelado, definición y ejecución de los procesos de negocio.

Los modelos de procesos de negocio (MPN) tienen un amplio rango de usos tales como el soporte a la re-ingeniería de procesos, la simulación o servir como base para el desarrollo de sistemas que automatizan dichos procesos. Estos modelos pueden ser creados o presentados usando diversos lenguajes, que son bastante diferentes entre sí, dado que cada uno tiene una manera diferente de ver los procesos

dependiendo del propósito para el cuál fue creado [5].

Entre los lenguajes para el modelado de procesos de negocio mencionados en la literatura que merecen especial atención, se encuentran: IDEF0 [7], IDEF3 [8], UML [9], UML 2.0 [10] y BPMN.

En especial, BPMN proporciona una notación gráfica para expresar procesos de negocio mediante un Diagrama de Proceso de Negocio (DPN) que está basado en una técnica de diagramas de flujo adaptada para la creación de modelos gráficos de las operaciones de procesos de negocio. Un DPN está compuesto de dos categorías básicas de elementos, en la primera se encuentran los elementos centrales con los cuales es posible desarrollar modelos de procesos simples, mientras que la segunda incluye los elementos que permiten la creación de MPN's complejos o de alto nivel.

Desde otro punto de vista, en [11] se presenta la definición del Metamodelo de Definición de Procesos de Negocio (BPDM), que proporciona un mecanismo de serialización y un metamodelo explícito para conceptos BPMN. La integración de BPMN y BPDM proporciona un conjunto integrado de estándares que cubren el modelo y la notación subyacentes para la *orquestración de procesos*.

En este sentido, basados en la notación estándar del BPDM, en este trabajo se propone la especificación con OCL de algunas métricas para modelos conceptuales de procesos de negocio representados con BPMN [12]. Estas métricas están basadas en el marco FMESP (*Framework for the Modeling and Evaluation of Software Processes*) [13] para la medición de procesos software.

Las métricas propuestas en [12] han sido definidas contando los diferentes tipos de elementos que componen un MPN representado con BPMN (como los objetos de flujo: constructores Evento, Actividad, Decisión o Unión). Las métricas definidas para modelos de proceso de negocio representados con el metamodelo BPMN

han sido agrupadas en dos categorías, siguiendo las indicaciones de la Ontología de la Medición del Software [14]: Medidas Base y Medidas Derivadas

Las medidas base consisten principalmente en contar los elementos significativos del modelo de proceso de negocio [12].

En función de estas medidas base, la propuesta de métricas para modelos de procesos de negocio incluye algunas métricas derivadas significativas que establecen las proporciones existentes entre los diferentes elementos del modelo.

Analizando los distintos paquetes de BPDM, encontramos que son de gran ayuda para la especificación de las métricas propuestas. Basados en dichos paquetes, proponemos una especificación de estas métricas, en la cual se da: una *descripción de la métrica*, los *Metamodelos Usados*, las *Metaclases Usadas* y una *Definición Formal en OCL*.

Las métricas propuestas permiten evaluar la complejidad estructural de un modelo, por ejemplo en función del número de actividades atómicas y compuestas. Estos parámetros son de utilidad a la hora de evaluar la complejidad de los modelos para decidir cuál es más adecuado para modelar un proceso de negocio, lo que redundará en una mejor performance del negocio a la hora de implementarlo.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

En la última década, los procesos de negocio han adquirido gran popularidad, y derivado de ello han surgido diversos estudios al respecto. Sin embargo, actualmente los estudios relacionados con la evaluación y medición de procesos de negocio que se encuentran en la bibliografía están dirigidos principalmente al nivel de ejecución de los procesos, es decir, se analizan aspectos tales como la evaluación de los resultados obtenidos, los tiempos de ejecución, costos del proceso, etc.

Proporcionar información objetiva acerca de la mantenibilidad de los modelos, facilita

la evolución de los procesos de negocio de aquellas empresas que evalúan constantemente sus procesos al estar involucradas en una mejora continua. Al mismo tiempo se estaría proporcionando soporte a la gestión de procesos de negocio al facilitar la evaluación temprana de ciertas propiedades de calidad de sus modelos, con lo cual las organizaciones se verían beneficiadas en dos maneras: 1) Garantizando el entendimiento y la difusión de los procesos, y su evolución, sin afectar su ejecución; 2) Reduciendo el esfuerzo necesario para cambiar los modelos con la consecuente reducción de los esfuerzos de mantenimiento y/o mejora.

Por lo anterior, el interés de nuestra investigación se centra en la utilización de métricas para la evaluación a nivel conceptual de los modelos de procesos de negocio. Esto puede ser un aspecto clave para obtener modelos de calidad que sirvan como soporte para mejorar el mantenimiento y la adaptabilidad de los procesos de negocio a los cambios constantes a los que dichos procesos están sometidos hoy en día.

En función de esto, proponemos realizar la especificación de métricas para modelos conceptuales de procesos de negocio utilizando OCL y basados en el estándar de BPDM. De manera de poder utilizar dicha propuesta como una herramienta de validación de los modelos analizados y así introducir mejoras en la calidad de los modelos en una etapa temprana, lo que redundará en una mejor calidad en los procesos de negocio a la hora de su implementación.

3. RESULTADOS OBTENIDOS/ESPERADOS

En esta primera etapa hemos realizado la especificación de métricas para modelos de procesos de negocio utilizando OCL y basados en el estándar de BPDM. Para realizar la especificación de dichas métricas consideramos los distintos aspectos que abarca el metamodelo BPDM, como por

ejemplo: Objetos de Conexión, Carriles y Artefactos; Control de Decisiones de los Objetos; Evento de los Objetos de Flujo, etc.

En la continuidad de este trabajo se prevee definir nuevas métricas para los modelos conceptuales de procesos de negocio. Además, la selección de OCL para la especificación de las métricas facilitará, en un paso posterior, la creación de una herramienta que de forma automática, permita la evaluación de modelos de procesos de negocio en función de las métricas propuestas.

4. FORMACION DE RECURSOS HUMANOS

Basados en la temática planteada en el presente trabajo, se están desarrollando tesis de Maestría cuyo objetivo es proveer herramientas que ayuden a la mejora de los procesos de negocio.

En estos momentos tenemos alumnos llevando a cabo el desarrollo de una herramienta para la aplicación de las métricas propuestas. Dicho trabajo se está desarrollando como tesis de Licenciatura de los alumnos.

5. BIBLIOGRAFIA

- [1] M. A. Rappa, "The utility business model and the future of computing services," *IBM Systems Journal*, vol. 43, pp. 32-42, 2004.
- [2] J. Becker, M. Rosemann, and C. von Uthmann, "Guidelines of Business Process Modeling," *Business Process Management, Models, Techniques and Empirical Studies (BPM'00)*. Springer, pp. 30-49, 2000.
- [3] V. Vitolins, "Business Process Measures," presented at Int. Conference on BALTIC DB&IS. Riga, Latvia., 2004.
- [4] C. Dewalt, "Business Process Modeling with UML," *Johns Hopkins University*, 1999.

- [5] T. Dufresne and J. Martin, "Process Modeling for e-Business," *Dr. Larry Kerschberg (ed) Information Systems Department, George Mason University*, 2003.
- [6] S. A. White, "Process Modeling Notations and Workflow Patterns," in *Workflow Handbook 2004*, L. Fischer, Ed.: Published in association with the Workflow Management Coalition (WfMC), 2004.
- [7] FIPS, "Integration Definition for Function Modeling (IDEF0), Standard.," *National Institute of Standards and Technology*, 1993.
- [8] R. J. Mayer, C. P. Menzel, and e. al., "Information Integration for Concurrent Engineering (IICE) IDEF3 Process Description Capture MethodReport.," *Interim Technical Report*, 1995.
- [9] H.-E. Erickson and M. Penker, "Business Modeling with UML-Business Patterns at Work," *ed. I. John Wiley & Sons. USA: Robert Ipsen.*, 2000.
- [10] OMG, "Unified Modeling Language (UML) Specification: Infrastructure", version 2.0," *Object Management Group.*, 2003.
- [11] "Business Process Definition MetaModel (BPDM), Beta 1 OMG Adopted Specification," <http://www.omg.org/docs/dtc/07-07-01.pdf>.
- [12] E. Rolon, F. Ruiz, F. Ó. Garcia Rubio, and M. Piattini, "Aplicación de Métricas Software en la Evaluación de Modelos de Procesos de Negocio," *Revista Electrónica de la Sociedad Chilena de Ciencia de la Computación*, vol. 6, 2005.
- [13] F. Ó. García Rubio, "FMESP: Marco de Trabajo Integrado para el Modelado y la Medición de los Procesos Software," in *Departamento de Informática. Ciudad Real. España: U.C.L.M.* Universidad de Castilla La Mancha. España, 2004, pp. 491.
- [14] F. Ó. García Rubio, M. F. Bertoa, C. Calero, A. Vallecillo, F. Ruiz, M. Piattini, and M. Genero, "Towards a Consistent Terminology for Software Measurement," *Information and Software Technology*, 2006.

Entornos para usar BPM en aplicaciones JAVA: un análisis comparativo

Javier Díaz ¹, Patricia Bazán ¹, Claudia Queiruga ¹, Gabriel Gotelli ², Facundo Rodríguez ²

¹ LINTI Facultad de Informática UNLP ² Facultad de Informática UNLP

CONTEXTO

Este es un trabajo de fin de carrera de Licenciatura en Sistemas de la Facultad e Informática de la UNLP, de los alumnos Gabriel Gotelli y Facundo Rodríguez, dirigidos por la Lic. Patricia Bazán y la Lic. Claudia Queiruga

RESUMEN

La gestión de los procesos de negocio o BPM (Business Process Management) es una metodología empresarial que permite representar las actividades de los procesos de negocios y las restricciones que se aplican sobre ellas. Automatizar esta tarea conlleva el beneficio de contar con un soporte que modela, simula y monitorea las actividades de dichos procesos de negocio.

Este trabajo tiene como objetivo comparar, mediante un conjunto de métricas preestablecidas, las herramientas o entornos de desarrollo actuales que soportan BPM particularmente con soporte para construcción de aplicaciones en JAVA. Se construirá un cuadro comparativo a partir de dichas métricas que permita describir las características más salientes de las herramientas y contar con algunos patrones a tener en cuenta a la hora de elegir una de ellas para evaluarla y, eventualmente, adoptarla.

Palabras clave: servicio, proceso, modelo de proceso, JAVA.

1. INTRODUCCION

Un proceso de negocio es una actividad del mundo empresarial que consta de un conjunto de tareas lógicamente relacionadas, que cuando se realizan en la secuencia apropiada y siguiendo las reglas del negocio, producen una salida válida para el negocio. Por ejemplo, realizar una transacción bancaria [1]. Actualmente, los procesos de negocios se están tornando ineficientes y obsoletos básicamente por su incapacidad de adaptarse a los cambios. Por ende, la necesidad de modelizar y optimizarlos es cada vez más importante para las organizaciones [2].

El cambio de un proceso de negocios involucra examinar el mismo con el objetivo de reducir el número de actividades, eliminando las tareas de menor importancia y simplificando el proceso en general. Con el objetivo de poder introducir cambios, es necesario adoptar un enfoque que permita rediseñar los procesos de negocios. Típicamente, las áreas problemáticas de un cambio en los procesos, solo pueden ser identificadas una vez que el mismo ha sido físicamente implementado. Por lo tanto, tener la capacidad de visualizar y evaluar un cambio en los procesos antes de su implementación puede tener un impacto positivo sobre la tasa de éxito de los futuros cambios. Una posible forma de lograr esto es utilizar un modelo dinámico de procesos de negocios [2].

Se llama Business Process Management (BPM) a la metodología empresarial cuyo objetivo es mejorar la eficiencia a través de la gestión sistemática de los procesos de negocio que se deben modelar, automatizar, integrar, monitorear y optimizar de forma continua [3].

BPMS (BPM Suite) es el conjunto de servicios y herramientas informáticas que facilitan la administración de procesos de negocio.

Este trabajo presenta un análisis comparativo de cuatro BPMSs actuales utilizando un conjunto de métricas específicas.

a. Definiciones y marco conceptual

Un proceso de negocio es un conjunto de actividades que se realizan en coordinación en un ambiente organizacional y técnico [4].

La gestión de procesos de negocio (BPM) incluye conceptos, métodos y técnicas para soportar el diseño, administración, configuración, representación y análisis de los procesos de negocio. Su objetivo fundamental es representar el proceso de negocio con sus actividades y las restricciones de ejecución que se aplican sobre las actividades [4].

Los sistemas de gestión de procesos de negocios (BMPS) son programas de software utilizados como herramienta para representar y coordinar las actividades involucradas en un proceso de negocio.

BPMS además contempla soporte para interacción humana e integración de aplicaciones, y es aquí la diferencia fundamental con la tecnología de *workflow* [5] existente.

Las soluciones basadas en *workflows* se limitan a definir el flujo de actividades humanas o de documentos, y a partir de él obtener el seguimiento de los procesos. En estos casos si un participante del proceso requiere como parte de sus actividades ingresar datos en una aplicación, se debe abandonar temporariamente del ambiente del *workflow*, ejecutar la aplicación en cuestión y una vez finalizada su operación regresar al *workflow* y registrar el cambio de estado o finalización de la actividad. En BPM todas estas actividades están integradas en el mismo flujo, lo que resulta más natural y cómodo para un participante ya que cuenta con la facilidad de completar su actividad dentro del flujo BPM e inmediatamente los sistemas involucrados se actualizan.

A su vez, un “valor agregado” destacable de BPM es que ofrece una solución completa que abarca todo el ciclo de vida de un proceso de negocio: análisis, modelado, ejecución y monitoreo de los procesos.

En BPM el modelo del proceso se convierte en el núcleo de la implementación del proceso como solución tecnológica. El modelo del proceso de negocio diseñado por el área de negocios de una

empresa es “en sí” lo que se ejecuta sobre el “servidor de procesos” (el motor de BPM). Dicho en otras palabras: la “lógica principal del negocio” que bajo la tecnología tradicional se debe programar y desplegar sobre un “servidor de aplicaciones”, con BPM se reemplaza por un modelo que se despliega en el “servidor de procesos” con mucho menos intervención del área de tecnología (menos programación).

En la práctica una buena solución BPM debería poder ejecutar un proceso modelado por el área de negocios, sin la necesidad de que el área de tecnología de la organización tenga que programar una sola línea de código y así obtener como solución algo equivalente a un *workflow* tradicional. Luego el área de tecnología debería tomar este *workflow* e implementar los formularios de entrada (de interacción con usuarios) y los “servicios” (las actividades automatizadas) para completarlo en un flujo BPM.

Hacer que un modelo se convierta en un proceso ejecutable requiere de varias tecnologías habilitantes, cuando estas tecnologías se proveen juntas se la llama BPMS (Business Process Management Suite). Las componentes tecnológicas de esta suite son:

- Motores de Orquestación: coordinan la secuencia de actividades según los flujos y reglas del modelo de procesos.
- Herramientas de Análisis y *Business Intelligence*: analizan la información producto de la ejecución del proceso en tiempo real.
- Motores de Reglas: ejecutan reglas que permiten abstraer las políticas y decisiones de negocio de las aplicaciones subyacentes.
- Repositorios: mantienen los componentes y recursos de los procesos (definiciones, modelos, reglas) disponibles para su reutilización en múltiples procesos.
- Herramientas de Simulación y Optimización: permiten a los administradores del negocio comparar el nuevo diseño de procesos con el desempeño operacional actual.
- Herramientas de Integración: permiten integrar el modelo con otros sistemas, como los sistemas legados de la empresa.

BPM también es considerada una disciplina de administración, que requiere que las organizaciones cambien hacia un pensamiento centrado en los procesos y que reduzcan su dependencia de estructuras funcionales tradicionales. Es un enfoque estructurado que emplea métodos, políticas, métricas, prácticas de administración, y herramientas de software para mejorar la agilidad y el desempeño operacional [6].

En los últimos años, las técnicas de BPM han ayudado a reducir errores, minimizar costos e incrementar la productividad en las organizaciones,

que van desde las empresas de fabricación de productos pasando por proveedores de telecomunicaciones y empresas de seguros hasta departamentos gubernamentales.

b. BPM y SOA

BPM se complementa con otros enfoques que proveen mejoras en las organizaciones, en particular, se relaciona muy bien con la filosofía de diseño de software llamado SOA (Service Oriented Architecture).

BPM ayuda a los directores de una empresa u organización a adaptarse a los cambios continuos de los procesos de negocios, mientras que SOA ayuda a los arquitectos de software a crear sistemas empresariales que se adapten a los continuos cambios en los requerimientos.

Una de las claves del éxito de BPM es que provee un conjunto de metodologías para el análisis, comprensión y documentación de los procesos de negocios. También cabe destacar que BPM es una especificación desarrollada por la OMG (Object Management Group) que es un organismo internacional sin fines de lucro encargado de desarrollar estándares de integración para un amplio rango de tecnologías [7].

2. LINEAS DE INVESTIGACION y DESARROLLO

Las líneas de investigación principales asociadas a este trabajo pueden sintetizarse en las arquitecturas orientadas a servicios (SOA) y la gestión de los procesos de negocios (BPM).

BPM es un tema altamente relevante desde el punto de vista práctico y ofrece nuevos desafíos tanto a desarrolladores como a científicos. El enfoque orientado a procesos explicita el conocimiento muchas veces oculto acerca de cómo, quién y cuándo se realizan las actividades de una organización. La gestión de los procesos de negocio permite cubrir todo el ciclo de vida de los procesos de negocios desde su modelado hasta su monitoreo, pasando por el ensamblado y el despliegue.

El actual interés en BPM es el combustible del área de tecnología para las arquitecturas orientadas a servicios

La idea principal de la orientación a servicios es capturar la funcionalidad del negocio más relevante y proveerla con el suficiente grado de detalle para que pueda ser consumida. Para favorecer dicho consumo, los servicios deben ser registrados y publicados, además de contar con interfaces bien definidas basadas en lenguajes estándares como XML.

Por otra parte, si bien BPM muchas veces se describe como una metodología despegada de la implementación, en la medida en que la solución BPM se ajuste mejor a la plataforma de desarrollo redunda en facilidades para la documentación, depuración y mantenimiento posterior. En este sentido las aplicaciones basadas en JAVA adhieren a metodologías de diseño y paradigmas orientados a

objetos, cuentan con un *framework* estándar y ampliamente usado para desarrollo de aplicaciones distribuidas como es JEE [8] y todos los *frameworks* arquitecturales basados en JEE que facilitan el desarrollo de aplicaciones distribuidas. De esta manera se garantiza modularidad, extensibilidad, adaptación a cambios y otras cuestiones relacionadas a los requerimientos no funcionales de aplicaciones distribuidas como es la seguridad, alta disponibilidad y tolerancia a fallas. Todos estos aspectos facilitan el desarrollo de aplicaciones basadas en procesos con orientación a servicios en tecnologías JAVA.

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo de este trabajo es evaluar algunos BPMS actuales, según un conjunto de métricas consideradas de interés a la hora de elegir una herramienta que le permitiría a la organización que la adopte, gestionar más adecuadamente los cambios en sus procesos y el impacto de los mismos en el área de tecnología.

Se definen a continuación los aspectos de comparación que se considerarán para cada una de las herramientas analizadas y la fundamentación de su elección:

a. **Requerimientos de hardware:** este aspecto describe las necesidades mínimas de hardware para instalación y uso de la herramienta. Su importancia radica en que permite dimensionar la disponibilidad de recursos a destinar por una organización para adherir a metodologías que mejorarían el desarrollo de su negocio.

b. **Licenciamiento:** este aspecto indica si el licenciamiento es de código fuente abierto o propietario. La decisión que la organización tome en este sentido tiene que estar alineada con la estrategia global de la empresa u organismo en este sentido ya que estaría en juego el mantenimiento y responsabilidad a futuro de la aplicación.

Las soluciones de código fuente abierto implican mayores libertades pero también obligan a identificar cómo se trabaja con tales soluciones y cómo se mantienen actualizadas las mismas, por otro lado soluciones propietarias implican la necesidad de contratos de mantenimientos que aseguren que los entornos se actualicen y no se vuelvan obsoletos o peligrosos desde el punto de vista de vulnerabilidades de seguridad.

c. **Soporte para todo el ciclo de vida de los procesos:** esta métrica mide las etapas del ciclo de vida y los roles de los actores en un proceso de negocio. Las etapas son: modelado, ensamblado, despliegue, monitoreo y el soporte de roles que faciliten la tarea del analista del negocio (para el modelado), del arquitecto de componentes (para el ensamblado) y del desarrollador (para el despliegue).

Es fundamental determinar si la herramienta soporta todas las fases y lo hace unificadamente,

sin necesidad de importación/exportación de archivos. El soporte unificado también garantiza una actualización consistente en todos los archivos fuentes teniendo en cuenta los cambios frecuentes en las múltiples fases.

En cuanto al soporte de roles es muy provechoso contar con múltiples visiones de un mismo proceso de negocio. Estas vistas garantizan que la funcionalidad disponible corresponda a las actividades propias del rol, delimitando las actividades habilitadas para cada caso.

d. **Soporte para SOA:** este aspecto evalúa en qué medida la herramienta brinda soporte para acoplarse a una arquitectura orientada a servicios. El soporte del lenguaje BPEL (Business Process Execution Language) [9] y la adhesión al estándar SCA (Service Component Architecture) [10] garantizan que los procesos de negocios se transformen en los consumidores de los servicios ofrecidos por SOA. Frecuentemente las organizaciones cuentan con un activo tecnológico desarrollado y estable que requiere ser integrado en términos de procesos de negocio. El soporte para SOA de la herramienta minimiza costos y es determinante a la hora de decidir incluir el paradigma BPM en la organización.

e. **Integración con IDEs JAVA:** esta característica mide la capacidad de integración con un IDE JAVA tanto de la comunidad de software libre como propietaria. Los entornos de desarrollo cuentan con características amigables, definición de múltiples visiones, editores contextuales, compiladores, depuradores y ejecutores integrados, así también como paletas de componentes gráficas para arrastrar-soltar, agilizan sin lugar a dudas la producción de software. Las herramientas analizadas pueden cubrir todas las etapas del ciclo de vida del software hasta su implementación. En este caso es importante medir en qué grado la herramienta BPM posee un entorno de desarrollo propio o la capacidad de integrarse alguno existente y popular dentro de la comunidad de desarrolladores.

Las herramientas evaluadas son las siguientes:

- JBPM de JBoss Enterprise SOA Platform <https://www.jboss.com/products/jbpm/>
- Bonita de Bull y OW2 Consortium <http://wiki.bonita.objectweb.org/xwiki/bin/view/Main/>
- GPA de Consist <http://www.consist.com.ar/>
- Oracle Business Process Analysis Suite 10g <http://www.oracle.com/technologies/soa/bpa-suite.html>
Oracle SOA Suite 10g <http://www.oracle.com/global/lad/technologies/soa/oa-suite.html>

	Hardware	Licenciamiento	Ciclo de vida		Soporte SOA	Integración con IDEs JAVA
JBPM	768 MB de RAM 800 MB de espacio en el disco rígido para la versión full (<i>platform</i>) o 220 Mb de espacio en disco rígido para la versión <i>standalone</i>	Lesser General Public License (LGPL)	Modelado	✓	Cuenta con componentes “ESB <i>service</i> ” dentro del diseñador de procesos para utilizar servicios que están disponibles en el ESB	Eclipse con plug-in http://www.eclipse-plugins.info/eclipse/plugin_details.jsp?id=1028 http://sourceforge.net/projects/jbpm/
			Ejecución/Despliegue	✓		
			Monitoreo	✓		
Bonita	Procesador de 1 GHZ es recomendado, con un mínimo de 512 de RAM	Lesser General Public License (LGPL)	Modelado	✓	No provee	Eclipse con plug-in http://forge.objectweb.org/project/showfiles.php?group_id=56
			Ejecución/Despliegue	✓		
			Monitoreo	✓		
GPA	Procesador: PC Pentium 233 MHz o superior Pentium III recomendado. 128 MB de RAM (o superior) Disco Rígido: 245 MB de espacio disponible	Licencia Consist-GPA	Modelado	✓	Posee un mecanismo propio para definir un adaptador y un componente para ejecutar un servicio web	No posee un entorno de desarrollo. Ofrece una plantilla de modelado sobre Microsoft Visio: http://office.microsoft.com/en-us/visio/default.aspx
			Ejecución/Despliegue	✓		
			Monitoreo	✓		
Oracle Business Process Analysis Suite 10g Oracle SOA Suite 10g	Procesador Intel Pentium IV 3.0 Ghz, 2 GB RAM, Tarjeta gráfica con resolución de 1024 x 768 con al menos 256 colores Red de alta velocidad (100 Mbit) entre el servidor de base de datos y Oracle business process repository.	OTN License Agreement	Modelado	✓ Orable BPA 10g	Utiliza como lenguaje de ejecución a BPEL.	JDeveloper 10g
			Ejecución/Despliegue	✓ JDeveloper 10g y Oracle SOA Suite 10g		
			Monitoreo			

Tabla 1. Comparación de herramientas

4. FORMACION DE RECURSOS HUMANOS

BPM es un modelo fundamental en la nueva ciencia de los servicios que se plantea como una línea de formación de recursos humanos necesarios para el área Tecnología Informática y de Comunicaciones, que en los últimos años evolucionó decididamente del producto a soluciones y servicios. El presente proyecto muestra una línea de trabajo que se inició el año pasado y donde se están formando alumnos para desarrollar su tesina e interactuar con docentes investigadores formados que están incorporando BPM como línea de acción.

5. BIBLIOGRAFIA

[1] Patricia Bazán (2007) "BPEL: una propuesta para el uso de Web Services". CACIC 2007

[2] Nayna Patel y Vlatka Hlupic "Dynamic Business Process Modelling (BPM) For Business Process Change". Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex.

[3] Definición de BPM:
http://es.wikipedia.org/wiki/Business_Process_Management

[4] Weske Mathias (2008) Business Process Management: Concepts, Languages, Architectures. Springer, 3-67.

[5] Definición de Workflow
<http://en.wikipedia.org/wiki/Workflow>

[6] BPM Group Snoop Consulting (2008) "BPM Group". <http://crm.snoop.com.ar/tikiwiki/tiki-index.php?page=BPMGroup>

[7] Object Management Group (OMG) "Business Process Management with OMG specifications".
<http://www.omg.org>

[8] Sitio Oficial de JEE:
<http://java.sun.com/javaee/technologies/javaee5.jsp>

[9] Definición de BPEL
<http://es.wikipedia.org/wiki/BPEL>

[10] Definición de SCA
http://en.wikipedia.org/wiki/Service_component_architecture

TRIPTICO DE LA INGENIERÍA DEL SOFTWARE. ANÁLISIS DESARROLLO Y APLICACIONES DE METODOLOGÍA RAISE

José H. Paganini, Sebastián M. Figueroa Juan C. Rodriguez, Hector P. Liberatori, José Eduardo Huanca, Nancy Cholele, Delia C. Cruz. Centro de Investigaciones Básicas y Aplicadas (CIBA) Facultad de Ingeniería Universidad Nacional de Jujuy UNJu

CONTEXTO

Este proyecto constituye un avance en la Argentina de la aplicación de la metodología RAISE, a tal punto que casi no hay antecedentes de aplicaciones similares en el país

A su vez es la investigación en aplicaciones y desarrollos que puedan enriquecer la Metodología RAISE con aportes desde lógicas deónticas, modales y paraconsistentes junto con otros provenientes del análisis funcional, para aumentar el espectro de formalización de dominios, entidades y procedimientos en el campo de las Ciencias de la Computación, genera una contribución al desarrollo de la informática, en su finalidad de obtener sistemas más confiables. En ese sentido la Facultad de Ingeniería de la UNJu, a través del CIBA, con este grupo de I/D contribuye a la investigación en temas de la Ingeniería del Software.

RESUMEN

El Tríptico de la Ingeniería del Software, compuesto por la Ingeniería del Dominio, Ingeniería de Requisitos y el Diseño del Software, tiene como fase más importante a la primera de ellas, que especifica el espacio donde será eventualmente aplicado un producto software particular.

Especificar según la metodología RAISE es expresar mediante formulaciones lógico matemáticas las entidades y sus relaciones en un dominio; logrando una sintaxis precisa y una semántica única para ello se emplea el lenguaje formal RSL (RAISE Specification Language)

El presente proyecto encara, por un lado, el análisis de la metodología RAISE, su vinculación a los procesos de Verificación y Validación del software y eventuales extensiones de la metodología a campos de la Lógica Modal y teoría de Probabilidades. Por otro se realizarán aplicaciones a casos concretos de diseños efectuados bajo el paradigma del Tríptico, con énfasis en la Ingeniería del Dominio

Palabras clave: RAISE, RSL, LÓGICA TRIPTICO DEL SOFTWARE

1. INTRODUCCION

El diseño de un producto software debe ser acorde a un conjunto de requisitos que debe satisfacer; estos requisitos a su vez están enmarcados, por decirlo así, en un dominio que se estableció para fijar su alcance

y visión, más específicamente el dominio en que el software debe servir. Entonces el proceso que desemboca en el diseño posee tres componentes, el establecer el dominio, que se lo denota con Ingeniería del Dominio; el establecimiento de los requisitos, conocido como Ingeniería de requisitos, y por último, el diseño del software emergente.

De esta forma, estos tres elementos, Ingeniería del Dominio, Ingeniería de Requisitos y el Diseño del Software constituyen un Tríptico. Concepto este, establecido por Dines Bjørner; y expuesto en [7], [8], [12], [13] y [14].

Este enfoque constituye un nuevo paradigma en la manera de producir productos software, bajo el cual se están elaborando proyectos software, principalmente, en la Comunidad Europea como el AMORE, que organiza la gestión de las distintas líneas férreas de los países de la Comunidad, y también aplicaciones a sistemas de control de vías férreas en China; junto con las diversas aplicaciones generadas en el UNU/ IIST. (United Nations University International Institute of Software Technology) sito en Macao, República Popular China.

Este nuevo paradigma, establece como necesario, como paso inicial y fundamental, la descripción formal del Dominio y de los Requerimientos, para poner en claro, sin las ambigüedades del lenguaje coloquial, de que se habla, a que se refiere cuando se designa algo, y el establecimiento de los requerimientos. Entonces, para la descripción formal del dominio y los requerimientos se emplea el Método RAISE, con un lenguaje de especificación, riguroso basado en la Lógica y Matemática, que es el RSL (RAISE Specification Language).

Los lineamientos básicos, del método formal y del lenguaje, se encuentran en los trabajos del investigador británico Christopher William George y un grupo de colaboradores (conocido como RAISE Group) [6] y [9].

Esto también supone para el grupo de I/D el estudio y análisis de sistemas formales, leyes de la lógica y matemática [1], [2], [3], [4] y [5].

Para estructurar la fase de Diseño del Software con las fases anteriores del Tríptico, se acude al empleo de herramientas software (EMACS y otras), desarrolladas en el UNU/IIST, que por un lado depuran las formalizaciones y establecen condiciones de confiabilidad de las mismas; y por otro avanzan a una etapa de implementación

produciendo código en lenguajes aplicativos cómo consecuencia lógica de las formalizaciones en RSL. Con esta fase culmina la aplicación del paradigma del Tríptico a través de esta metodología que parte de una especificación formal y termina en un código en un lenguaje aplicativo

Esta metodología constituye un aporte más, entre muchos otros, a la Ingeniería del Software, constituyéndose esta última en una de las ramas de la Informática que mayor crecimiento e importancia ha adquirido en los últimos años.

En efecto siendo que el desarrollo de sistemas es uno de los aspectos económicamente más importante de la Informática y que dicho desarrollo es crucial para las aplicaciones informáticas que tiende a cerrar la brecha entre las crecientes necesidades de desarrollo y las prestaciones que implican velocidad en producir y llegar a tiempo al mercado con nuevos productos, garantizando la seguridad y confiabilidad de los desarrollos, ajustándose a las previsiones de costo, y que la Ingeniería del Software es la rama de la Informática que trata con las herramientas, métodos y principios necesarios para el desarrollo de sistemas, cuya importancia no puede negarse.

1.4.1. Objetivos Generales y Particulares

El proyecto tiene como objetivos generales, en el marco de la Ingeniería del Software, los siguientes:

- I. El estudio, investigación y desarrollo de diversas técnicas y herramientas, dentro de la metodología RAISE que sean útiles en el campo del desarrollo de sistemas.
- II. La investigación básica en Lógica y Matemática para el enriquecimiento de la Metodología RAISE y de los Métodos Formales.
- III. La preparación de estudiantes de grado y tesis de postgrado en estas técnicas.

Los objetivos particulares son los siguientes:

- I. El ensayo de herramientas software para la metodología RAISE.
- II. El eventual adaptación y/o diseño de herramientas software para la metodología RAISE
- III. La aplicación concreta de diseño de un software complejo con la metodología RAISE y bajo el paradigma del Tríptico para el Control de Gestión de la Logística en el Corredor Bioceánico (Atlántico Pacífico) del Mercosur, que pasa por la Provincia de Jujuy en la Argentina.

1.4.2. Metodología.

Como el presente es un proyecto de investigación que presenta los aspectos fundamentales siguientes:

1. Estudio y Análisis del Método RAISE y de su aplicación en el Paradigma del Tríptico.
2. Análisis e Investigación de modificaciones y/o herramientas de aplicación al Método RAISE.

3. Aplicaciones de Diseño Software a casos concretos con metodología RAISE y bajo el Paradigma del Tríptico,

Entonces dado que los dos primeros aspectos tienen un sesgo teórico, la metodología a que se emplea es la de estudiar bibliografía, asistir a cursos sobre la temática y el tratar de producir resultados y publicaciones que contribuyan a difundir y/o perfeccionar esta técnica,

Respecto al tercer aspecto, el grupo de I/D decidió establecer que la aplicación se realizaría sobre la Logística del Corredor Bioceánico (Atlántico Pacífico) del Mercosur; y que esta etapa empezaría en Julio de 2009. Por lo consiguiente en esa fecha se empezaría recopilando información e iniciando los contactos con entidades gubernamentales que ejercen el control en la logística de esa área geográfica.

2. LINEAS DE INVESTIGACION y DESARROLLO

Esta línea de investigación pretende moverse sobre el eje de hallar parámetros formales que establezcan condiciones cada vez más confiables para el diseño de productos software.

Dado que el software es una creación humana, es una aproximación a la realidad bajo una visión particular; el desarrollo de productos es acorde a una cultura modelada de acuerdo a un conjunto de paradigmas.

Entonces software de una forma particular, es un problema cultural, la implementación de un diseño de un producto software con un proceso riguroso metódico validable y verificable, con toda la documentación pertinente y necesaria, con el dominio de aplicación y los requerimientos estrictamente formalizados, constituye la adopción de nuevos paradigmas que producen un cambio cultural.

Este cambio cultural, reviste caracteres de importancia si es aplicado como norma de adquisición de software en reparticiones del estado provincial; elevando, en consecuencia el estándar de calidad de desempeño de estos productos, y contribuye a fijar criterios para elaboración de disposiciones y leyes regulatorias en el mercado de los sistemas informáticos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados obtenidos concuerdan con los esperados y a esto hay que sumarle el hecho de que en la Facultad de Ingeniería de la UNJu se dicta un Maestría en Ingeniería del Software de la UNSL, en donde docentes de este proyecto colaboran en el dictado contribuyendo a la formación de alumnos de maestría y especialidad. Además se tramitan becas internacionales para estudiar esta temática en el International Insntitute for Software Technology, de las Naciones Unidas UNU/IIST, sito en Macao

República Popular China. Exponer los resultados alcanzados y los objetivos en curso.

4. FORMACION DE RECURSOS HUMANOS

En la formación de recursos humanos, con trabajos contributivos a la investigación dentro de la temática del proyecto, se destacan los siguientes:

1. Tesis de Maestría, en la Universidad Nacional de San Luis UNSL “Dominio de un Sistema de Viaje” de José Humberto Paganini
2. Tesis de Maestría, en la Universidad Nacional de San Luis UNSL “Especificación de requerimientos en un Sistema de Viaje” de Sebastián Marcos Figueroa.
3. Proyecto final de carrera de Ingeniería Informática de la Universidad Nacional de Jujuy, de José Eduardo Huanca.
4. Proyecto final de carrera de Ingeniería Informática de la Universidad Nacional de Jujuy, de Nancy Cholele.
5. Proyecto final de carrera de Ingeniería Informática de la Universidad Nacional de Jujuy, de Delia Cristina Cruz

5. BIBLIOGRAFIA

1. Lógica para Matemáticos. Hamilton, A. G.. Paraninfo. Madrid. 1981.
2. Foundations of Logic Programming. Second extend Edition. Lloyd, J. W. Springer Verlag New York. 1987
3. Fundamentos de Lógica Computacional. Juan Fausto Solís. Gildardo Sánchez Ante. Editorial Trillas. México. ISBN 968-24-6100-6, año de edición 2000.
4. Introducción a la Lógica Matemática. P. Suppes; S. Hill. Editorial Reverté. México. ISBN 968-6708-01-4, año de edición 1999.
5. Introducción a la Lógica. Inving M. Copi. Eudeba manuales. Argentina. ISBN 950-23-0040-8, año de edición 1984.
6. The RAISE Specification Language. Chris George; P. Haff; K. Havenlund; A. E. Haxthausen; R. Milne; C. B. Nielsen; S. Prehn; K. R. Wagner. Prentice Hall. UK., año de edición 1996.
7. Software Engineering A New Approach. Dines Bjøner.. Technical University of Denmark. Dinamarca., año de edición 2002.
8. The SE Book (Software Engineering Book) Dines Bjøner. Technical University of Denmark. Dinamarca., año de edición 2004.
9. The RAISE Development Method. Chris George. Springer Verlag. London 2000
10. Decision Support Systems in the Twenty-First Century. George Marakas. Prentice Hall. USA ISBN 0-13-744186-X, año de edición 2000.
11. Fundamentals of Software Engineering. C. Ghezzi; M. Jazayeri. D. Mandrioli. Prentice Hall. USA. ISBN -0-13-820432-2, año de edición 1999
12. Software Engineering 1 Abstraction and Modeling. Dines Bjøner. SPRINGER 2006 ISBN 10-3540211497
13. Software Engineering 2 Specification System and Languages. Dines Bjøner. SPRINGER 2006 ISBN 10-3540211300
14. Software Engineering 3 Domain Requirements and Software Design. Dines Bjøner. SPRINGER 2006 ISBN 10-3540211518

Imputación de datos con redes neuronales

María E. Valesani
Mgter. en Informática y Computación
Profesor Adjunto
evalesani@exa.unne.edu.ar

Osvaldo P. Quintana
Experto en Estadística y Computación
Docente
oquin@mecon.indec.gov.ar

Oscar A. Vallejos
Mgter. en Informática y Computación
Profesor Adjunto
ovallejos@exa.unne.edu.ar

Universidad Nacional del Nordeste
Facultad de Ciencias Exactas y Naturales y Agrimensura
Dpto. de Informática
9 de Julio 1449
(3400) Corrientes, Argentina
(03783)-15-679884; (03783)-15-405117

Contexto

Miembros estudiantes de doctorado en Ingeniería de Software de la Universidad de Málaga (España) en convenio con las Universidades del Nordeste y Universidad de Misiones.

Resumen

El presente trabajo tiene por objeto la aplicación de Redes Neuronales Artificiales (RNA) como métodos de imputación, para ser utilizados sobre una base de datos real. Donde se simulo perdida de datos en distintos porcentajes, aplicando la técnica MCR (*Missing completely at random*). Estos datos faltantes o perdidos se completan mediante la aplicación de distintos modelos y en distintas situaciones, con el propósito de valorar el comportamiento de los mismos a través de distintos parámetros de eficiencias como MAE, MSE, y Regresión, se pretende determinar si RNA brinda una herramienta adecuada para la imputación en este caso en particular aplicados a datos de Censos Ganaderos.

Palabras clave: Imputación de datos, Redes neuronales artificiales, perceptrones multicapa, aprendizaje supervisado, imputación de datos en ganadería.

1. Imputación con redes Neuronales

En los últimos años también se ha abordado el problema de los datos faltantes mediante redes neuronales artificiales (RNA). Las RNA se definen como un sistema de procesamiento de información, formado por un conjunto de unidades simples o procesadores organizadas en paralelo, que operan sólo con la información disponible localmente que reciben a través de las conexiones con otras unidades por las que fluye información de tipo numérico [5]. Una tipología de RNA que se emplea habitualmente en la generación de modelos de clasificación y predicción son las denominadas RNA supervisadas, entre las cuales destacan, tanto por el número de trabajos que las utilizan como por su amplia aplicabilidad, las redes perceptrón multicapa (MLP), consideradas aproximadores universales de funciones [4]. No nos detendremos aquí a exponer los detalles de la arquitectura de este tipo de RNA, ya que se trata de una línea de investigación consolidada.

Generalmente los datos recopilados en censos y/o encuestas son procesados mediante técnicas estadísticas tradicionales. El inconveniente frecuente se presenta debido a la falta de respuesta por parte de los entrevistados. Se pensó que esta técnica de la IA presentaba una alternativa al momento de proporcionar imputación y/o predicciones sobre la existencia de ganado vacuno por medio de imputaciones anteriores en variables auxiliares, a partir de otros métodos de imputación, un archivo histórico u otros medios como opiniones de expertos temáticos.

Los datos empleados en el presente trabajo corresponden a los registrados para la provincia de Corrientes durante el Censo Nacional Agropecuario 2002.

En este trabajo, se describe el comportamiento de modelos de RNA supervisadas aplicadas en la imputación de Cabezas de Ganado Vacuno, partiendo como entrada a la red con los datos de Superficie dedicada a la Ganadería y el coeficiente de Hectáreas/Equivalente Vaca.

2. Líneas de investigación y desarrollo

La línea de investigación es imputación de datos con técnicas tradicionales [3] y no tradicionales como operadores de agregación de la mayoría [6] y RNA.

2.1. Descripción del experimento

Se describe la metodología aplicada en el diseño, desarrollo y entrenamiento de modelos de RNA supervisadas. A continuación se sintetizan las etapas consideradas:

Formulación del problema. La formulación del problema se concretó mediante la aplicación, en base a dos variables auxiliares, Superficie Dedicada a la Ganadería, y la relación de la superficie y el Equivalente Vaca Ha/EV, la incorporación de esta última variable auxiliar surgió como consecuencia de una entrevista con especialista de la temática específica, quien sugirió que actualmente se emplea como indicador importante el *equivalente vaca* que consiste en determinar la capacidad que tiene un campo para la ingesta de materia seca como alimentación del rodeo y la misma podría ser utilizada para predecir el número de ganado, de acuerdo a la zona donde se encuentre la explotación agropecuaria. Cabe aclarar que esta elección de variables evidenciales o variables de entrada se fundamenta en la no respuestas por diversos motivos de parte de los productores del sector a censos y encuestas específicas.

Selección de las variables evidenciales. Las variables relevantes requeridas para predecir el número de cabezas de ganado fueron obtenidas en base al conocimiento de

técnicos. Los valores que asumen las variables evidenciales corresponden a registros del Censo Nacional Agropecuario 2002.

Estudio para la implementación de una RNA: Se realizó un estudio referente a conceptos relacionados con la topología de una red neuronal, dimensionamiento de la red, arquitectura de una RNA, algoritmos de aprendizaje. Se seleccionó el aprendizaje supervisado por considerarse el más adecuado al problema y el que mayor validez tiene en amplios dominios del conocimiento.

Selección de las herramientas informáticas. Esta etapa se concretó mediante la elección de un toolbook para modelizar sistemas RNA. Estos toolbooks o herramientas de software, son de carácter general, motivo por el cual deben evaluarse a fin de verificar que las opciones disponibles son adecuadas para el tratamiento de un problema en particular. Se examinaron las aplicaciones para la creación y tratamiento de modelos de redes neuronales artificiales disponibles. Los toolbooks o herramientas existentes están destinadas a la predicción y/o clasificación. En la selección de las mismas se mencionan características como: la velocidad de ejecución, los requerimientos del sistema, las interfaces, el máximo número de variables y capas a especificar. Se seleccionó nntool de Matlab 7.4. Esta herramienta construye un modelo en base a los datos. Permite especificar diversos parámetros de aprendizaje. La conjunción de estas características se empleó para la definición de distintos modelos de entrenamiento de la RNA.

Construcción de modelos de RNA

a. **Arquitectura.** El formato del archivo para registrar los datos recolectados depende del modo de funcionamiento del software de RNA seleccionado. En primer lugar para crear los conjuntos de datos de entrenamiento y comprobación se diseñó el patrón (ejemplar o ejemplo). Cada patrón consta de dos partes. La primera parte es, un conjunto de números que representan los valores de las variables de entrada o evidenciales, empleadas para estimar los resultados. Si hay m variables predictoras, los primeros n elementos de cada patrón del archivo de datos de entrenamiento, segunda parte, la sección del "criterio" o "resultado", que consta de uno o más números, cada uno representa los valores de las variables de salida, es decir, los valores observados y que deberá predecir el modelo de RNA.

b. Definición de nodos de entrada y nodos de salida. Las variables evidenciales, SG = Hectáreas dedicada a la ganadería por cada explotación agropecuaria y Ha/EV = Hectáreas sobre Equivalente Vaca, se definieron como nodos de entrada y CG = el número de cabezas de ganado registrados en el Censo 2002 como valores esperados del nodo de salida.

c. **Diseño de los archivos de datos.** Se dividió el conjunto de datos en dos partes, una destinada al entrenamiento 75% y otros datos reservados para la comprobación de los modelos de RNA, 25%.

d. **Entrenamiento de los modelos.** Consiste en la definición de la topología de la RNA. No existe un procedimiento específico para definir *a priori* del número de capas ocultas y el número de neuronas por capa necesarios para lograr el aprendizaje de la RNA. Se propusieron distintas configuraciones modificando el número de nodos y capas intermedias, funciones de activación y parámetros.

e. **Ejecución de los modelos,** se entrenaron distintos modelos de RNA utilizando el software nntool de Matlab 7.4.

f. **Validación de los modelos.** En el aprendizaje supervisado, una medida de calidad del modelo está dada en términos de los valores del error estándar MSE y el MAE [2] y la regresión entre salida de la red y salida real que proporciona la herramienta utilizada para el modelado de la red.

Validación del software. Finalizado el desarrollo de los modelos, es imprescindible verificar el correcto funcionamiento del mismo. Se deben implementar validaciones internas y validaciones externas. Las primeras fueron realizadas por los autores del trabajo. Las segundas se llevarán a cabo con especialistas en la temática.

3. Resultados obtenidos

A partir de los datos proporcionados por el Censo Nacional Agropecuario 2002, se buscó modelizar RNA para imputar la cantidad de cabezas de ganado de una explotación a partir de los datos de Superficie Total de la explotación Agropecuaria, Superficie Dedicada a la Ganadería y Estrato de tamaño entre otras variables disponibles, se descartaron muchos modelos entrenados, por considerar que los resultados obtenidos no fueron aceptables en función a los valores de los parámetros utilizados para medir su eficiencia como el error cuadrático medio, el error absoluto y regresión.

Posteriormente y ante la consulta con expertos en el tema se considero finalmente como variable evidenciales a la Hectáreas Dedicada a la Ganadería (SG) y el Coeficiente Hectárea sobre Equivalente Vaca (Ha/EV), consiguiendo con este modelo una RNA que brindo restados muy aceptables para la imputación de Cabezas de Ganados (CG).

Entrada

Superficie dedicada a la ganadería = SG
Hectárea por Equivalente Vaca = Ha/EV

Salida

Cantidad de Cabezas de Ganado = CG

Se efectuó en principio un análisis exploratorio de los datos mediante el BoxPot y se descartaron los datos fuera del rango para el ítems relación Ha/EV donde se separo alrededor de 100 establecimiento, que representaba el 0,22 % del total de cabezas, Se verifico si existía algún de relación entre las variables de entradas y se observo que no estaban correlacionadas ya que el coeficiente de Pearson tenia el valor de 0.078, prácticamente igual a cero, se realizo la misma comprobación entre la variable de entradas SG, Ha/EV y la salida CG, encontrando que la SG tenia un relación alta que rondaba el valor de 0.90 de mismo coeficiente, mientras que la variable de salida con la relación Ha./EV tenia una relación baja.

Se experimentó con redes Multicapas con distintas estructuras de capas, con 1,2 y 3 capas intermedias, distintas funciones de transferencias y aprendizajes supervisados de retropropagación, con funciones varias funciones de aprendizaje, con datos reales sin transformaciones y posteriormente con la normalización de los datos con sus mínimos y máximos, hasta encontrar una red que con tres capas intermedia de topología 2-14-8-4-1 con función de aprendizaje LEVENBERG –MAQUARD y la función Logística como función de transferencias para todas las capas, la cuál brindo resultados satisfactorios para la etapa de entrenamiento con el 75% de los datos consiguiéndose una buena convergencia del error como puede observarse en la Figura 1, aplicando la misma al 25% restante que se reservó para realizar el test de comprobación, al cual se aplico una medida de regresión entre la salida real y la obtenida por la RNA , consiguiendo un valor de $R= 0.982$, que se considera muy aceptable en términos estadísticos, como se observa en la Figura 2.

4. Conclusiones

La imputación mediante RNA se diferencia fundamentalmente lo métodos estadísticos tradicionales, en que los primeros no realizan condicionamiento, ni ninguna hipótesis sobre

La distribución de los datos a estudiar y que en este caso en particular sobre la imputación de Cabezas de Ganado Vacuno brinda una herramienta que merece ser tenida en cuenta para el tratamiento de datos faltante como método de imputación en Encuestas y Censos del sector, principalmente si se cuenta con las variables auxiliares necesarias para su aplicación, de las cuales van a depender directamente la imputación a realizar, las que se pueden obtener por

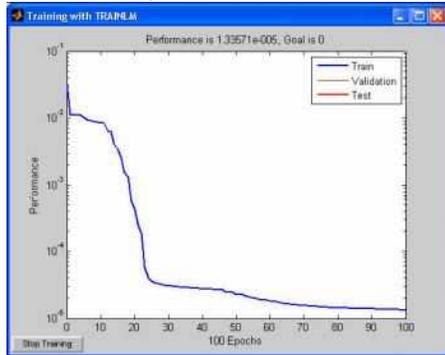


Figura 1: Error de Entrenamiento de la Red (100 Epochs)

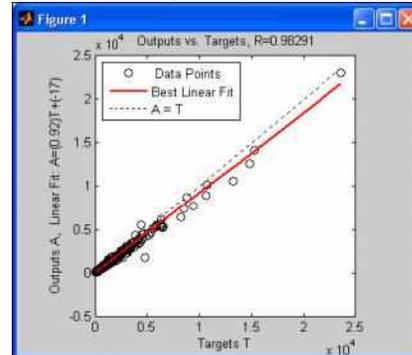


Figura 2: Comparación salida de la RNA y salida Real

medio de trabajos realizados con anterioridad, por consultas con expertos, mediante estudio de cartografía digitales con fotointerpretación de las explotaciones, material del cual se dispone en la provincia donde se realizó el estudio o por imputación con otros métodos como imputación multitarea con redes neuronales [1] siendo la misma una línea de futuros trabajos, con el fin de poder contar con datos más precisos sobre el tema tratado en este caso, ya que representa de sumo interés en la actualidad debido a la preponderancia que tienen los productos de origen primarios en el sistema económico nacional y global.

5. Referencias

- [1] Caruana R. (1993), Multitask learning: a knowledge-based source of inductive bias. Proceedings of the 10th International Conference of Cognitive Science, pp. 41-48.
- [2] Castillo, E., Cobo, A.; Gutiérrez, J. M. y Pruneda, R. E. (1999). "Introducción a las redes funcionales con aplicaciones. Un nuevo paradigma funcional". Ed. Paraninfo.
- [3] Doña J.M., Quintana O.P., Valesani M.E., Vallejos O.A. (2008) "Analysis of Aggregation Methods in Incomplete Database Systems". Information Processing and Management of Uncertainty in Knowledge-Based System (IPMU 2008). ISBN: 978-84-612-3061-7
- [4] Hornik, K., Stinchcombe, M. y White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks, 3, 551-560.
- [5] Rumelhart, D.E. y McClelland, J.L. (Eds.). (1992). Introducción al procesamiento distribuido en paralelo (García, J.A., Trad.). Madrid: Alianza Editorial. (Traducción del original Paralell distributed processing, 1986).
- [6] Vallejos, Oscar A, Valesani, Maria E, Quintana, Osvaldo (2008) "Imputación de datos con operadores OWA de la mayoría". X Workshop de Investigación en Ciencias de la Computación Gral Pico La Pampa Argentina 5 Y 6 de Mayo 2008 Org Universidad Nacional de La Pampa - Facultad de Ingeniería.

Diseño de un Profile para el Modelado de Aplicaciones Paralelas y Concurrentes

Universidad Nacional de La Matanza
Departamento de Ingeniería e Investigaciones Tecnológicas
Instituto de Investigación y Desarrollo
Florencio Varela 1704. San Justo, Buenos Aires, Argentina

Daniel A. Giulianelli
dgiulian@unlam.edu.ar

Rocío A. Rodríguez
rodri@unlam.edu.ar

Pablo M. Vera
pablovera@unlam.edu.ar

Victor M. Fernandez
vfernandez@unlam.com.ar

Contexto

Este artículo se basa en el trabajo realizado por el equipo de investigación de MOA (modelado de aplicaciones) de La Universidad Nacional de La Matanza. Este equipo trabaja en el desarrollo de extensiones, es decir nuevos artefactos UML (Lenguaje Unificado de Modelado) para la mejor representación de problemáticas de aplicaciones paralelas y concurrentes.

Resumen

UML es un lenguaje de representación de amplio propósito general, el cual para tal fin cuenta con vocabulario gráfico. En algunos casos cuando se quiere modelar un tipo de aplicación particular el vocabulario gráfico de UML resulta ser muy reducido. Por esta razón es necesario extender el lenguaje con nuevos artefactos que permitan modelar las características particulares del dominio en cuestión.

El equipo de investigación trabaja con aplicaciones paralelas las cuales poseen procesos concurrentes y distribuidos. Al intentar modelar este tipo de aplicaciones con UML surge la necesidad de crear nuevos artefactos, los cuales deben ser especificados formalmente para lo cual se genera un Profile.

Palabras clave: UML, Dominio, Artefactos, Extensiones, Aplicaciones, Procesos, Paralelos, Concurrentes, Distribuidos, Recursos, Profile, Meta-Modelado.

1. Introducción

OMG (Object Management Group) ha creado perfiles como solución para la falta de

elementos de representación de dominios específicos. Un profile es un mecanismo para extender un lenguaje a fin de expresar conceptos más específicos de ciertos dominios de aplicaciones. Según OMG “un profile es un subconjunto del meta-modelado de UML, este subconjunto del meta-modelado determina las reglas para representar a este subconjunto del UML” [3]. Para aclarar la definición es posible considerar el siguiente ejemplo: Si a través de UML se modela el diagrama de actividades con un diagrama de clases, este diagrama de clases que representa características del diagrama de actividades es un meta-modelado de UML, un profile es un subconjunto de este meta-modelado.

Al momento de modelar para dominios específicos surgen problemas en la expresividad de UML, por lo que es necesaria la extensión del lenguaje. Esta extensión del lenguaje permitirá crear nuevos artefactos dedicados a una tarea o con un significado determinado en el dominio de la aplicación y por consiguiente permitirá modelar aquellas características que no eran contempladas por la concepción original de UML. Por esta razón UML nos provee un mecanismo de extensibilidad para poder ampliar el vocabulario, estos mecanismos se encuadran dentro de la definición de los perfiles:

- Estereotipos: Permiten la creación de nuevos tipos de bloques de construcción que derivan de otros existentes pero no son específicos de un problema particular. Estos son definidos por un nombre y un grupo de elementos del meta-modelado. Los estereotipos representan una nueva característica agregada al UML para extender el lenguaje.

- **Valores Etiquetados:** Los valores etiquetados son propiedades nuevas para elementos existentes, estos son meta-atributos que son asociados a una meta-clase de un meta-modelado extendido del profile. Cada valor etiquetado tiene un tipo y es asociado a un estereotipo.
- **Limitaciones o Restricciones:** Forman reglas (de consistencia o de negocios) sobre los elementos y sus propiedades. Las limitaciones son asociadas a los estereotipos, imponen condiciones a los elementos del meta-modelado que fueron estereotipados. Las limitaciones son escritas en un lenguaje natural denominado OCL (Object Constraint Language).

En el sitio de OMG es posible encontrar los siguientes ejemplos de Profiles extendidos para propósitos particulares [2]:

- **UML Profile for CORBA ®:** Permite expresar la semántica de CORBA a través de herramientas UML.
- **UML Profile for Data Distribution:** El propósito de este Profile UML es brindar soporte al análisis y diseño de sistemas orientados a objetos usando servicios de datos distribuidos.
- **UML Testing Profile:** Define un lenguaje para diseño, visualización, especificación, análisis, construcción y documentación de artefactos de testing de sistemas. Este profile es un lenguaje de modelado de testing que puede ser usado con la mayoría de los objetos y componentes y a la vez ser aplicado al testing de sistemas de varios dominios de aplicación. El UML Testing profile puede ser usado de manera autónoma para el manejo de artefactos de test o ser integrado a UML para el manejo y artefactos de test juntos.
- **UML Profile for Enterprise Application Integration (EAI):** El objetivo de este profile es simplificar la integración de aplicaciones estandarizando los metadatos de invocación y translación de información entre aplicaciones.
- **UML Profile for System on a Chip (SoC):** Esta especificación define un profile para sistemas sobre chip, permite la

representación jerárquica de módulos y canales, los cuales son elementos fundamentales de los sistemas sobre chips.

2. Línea de Investigación y Desarrollo

Bajo la línea inicial de investigación que consistió en analizar las extensiones que resultan necesarias para modelar aplicaciones con PRocesos COncurrentes y DIstribuidos se realiza un vocabulario gráfico que conforma un lenguaje al que se denomina PROCODI. Luego se desarrolla un profile que permita definir formalmente al nuevo lenguaje. PROCODI se basa en UML e incorpora nomenclatura para:

- Acceso a los recursos compartidos (Semáforos y Monitores).
- Diferencia entre mensajes y RPC (Remote procedure Call – Llamada a procedimientos Remotos).
- Identificación de área de memoria compartida.
- Cardinalidad para tareas que se harán en forma idéntica en distintos threads.
- Timer para aquellos casos en que las acciones repetitivas tienen un tiempo establecido de repetición.
- Diferenciación entre RPC Sincrónicos y Asincrónicos.
- Indicar el uso de cola de mensajes
- Identificar procesos que requieran independencia en la ejecución paralela de manera de resaltar claramente aquellos procesos cuya ejecución no puede ser segmentada en distintos procesadores.
- Identificar web services.

Tomando los elementos propios del diagrama de actividades es posible hacer una adaptación la cual permitirá:

- Diferenciar nodos mediante las calles del Diagrama de Actividades.
- Indicar subcalles para distinguir hilos.
- Establecer tipo de conexión entre nodos.

Para la definición de este nuevo lenguaje (sub-lenguaje basado en UML el cual extiende al mismo añadiéndole expresividad para este dominio en particular) es necesario

generar un profile. Los pasos para la construcción del profile consisten en definir:

1) Un meta modelado del dominio de la aplicación: Si no existe, entonces es necesario definirlo utilizando los mecanismos del propio UML [1]. Para lo cual habrá que incluir la definición de las entidades propias del dominio, las relaciones entre ellas, así como las restricciones que limitan el uso de estas entidades y de sus relaciones.

2) El profile: Para lo cual un estereotipo debe ser creado por cada elemento del meta modelado. Es conveniente que los estereotipos tengan el mismo nombre que los elementos del meta-modelado.

Estableciéndose de esa forma una relación entre el meta modelado y el profile. En principio cualquier elemento que se necesite para definir el meta modelado puede ser etiquetado posteriormente con un estereotipo.

3) Los elementos UML que van a ser extendidos para cada estereotipo: Es importante tener en claro cuales son los elementos del meta modelado de UML que se están extendiendo sobre los que es posible aplicar un estereotipo. Ejemplos de estos son: las clases, sus asociaciones, sus atributos, las operaciones, las transiciones etc. De esta forma cada estereotipo se aplicará a la meta-clase de UML que se utilizó en el meta-modelo del dominio para definir un concepto o una relación.

4) Los valores etiquetados. En este paso los valores etiquetados deben ser agregados para cada atributo del meta modelado que es asociado al profile.

5) Las limitaciones del profile, a partir de las restricciones del dominio: Esto se realiza mediante OCL (Lenguaje de restricción de objetos) [6].

3. Resultados Obtenidos/Esperados

En este apartado se muestra el resultado del seguimiento de los pasos para la construcción del profile de PROCODI:

1) Se debe identificar los elementos que forman la extensión del lenguaje UML para poder determinar el meta-modelado del mismo. Los componentes introducidos por PROCODI son los siguientes:

- Semáforos y Monitores: Cuando se utilizan hilos es muy posible que existan recursos que se deben compartir y por lo tanto es necesario administrar su acceso, ya que solamente un hilo puede utilizarlo en un momento dado.
- Cardinalidad: Esta construcción se propone para el caso de que varios hilos realicen una operación idéntica.
- Timer: Extendiendo la cardinalidad se puede realizar una construcción que represente actividades que se realicen cada determinado tiempo.
- Mensajes y RPC: El procesamiento distribuido requiere que los procesos alojados en distintos host, se comuniquen de alguna manera para poder intercambiar información. Existen dos formas distintas de realizar dicha comunicación: mediante el envío de mensajes y mediante la utilización de RPC.
- Colas de Mensajes: Para aplicaciones con comunicación asíncrona muchas veces los mensajes recibidos desde distintos nodos y aplicaciones se van encolando para luego ser procesados cuando el recurso esté disponible. Ejemplos de implementaciones de esta tecnología son MQSeries de IBM [4] y Microsoft Message Queuing (MSMQ) [5].
- Web Services: Actualmente tanto en entornos GRID como en entornos empresariales para lograr la reutilización de funcionalidades entre aplicaciones es posible exponer una funcionalidad determinada por medio de web services, lo que permite a otras aplicaciones consumirlos mediante una simple llamada http.

Los siguientes elementos surgen de hacer adaptaciones a los elementos preexistentes del diagrama de Actividades.

- Nodos: Cada calle del diagrama de actividades será un nodo de aplicación.
- Hilos: Se propone una notación mas clara, cuando se mencionan estados dispares de cada hilo a través de líneas punteadas.

Por otra parte es muy común en las aplicaciones paralelas y procesamiento concurrente la utilización de memoria

compartida. Seguidamente se detallan dos tipos específicos de las mismas.

- Memoria compartida: una memoria única y global accesible desde todos los procesadores.
- Memoria compartida – distribuida: la memoria está físicamente distribuida pero lógicamente compartida.

2) Una vez identificados los elementos para el metamodelado se realiza la definición de los estereotipos para cada elemento que se está extendiendo.

Es importante tener en cuenta que elementos del metamodelado de UML se están extendiendo y sobre los que es posible aplicar un estereotipo. Ejemplos de tales son clases, asociaciones, relaciones, operaciones, atributos, etc. De esta forma el estereotipo se aplicara a una metaclass de UML. En la tabla 1 se puede observar la asociación entre los elementos de la extensión y las metaclasses en la cual se definen los estereotipos para los

mismos. A su vez los estereotipos definidos generan nuevas metaclasses que también pueden ser extendidas, por ejemplo en la tabla 1 se muestra una Cola de Mensajes utilizando como metaclass ProMemoria que es un estereotipo propio del profile.

Tabla 1. Asociación entre los elementos de extensión y las metaclasses

Elemento de ProCodi	Estereotipo	UML Metaclass
Swimlanes	ProSwimlanes	Colaboración
Memoria	ProMemoria	Clase
Actividades	ProActividad	Operación
Semáforos	ProSemaforo	Clase
Monitor	ProMonitor	Clase
Timer	ProTimer	Parámetro de Actividad
Cardinalidad	ProCardinalidad	Operación
Cola de Mensajes	ProColaMsg	ProMemoria
Web Services	ProWebServices	Operación
SubCalles	ProSubCalles	Colaboración

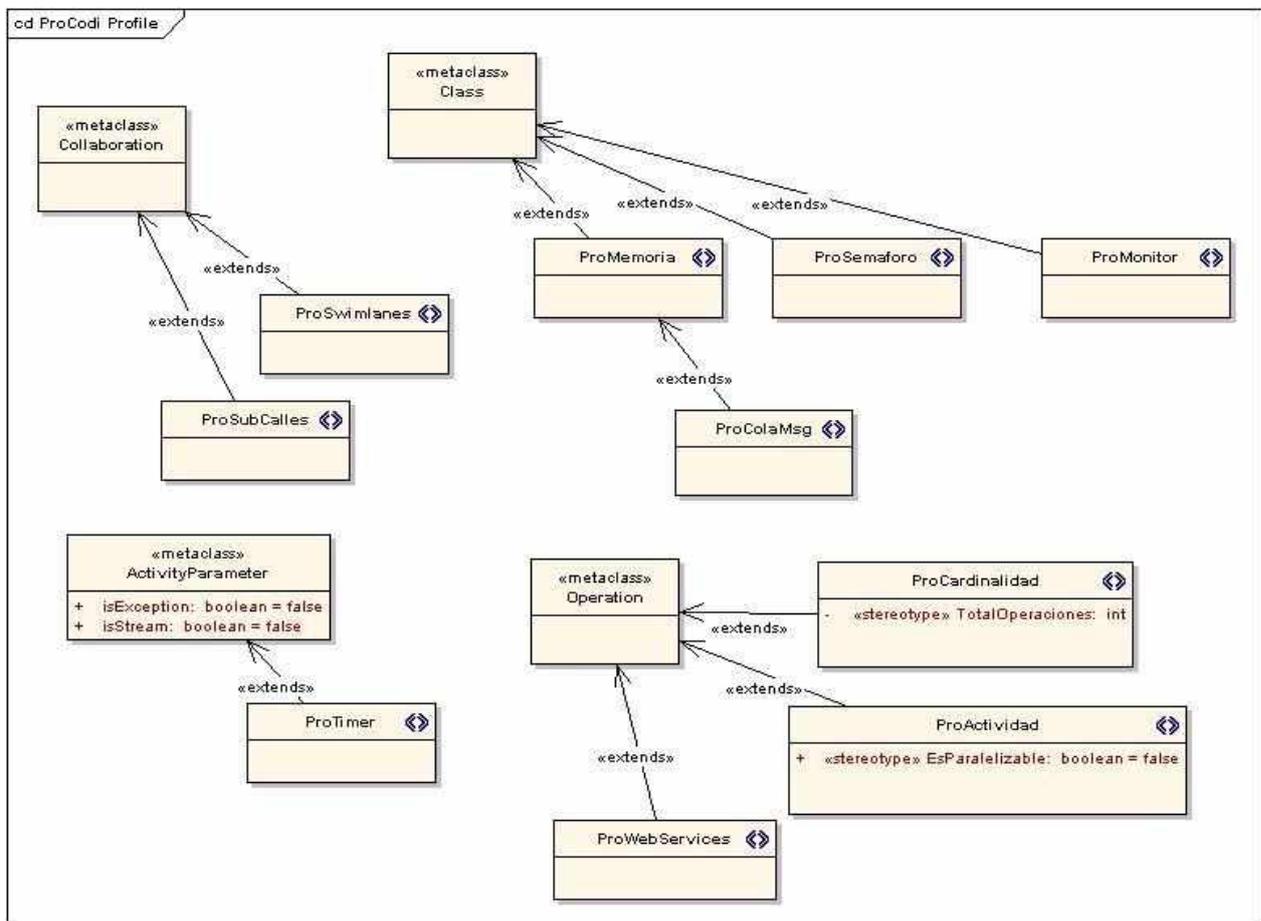


Figura 1. Determinación de los valores etiquetados para cada elemento del Profile

- 3) Se procede a la determinación de los valores etiquetados de los elementos del profile, estos valores etiquetados son los atributos de los elementos que estamos extendiendo. Deben incluir la definición de sus tipos, y sus posibles valores iniciales.

En la figura 1 se pueden observar los estereotipos derivados de las metaclasses: Class, Collaboration, Operation, ActivityParameter, así como algunos valores etiquetados en Cardinalidad para representar la cantidad total de Operaciones.

Como resultado final se obtiene el paquete del Profile mostrado en la figura 2. Dentro del paquete se incluyen todos los estereotipos usados y los creados por el profile. Este paquete es creado automáticamente por la herramienta utilizada.

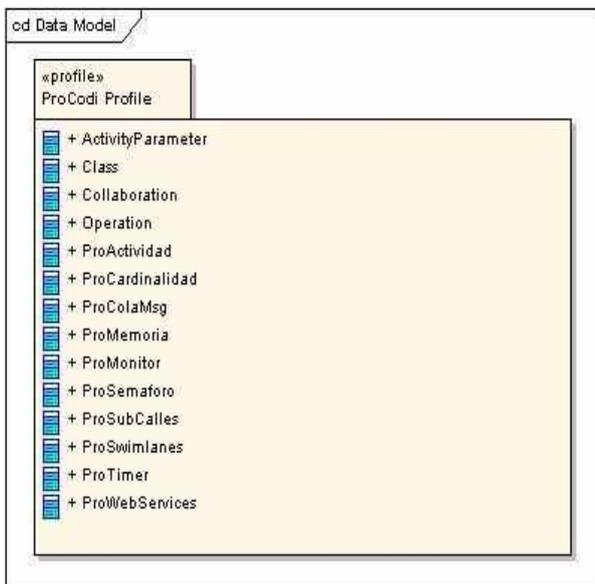


Figura 2. Paquete del profile

Actualmente el grupo de investigación se encuentra en desarrollo de las restricciones del profile, estas restricciones son escritas en OCL. Mediante OCL se podrán escribir las restricciones de las extensiones a través de un lenguaje formal que facilita la eliminación de ambigüedades.

4. Formación de recursos humanos

El equipo de investigación está formado actualmente por 6 personas, entre los cuales se encuentran docentes, egresados y colaboradores externos vinculados con

empresas privadas. Tres de los miembros del equipo se encuentran haciendo un doctorado en ciencias informáticas, el tema de tesis de uno de ellos está asociado con la presente investigación. El equipo trabaja en la definición de restricciones mediante OCL y prevé para lograr finalizar el profile incluir 3 personas más al equipo de trabajo (alumnos avanzados o becarios).

Esta línea de investigación comienza en el 2005 y ha ido evolucionando permitiendo obtener varios resultados intermedios los cuales han sido presentados en diversos congresos nacionales e internacionales. A fines del 2008 un paper presentado por el equipo de trabajo ha sido elegido en la Jornada Chilena de Computación entre los 4 mejores papers publicados en el “Workshop de Sistemas Distribuidos y Paralelismo”, obteniendo la invitación de escribir un artículo extendido para su publicación en una revista digital.

5. Bibliografía

[1] Booch G, Rumbaugh J y Jacobson I. El proceso unificado de desarrollo de software. Addison Wesley, 2001.

[2] Catalog of UML Profile of Specification http://www.omg.org/technology/documents/profile_catalog.htm

[3] Fuentes L. y Vallecillo A. Una Introducción a los Perfiles UML. Depto. de Lenguajes y Ciencias de la Computación, Universidad de Málaga Campus de Teatinos. España

[4] IBM, MQSeries. Disponible en: <http://www-306.ibm.com/software/integration/wmq/>

[5] Microsoft, Message Queuing. Disponible en: <http://www.microsoft.com/windowsserver2003/technologies/msmq/default.mspx>

[6] Warmer J., Kleppe A. The Object Constraint Language: Precise Modeling With UML.

Relevamiento de los Pilares de Gobernabilidad Electrónica en Sitios Web Municipales

Universidad Nacional de La Matanza
Departamento de Ingeniería e Investigaciones Tecnológicas
Florencio Varela 1903, San Justo, Buenos Aires, Argentina

Rocío A. Rodríguez
rrodri@unlam.edu.ar

Daniel A. Giulianelli
dgiulian@unlam.edu.ar

Pablo M. Vera
pablovera@unlam.edu.ar

Artemisa Trigueros
artemisa@unlam.edu.ar

Isabel B. Markos
marko@unlam.edu.ar

CONTEXTO

Este artículo se basa en el trabajo realizado por el equipo de investigación de Gobierno Electrónico (e-governance) de la Universidad Nacional de La Matanza. Proyecto subvencionado por el Programa de Incentivos al Docentes Investigador 55-C073 durante los años 2007 y 2008. Tiene continuidad durante los años 2009 y 2010 con otro proyecto vinculado denominado “aseguramiento de la calidad para aplicaciones móviles de gobierno electrónico” (m-governance).

RESUMEN

La gobernabilidad electrónica (e-governance) tiene como propósito proporcionar a los ciudadanos una mejor calidad de gobierno, proveyendo servicios e información y también fomentando la comunicación interactiva entre ambos. Este objetivo se logra mediante la implementación de herramientas proporcionadas por las TICs (Tecnologías de la Información y la Comunicación) en el diseño y contenido de los sitios web gubernamentales. Esta forma de comunicación permite una redefinición del rol tradicional que cada uno juega en la relación. El gobierno como proveedor de servicios, información, transparencia y formas de comunicación activa y pasiva. Los ciudadanos como sujetos activos dentro de su gobierno, utilizando servicios, recibiendo información y también controlando las acciones de gobierno y brindando retroalimentación a sus gobernantes. La retroalimentación incluye opiniones, críticas y sugerencias utilizando los medios

tecnológicos ofrecidos por medio de un sitio web.

La investigación analizó los sitios web municipales y verificó si la aplicación de las TICs de acuerdo a normas internacionales y nacionales de diseño y contenido, contribuyen a lograr un alto nivel de: Democracia electrónica, Servicios electrónicos, Transparencia electrónica y Comunicación activa- pasiva, todos ellos conceptos básicos necesarios para elevar la calidad de gobernabilidad electrónica.

Palabras clave: Gobernabilidad Electrónica, Democracia electrónica, Servicios electrónicos, Transparencia, Comunicación

1. INTRODUCCION

“La gobernabilidad electrónica se refiere al uso de las tecnologías de la información y la comunicación por parte del sector público con el objetivo de mejorar el suministro de información y el servicio proporcionado. De esta manera, se trata de estimular la participación ciudadana en el proceso de toma de decisiones, haciendo que el gobierno sea más responsable, transparente y eficaz”. [1]

Es posible plantear la gobernabilidad electrónica como la intersección de dos fenómenos sociales contemporáneos: la comunicación global, generada a partir del desarrollo de las TICs y una nueva visión de gobernabilidad, donde el ciudadano asume un rol activo, en vez del tradicional rol de receptor pasivo de las acciones de gobierno. Esta concepción de gobernabilidad ha ido

perfeccionándose con el tiempo y la experiencia basada en la implementación en distintos países y comunidades. Uno de los conceptos que ha evolucionado es el de comunicación gobierno/ ciudadano. Dicha comunicación puede ser pasiva, donde el papel del gobierno municipal es el de proveedor/emisor de noticias, informaciones y servicios; mientras que el papel del ciudadano es simplemente de receptor pasivo de lo que el gobierno decide entregarle, siendo esta concepción primitiva y antigua.

A partir de las nuevas posibilidades brindadas por las TICs, se considera que una implementación más efectiva de gobernabilidad electrónica, conlleva a una comunicación interactiva entre gobierno/ciudadano, donde ambos representan el papel de emisor/receptor.

El gobierno como:

- Emisor de servicios, informaciones y noticias. Además como expositor de sus actos de gobierno, licitaciones, compras, decretos y toda otra información que implique transparencia de los actos de gobierno.
- Receptor de sugerencias, opiniones, críticas y felicitaciones emitidas por los ciudadanos.

El ciudadano como:

- Emisor de opiniones sobre los actos de gobierno, apoyo o rechazo de consultas, fuente de sugerencias, críticas y felicitaciones. Formador de grupos de debate, noticias e interés de los vecinos.
- Receptor de noticias, informaciones y servicios.

El ciudadano en lugar de ser un objeto pasivo que sólo recibe, o en muchos casos acepta medidas gubernamentales, a través de la gobernabilidad electrónica puede convertirse en sujeto activo y participante de su gobierno. En esta comunicación activa es donde se pone en juego el verdadero concepto de democracia, más evidente aún en los municipios donde las decisiones de quienes ejercen el gobierno afectan directamente, en forma positiva o negativa, la vida de cada uno de los ciudadanos y sus familias. La interacción reporta grandes ventajas a ambos

participantes. Al ciudadano le permite ejercer sus derechos, controlar los actos de gobierno y participar activamente en la toma de decisiones, así como consultar información, realizar pagos, efectuar trámites y consultas. Al gobierno le permite hacer efectiva la transparencia de su gestión gubernamental y proporcionar servicios más eficientes ahorrando tiempo y dinero.

La transparencia dentro de este contexto, se vuelve un elemento clave en el perfeccionamiento de la gobernabilidad electrónica, ya que permite reducir la corrupción, aumentar la credibilidad política así como la responsabilidad de los gobernantes, favoreciendo la participación y la consulta popular. Implica para el gobierno, hacer públicas sus acciones y reglamentaciones para que los ciudadanos puedan tener acceso a ellas y conocer que fueron hechas dentro del marco legal. Dicha información debe ser accesible a todo ciudadano en forma libre, amigable, veraz, fácil de entender y de encontrar.

La siguiente clasificación, basada en la especificación de la UNESCO [8], muestra los pilares básicos y paralelamente campos de aplicación de la gobernabilidad electrónica:

1. Administración electrónica (e-administración): Se refiere a la mejora de los procesos gubernamentales internos, la gestión de los funcionarios del sector público, los procesos de ejecución e información.
2. Servicios electrónicos (e-servicios): Se refiere a la mejora en el acto de proveer información y ofrecer acceso a servicios públicos a los ciudadanos. También están incluidas las informaciones sobre eventos, espectáculos, transporte público, bolsa de trabajo, políticas de empleo, licitaciones, mapas, etc. Como ejemplos de servicios interactivos se pueden mencionar: solicitudes de documentos públicos, solicitudes de documentos legales y certificados, expedición de permisos y licencias, otorgamiento de turnos, pagos online de impuestos, tasas y servicios...

3. Democracia electrónica (e-democracia): Implica una mayor y más activa participación ciudadana en el proceso de toma de decisiones gracias a las TICs. Como ejemplos se pueden mencionar: encuestas, foros, chat, blogs, paneles, referéndums, listas de correo, boletín por mail, contacto directo con autoridades, responsables de áreas y representantes, opiniones y sugerencias de los ciudadanos, libro de quejas.

De acuerdo a las características de la buena gobernabilidad enunciadas en [1], el equipo de investigación agregó a los precedentes, los siguientes pilares básicos:

4. Transparencia electrónica (e-transparencia): Recorre, los conceptos 2 y 3 enunciados, en forma transversal. Puede verse como e-servicios al proveer información de actos y decisiones de gobierno ya que a partir de esa información los ciudadanos conocen las acciones de gobierno y sus motivaciones. Y por otra parte también puede entenderse como e-democracia ya que el ciudadano, al informarse y tener los medios de comunicación adecuados, puede participar activamente en las decisiones y control de su gobierno.
5. Comunicación activa y pasiva (desarrollado anteriormente al referirnos a comunicación gobierno / ciudadano).

El equipo de trabajo se basó en el análisis de 4 pilares. Se excluyó el pilar de e-administración por ser considerado una función interna de la municipalidad, no evaluable a través de los sitios web.

2. LINEAS DE INVESTIGACION y DESARROLLO

El grupo de investigación define la premisa “La aplicación de las herramientas proporcionadas por las TICs de acuerdo a normas nacionales e internacionales para la construcción de sitios web, incrementa la implementación y cumplimiento de los pilares básicos de la gobernabilidad electrónica”. Bajo esta motivación se analiza la situación de los sitios web municipales que conforman el Conurbano Bonaerense para observar si en

ellos se han aplicado las herramientas proporcionadas por las TICs de acuerdo a normas nacionales e internacionales y evaluar el nivel de gobernabilidad electrónica.

Las 30 Municipalidades que conforman el conurbano bonaerense, son habitadas por aproximadamente 9.000.000 de personas, lo que equivale al 25 % de la población de la Argentina (datos obtenidos del Censo 2001, realizado por el INDEC [3]).

Para lograr los objetivos precedentemente enumerados, los sitios web gubernamentales no deben ser un sitio más, sino que deben poseer características de diseño y contenido especiales que favorezcan la participación y responsabilidad ciudadana, así como una comunicación fluida, confiable y accesible entre sus miembros. Se evalúa en qué grado los sitios web de los municipios del conurbano bonaerense, cumplen con los pilares de gobernabilidad electrónica: e-servicios, e-democracia, e-transparencia y comunicación activa-pasiva. Para lo cual se enuncian aspectos que contribuyan a uno o más pilares de la gobernabilidad electrónica. Los aspectos seleccionados para la evaluación de cada uno de los pilares, fueron extraídos de distintas fuentes:

- A nivel internacional: W3C (World Wide Web Consortium) [10]
- A nivel nacional: ONTI (Oficina Nacional de Tecnologías de Información) [4]
- Publicaciones académicas de: Australia[2], Chile[5] [7], España[9], Nueva Zelanda[6], Estados Unidos[11]

A los mismos se añaden aspectos consignados por el equipo de investigación los cuales han surgido en base a la experiencia del grupo de trabajo en diseño y desarrollo de sitios web.

Se logran consignar 80 aspectos los cuales tienen influencia al menos en uno los pilares. A cada aspecto se le otorga un puntaje que permita mostrar la relevancia del mismo en cada pilar en el que influye. La escala de puntaje a otorgar a cada aspecto es la siguiente: Indispensable: 5 puntos, Muy Importante: 4 puntos, Importante: 3 puntos, Deseable: 2 puntos y Optativo: 1 punto.

La clasificación indispensable (5 puntos), se asigna a aquellos aspectos esenciales para el

cumplimiento de algunos de los conceptos de gobernabilidad electrónica. El nivel de relevancia decrece a medida que el aspecto sea menos relevante. Es importante destacar que un mismo aspecto puede influir en varios pilares a la vez, ejemplo: “Posee formulario de contacto o mail” influye en los conceptos de e-democracia, comunicación activa y e-transparencia. Respecto a relevancia ejemplificamos a través del aspecto “Se incluyen los CV de las autoridades”, el cual obtiene una puntuación de 3 para e-servicios (información) y también 3 puntos para comunicación pasiva, pero posee 5 puntos para e-transparencia.

3. RESULTADOS OBTENIDOS

La evaluación se realizó por medio del análisis y observación de los 30 sitios web gubernamentales, llevada a cabo por el equipo de investigación. Se chequeó el cumplimiento o no de cada aspecto. Si cumple con el aspecto el sitio recibe el puntaje correspondiente, de lo contrario no se asigna puntaje. La suma del

puntaje considerando e-servicios, e-democracia, e-transparencia y comunicación, permite establecer el ranking de posicionamiento mostrado en la figura 1.

Se puede observar que aparecen mencionadas 28 municipalidades, ya que al momento de realizarse la evaluación las municipalidades de Ezeiza y San Miguel, no tenían sitio web oficial. Los 3 sitios con mayor puntaje son: San Fernando, Pilar y Morón, siendo San Fernando el que mayor porcentaje logró en e-democracia (72,13%) y Pilar en e-servicios (78,67%).

Paralelamente el equipo de investigación realizó una evaluación por cada uno de los 4 pilares de Gobernabilidad Electrónica pudiéndose observar que:

- En veintitrés sitios web, el porcentaje de e-servicios supera al de e-democracia. Para ninguno de los pilares existe sitio que supere el 80%.
- Sólo cuatro sitios web superan el 60% del puntaje total de e-servicios y e-democracia a la vez.

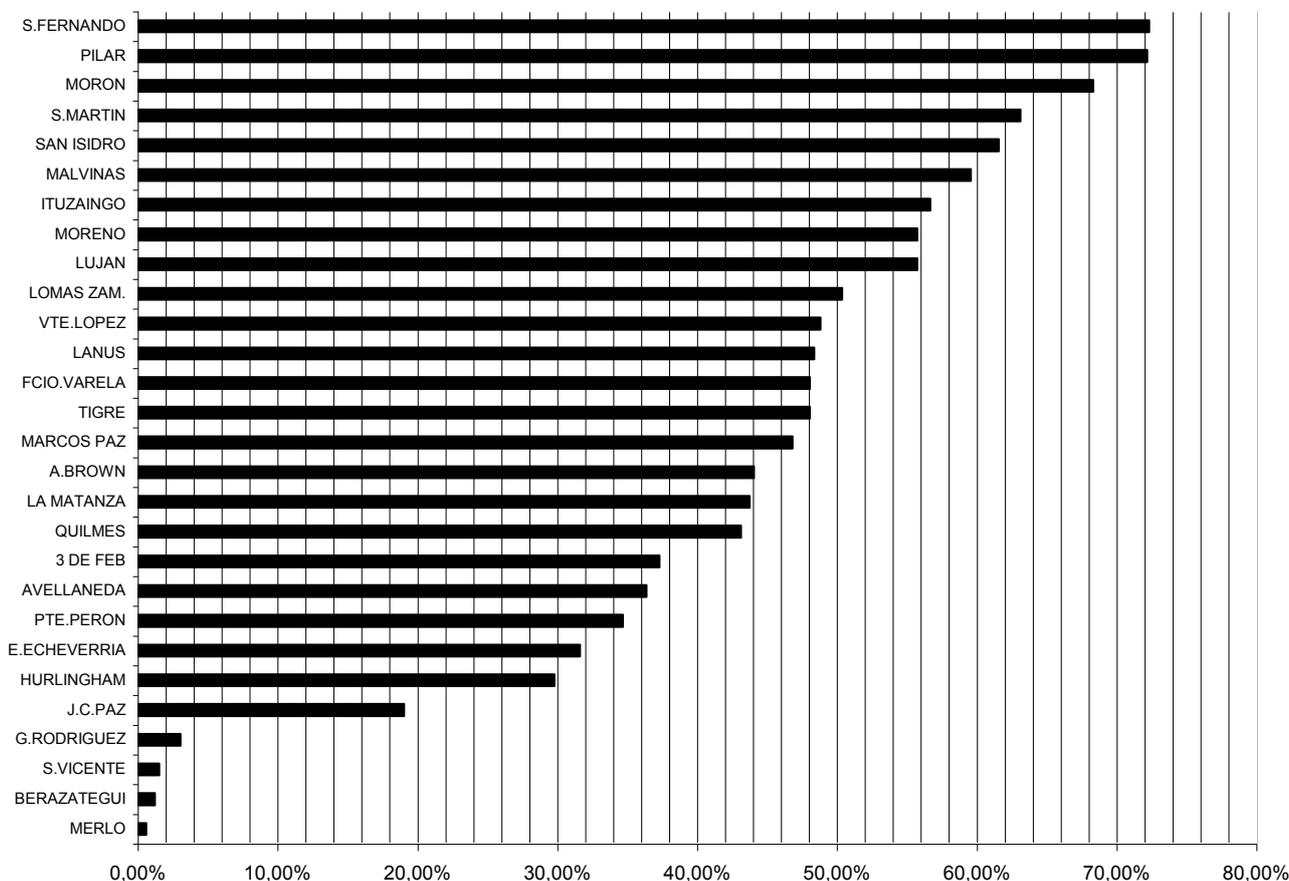


Figura 1. Porcentajes de Gobernabilidad alcanzado por los sitios web municipales

- En cuanto a e-transparencia los Municipios de San Fernando (70,09%), Morón (68,22%) y Pilar (65,42%) obtuvieron los porcentajes más altos. Habiendo en contrapartida 4 sitios que no satisfacen ningún aspecto de e-transparencia.
- En la totalidad de sitios web municipales hay mayor nivel de comunicación pasiva que activa. El máximo porcentaje en comunicación pasiva es compartido por Pilar y Morón, mientras que en comunicación activa la mejor puntuación fue obtenida por San Fernando y Pilar.

El equipo de investigación desarrolló una plantilla como propuesta de página principal en donde se sitúan las posiciones de los menús, botones y se propone un estándar de diseño unificado.

Por otra parte dada la inserción creciente de celulares y otros dispositivos móviles, es importante comenzar a considerar la posibilidad que los sitios puedan brindar servicios por medio de este tipo de equipos. Por ello se prevé realizar una propuesta de investigación en base a las posibilidades de gobernabilidad móvil (m-government) aplicado a sitios web municipales.

4. FORMACION DE RECURSOS HUMANOS

El grupo de investigación ha incorporado a dos docentes y tres alumnos sin experiencia en investigación. eGovenance es tema de tesis doctoral de uno de los investigadores que conforma el equipo de trabajo y la aplicación de mGovernment al ámbito académico es tema de tesis de maestría de otro miembro del grupo.

Es importante destacar que la presente investigación ha arrojado resultados intermedios los cuales han sido publicados por medio de papers en diversos países obteniendo además la invitación de realizar un capítulo de libro que se publicará a mediados del 2009 en Estados Unidos.

5. BIBLIOGRAFIA

[1] Castro Sardi Ximena and Mlikota Kristina. Overview On E-Governance. 2002. http://portal.unesco.org/ci/en/files/6532/10391876090Overview_on_e-

[governance_working_paper.doc/Overview%20Bon%20e-governance%20working%20paper.doc](http://www.unesco.org/ci/en/files/6532/10391876090Overview_on_e-governance_working_paper.doc/Overview%20Bon%20e-governance%20working%20paper.doc)

[2] Department of the Premier and Cabinet - Office of e-Government; Guidelines for State Government Websites Version 2.1; Australia; 2006. http://www.egov.dpc.wa.gov.au/documents/WebGuidelinesVersion2.1_final.doc

[3] INDEC (Instituto Nacional de Estadísticas y Censos). Censo Poblacional, 2001. <http://www.indec.mecon.ar/>

[4] ONTI (Oficina Nacional de Tecnologías de Información); Plan de Gobierno Electrónico Decreto 378/2005; Argentina, 2005. http://www.sgp.gov.ar/contenidos/onti/productos/pnge/docs/pnge_decreto_378_2005.pdf

[5] Pontificia Universidad Católica de Chile; Cátedra e-Government; Sitios Web Municipales e-Government Chile; 2006. <http://www.cetiuc.cl/wp-content/uploads/2007/01/presentacion-estudio-municipalidades.pdf>

[6] State Services Commission; NZ Government Web Standards and Recommendations; Version 1.0; Nueva Zelanda; 2007. <http://www.e.govt.nz/standards/web-guidelines/web-standards-v1.0/web-standards-v1.0.rtf>

[7] Subsecretaría de Desarrollo Regional y Administrativa; Ministerio del Interior; Gobierno de Chile; Resultados Segunda Encuesta Realidad Tecnológica Municipal; 2004. http://www.subdere.gov.cl/1510/articles-68168_recurso_1.ppt

[8] UNESCO (United Nations Educational, Scientific and Cultural Organization). Gobernabilidad electrónica: Fortalecimiento de capacidades de la gobernabilidad electrónica. <http://portal.unesco.org/ci/en/files/14896/11412266495e-governance.pdf/e-governance.pdf>

[9] Universidad Pompeu Fabra; Evaluación de la usabilidad en sistemas web municipales, metodología de análisis y desarrollo; Barcelona; 2005. <http://www.semanticweb.net/archives/2005/evaluacion-municipales-isko.pdf>

[10] W3C; Guía Breve de Accesibilidad Web. <http://www.w3c.es/divulgacion/guiasbreves/Accesibilidad>

[11] Web Managers; Requirements and Best Practices Checklist for Government; USA; 2005. http://www.usa.gov/webcontent/reqs_bestpractices/checklist/long.pdf

Construcción de una ontología para gramáticas formales y máquinas abstractas utilizando Protégé para la elicitación de requerimientos

Ing. Marcelo Marciszack/ Facultad Regional Córdoba/ Universidad Tecnológica Nacional

Dr. Manuel Pérez Cota / Facultad / Universidad de Vigo

Mg. Leandro Antonelli / Universidad Nacional de La Plata

Dra. Roxana Giandini / Universidad Nacional de La Plata

Marina Cardenas/ Facultad Regional Córdoba/ Universidad Tecnológica Nacional

CONTEXTO

El presente proyecto se encuentra consolidado dentro de la línea de investigación encabezada por el proyecto "Modelos de especificación de requerimientos para la obtención de esquemas conceptuales en un dominio restringido: comparación de metodologías" desarrollado en el Dpto. de Sistemas de Información de la UTN FRC, el cual tiene como objetivo la comparación de diferentes metodologías y herramientas para la especificación de requerimientos con el fin de determinar un esquema conceptual sobre un dominio de aplicación determinado: Máquinas Abstractas y las Gramáticas formales.

RESUMEN

El presente trabajo describe el desarrollo de un modelo de representación conceptual del dominio de la teoría de gramáticas formales y máquinas abstractas, a través del modelado ontológico, con el objetivo de evaluar su utilización como herramienta de soporte a la elicitación de requerimientos de software basándose en una metodología de modelado ontológico.

Este desarrollo es parte de un proyecto que nos permitirá comparar entre sí las diferentes metodologías y herramientas que suelen utilizarse en la etapa de relevamiento de requerimientos del proceso de desarrollo de software, y por otro lado, permitirá establecer el grado de correspondencia entre la conceptualización de las Máquinas y Gramáticas, ya que al existir un isomorfismo entre ambos dominios, éste debería continuar en los esquemas

conceptuales resultantes de la aplicación de cada una de las metodologías comparadas.

Palabras clave: Ontología, modelo conceptual, elicitación, maquinas abstractas, gramáticas formales, captura de requerimientos, Protégé 2000.

1. INTRODUCCION

Una ontología es un sistema de representación del conocimiento acerca de un dominio o ámbito específico, con el fin de obtener una representación formal de los conceptos que contiene y de las relaciones que existen entre dichos conceptos.

Además, una ontología se construye en relación a un contexto de utilización especificando una conceptualización, por lo que cada ontología incorpora un punto de vista. Todas las conceptualizaciones (definiciones, categorizaciones, jerarquías, propiedades, herencia, etc.) de una ontología pueden ser procesables e interpretadas por una computadora o un ser humano.

Un concepto también asociado a este sistema de representación es la Web Semántica, la cual tiene como objetivo principal la creación de un medio universal para el intercambio de información basado en representaciones del significado de los recursos que se encuentran en la Web, de una manera inteligible para las máquinas. Para que esto pueda llevarse a cabo, se necesita que el conocimiento de la web esté representado de forma que sea legible por las computadoras, esté consensuado, y sea reutilizable y es por ello que las ontologías proporcionan la vía para representarlo.

Para el diseño del modelo ontológico se establece como dominios a modelar las Gramáticas Formales y Máquinas Abstractas. La elección de estos dominios tiene un doble propósito: por un lado nos permitirá comparar entre sí las diferentes metodologías y herramientas de la etapa de relevamiento de requerimientos del proceso de desarrollo de software y por otro lado, permitirá establecer el grado de correspondencia entre la conceptualización de las Máquinas y Gramáticas, ya que al existir un isomorfismo entre ambos dominios, éste debería continuar en los esquemas conceptuales resultantes de la aplicación de cada una de las metodologías comparadas.

Elementos de las ontologías

Las ontologías proporcionan un vocabulario común de un área y definen, a diferentes niveles de formalismo, el significado de los términos y relaciones entre ellos.

El conocimiento en ontologías se formaliza principalmente usando cinco tipos de componentes: conceptos, relaciones, funciones, axiomas e instancias.

Los conceptos, entidades o clases en la ontología se suelen organizar en taxonomías. Se suele usar tanto el término clases como conceptos. Un concepto puede ser algo sobre lo que se dice algo y, por lo tanto, también podría ser la descripción de una tarea, función, acción, estrategia, proceso de razonamiento, etc.

Las relaciones representan un tipo de interacción entre los conceptos del dominio. Como ejemplos clásicos de relaciones binarias podemos mencionar: “subclase de” y “conectado a”.

Las funciones son un tipo especial de relaciones en las que el n-ésimo elemento de la relación es único para los “n-1” precedentes.

Los axiomas son expresiones que son siempre ciertas. Pueden ser incluidas en una ontología con muchos propósitos, tales como definir el significado de los componentes ontológicos, definir

restricciones complejas sobre los valores de los atributos, argumentos de relaciones, etc. verificando la corrección de la información especificada en la ontología o deduciendo nueva información.

Las instancias se usan para representar elementos específicos de la ontología.

Herramienta de modelado de ontologías: Protégé

Las ontologías requieren de un lenguaje lógico y formal para ser expresadas.

En la inteligencia artificial se han desarrollado numerosos lenguajes para este fin, algunos basados en la lógica de predicados y otros basados en frames (taxonomías de clases y atributos), que tienen un mayor poder expresivo, pero menor poder de inferencia; e incluso existen lenguajes orientados al razonamiento. Todos estos lenguajes han servido para desarrollar otros lenguajes aplicables a la Web. Es por ello que para el desarrollo de este proyecto se ha optado por utilizar la herramienta Protégé [2] que implementa el lenguaje OWL (Ontology Web Language) [3][4] para el modelado de ontologías basadas en Frames. La misma ha sido desarrollada por la Universidad de Stanford y se utiliza para el desarrollo de Ontologías y Sistemas basados en el conocimiento por medio de una interfaz de usuario que facilita la creación de estructuras de frames con clases, slots e instancias de una forma integrada.

Protégé permite:

- Modelar una ontología de las clases que describen un tema particular.
- Creación de una herramienta de adquisición de conocimiento para recoger conocimiento.
- Entrar en casos específicos de datos y de la creación de una base de conocimiento.
- La ejecución de usos.

Metodología

Debido a la existencia de diversas metodologías de desarrollo de ontologías,

se ha optado por la descrita en [1] ya que se adapta mejor a los requerimientos de modelado, desde el punto de vista de la simplicidad y completitud de la misma.

Antes de proceder a la explicación de dicha metodología, es preciso definir una serie de reglas que ayudarán a tomar decisiones de diseño y modelado:

1. No existe una forma correcta y única de modelar un dominio, es por ello que la mejor solución casi siempre depende de la aplicación que se le dará a la ontología una vez modelada.

2. El desarrollo de ontologías es un proceso necesariamente iterativo. La ontología inicial evoluciona y se refina a través de las sucesivas iteraciones.

3. Los conceptos en la ontología deben ser cercanos a los objetos (ya sean físicos o lógicos) y relaciones en el dominio de interés.

Tomando como base lo dicho anteriormente, a continuación se explicará la metodología adoptada para realizar el modelado de la ontología:

Paso 1. Determinar el dominio y alcance de la ontología.

Paso 2. Considerar la reutilización de ontologías existentes.

Paso 3. Enumerar términos importantes para la ontología

Paso 4. Definir las clases y la jerarquía de clases.

Paso 5. Definir las propiedades de las clases: slots.

Paso 6. Definir las facetas de los slots.

Paso 7. Crear instancias.

Aplicación

Debido a que las gramáticas proporcionan las reglas utilizadas en la generación de las

cadenas de los lenguajes, es inmediata la relación entre estas gramáticas y las Maquinas Abstractas capaces de aceptarlos. Es así que los lenguajes son el puente que vincula gramáticas y maquinas. En el modelo ontológico obtenido a partir de la metodología anterior, se puede observar que se ha representado este tipo de relaciones con el objetivo de determinar el isomorfismo entre los conceptos mencionados anteriormente.

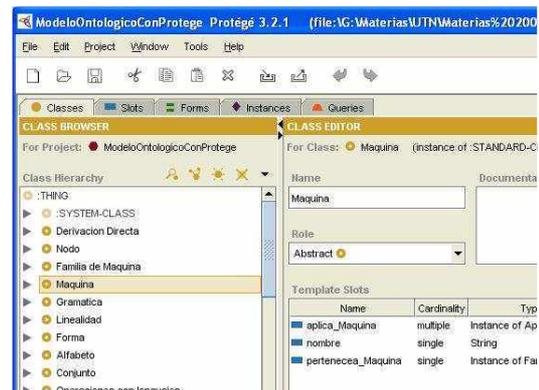


Figura 1. Modelo de clases en Protégé 2000

Utilizando Protégé 2000 para el modelado de la ontología diseñada en el presente trabajo, se procedió a la construcción de la jerarquía de clases conceptuales vinculadas al dominio.

Subsecuentemente se incorporaron los Slots a las clases conjuntamente con las facetas asociadas a cada uno. Las relaciones jerárquicas de herencia se pueden visualizar al hacer clic sobre una clase Padre (o también llamadas clases base).

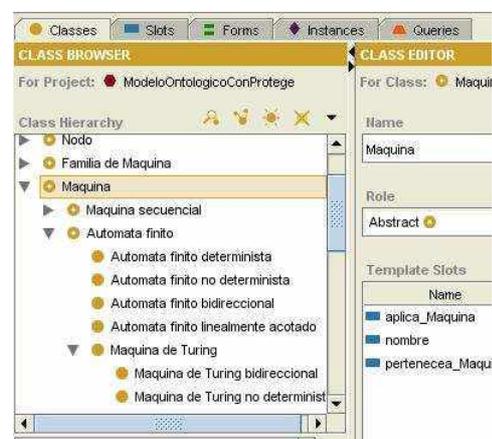


Figura 2. Taxonomía de clases en Protégé

Aquellas relaciones que no son de herencia entre los conceptos, son representadas a través de slots del tipo Instance, en el cual

se define en su faceta, el tipo de clase con la cual se relaciona.

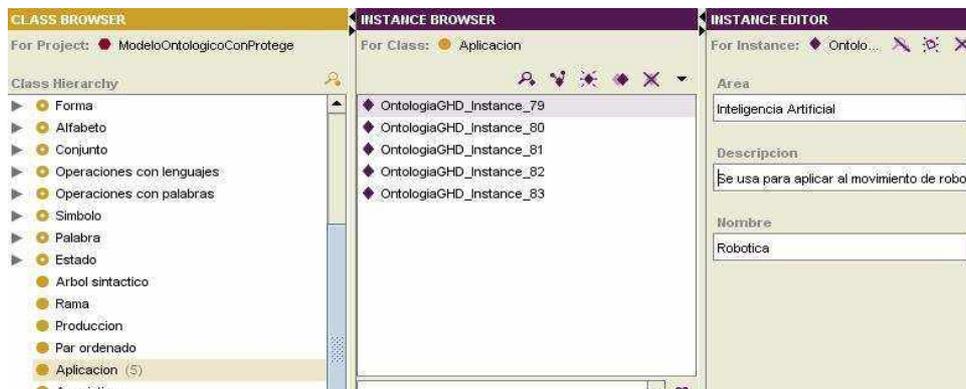


Figura 3. Creación de Instancias con Protégé 2000

Una vez que se realizó la carga del modelo completo, se crearon nuevas instancias para algunas clases con el objeto de efectuar pruebas, para lo cual se ingresó una consulta en la pestaña de la ventana principal llamada Queries.

Se prevé en trabajos futuros, la utilización de programas razonadores de ontologías para poder inferir nuevo conocimiento a partir el modelo ontológico.

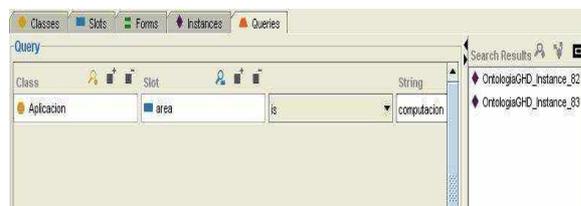


Figura 4. Creación de Queries con Protégé 2000

2. LINEAS DE INVESTIGACION y DESARROLLO

Como se mencionó inicialmente, el presente proyecto forma parte de la línea de investigación encabezada por el proyecto "Modelos de especificación de requerimientos para la obtención de esquemas conceptuales en un dominio restringido: comparación de metodologías", el cual tiene como objetivo la comparación de diferentes metodologías y herramientas para la especificación de requerimientos con el fin de determinar un esquema

conceptual [5] sobre el dominio de aplicación de las Máquinas Abstractas y Gramáticas Formales.

Dichas metodologías y herramientas derivarán en proyectos centrados en la obtención de modelos conceptuales comparables según una serie de parámetros que se definirán en base a una matriz comparativa predefinida oportunamente al terminar el desarrollo de dichos modelos. Dentro de la línea de investigación, cada proyecto se basará en las siguientes metodologías / herramientas respectivamente:

- ✦ LEL (Léxico Extendido del Lenguaje) [6], Escenarios y Tarjetas CRC, utilizando como herramienta de descripción al BMW (Baseline Mentor Workbench).
- ✦ Casos de Uso obtenidos a partir de la metodología RUP/UML [7] (Rational Unified Process) con la utilización de Rational Rose.
- ✦ Ontologías utilizando Protégé-2000 como herramienta de modelado y edición de Ontologías.
- ✦ Modelo de Objetos y Diagrama de Clases obtenido a partir de una definición Ad-hoc.

3. RESULTADOS OBTENIDOS / ESPERADOS

En base al modelo ontológico obtenido, se prevé la ampliación y profundización de los conceptos representados, continuando con el proceso iterativo que caracteriza a la metodología utilizada para su implementación. A partir de ello, se incorporará y refinará el dominio del problema que delimita el alcance de la ontología y se construirá un conjunto de queries de prueba del modelo ontológico para comprobar su performance. Una vez obtenido el modelo ontológico final, se procederá a compararlo con los obtenidos de otras metodologías y herramientas tales como las siguientes: LEL, Escenarios y Tarjetas CRC, utilizando como herramienta de descripción al BMW, Casos de Uso obtenidos a partir de la metodología RUP/UML con la utilización de Rational Rose; y un modelo de objetos y diagramas de clase obtenido a partir de una definición Ad-hoc.

4. FORMACION DE RECURSOS HUMANOS

Esta línea de investigación forma parte del proyecto mencionado en la sección 2 el cual también se encuentra enmarcado dentro del ámbito de la temática adoptada por uno de sus integrantes como parte de su tesis de maestría en ingeniería de software y como parte de un informe técnico de investigación de su carrera de doctorado.

Además, se ha incorporado como parte integrante del proyecto, una beca BINIT para fomentar el acercamiento de jóvenes graduados a las actividades de investigación.

Por otra parte, también participan del proyecto, alumnos del último nivel la carrera de Ingeniería en Sistemas de Información de la UTN FRC, próximos a recibirse y con perspectivas de iniciarse en una carrera de posgrado o doctorado, con lo cual, uno de los objetivos del proyecto es el

contribuir a la formación de dichos alumnos.

5. BIBLIOGRAFIA

[1] Ontology Development 101: A Guide to Creating Your First Ontology. Noy, Natalya F., McGuinness, Deborah L.(2005). Stanford University, Stanford.

[2] Ontologías 2. Ontologías en acción. Protégé – OWL. Alberto Barrón Cedeño (2005).

Homepage: <http://theory.lcs.mit.edu/~rajsbaum/cursos/web/ontologias2.pdf>

[3] W3C. OWL Web Ontology Language. Overview. <http://www.w3.org/TR/owl-features/>

[4] OWL. Homepage: <http://www.hipertexto.info/documentos/owl.htm>

[5] Modelado de Requisitos para la Obtención esquemas conceptuales. Emilio Insfrán, Isabel Díaz y Burbano Margarita. <http://www.dsic.upv.es/~einsfran/papers/39-ideas2002.pdf>

[6] Herramienta para implementar LEL y Escenarios (TILS). Gustavo Gil, Alejandro Oliveros, Gustavo Rossi. Tesis de Maestría Universidad Nacional de la Plata.

[7] The Unified Language User Guide. G. Booch, J. Rumbaugh, I. Jacobson. Addison Wesley 1998.

Evaluación de Atributos de Sitios de Gobierno Electrónico

Marcelo Castro^{*}, Aristides Dasso[§], Ana Funes[§]

^{*}Universidad Nacional de Jujuy; [§]Universidad Nacional de San Luis

CONTEXTO

Este trabajo se encuentra enmarcado dentro del trabajo de investigación “TIC’s: automatización y estandarización del proceso de Gobierno electrónico” perteneciente al grupo de Investigación y Desarrollo en Gobierno Electrónico, perteneciente al Área de Ingeniería Informática de la Facultad de Ingeniería, código D0047 de la Secretaría de Ciencia y Técnica - Universidad Nacional de Jujuy y del proyecto de incentivos de la Universidad Nacional de San Luis, código 22/F822: “INGENIERIA DE SOFTWARE: CONCEPTOS, METODOS Y HERRAMIENTAS EN UN CONTEXTO DE INGENIERÍA DE SOFTWARE EN EVOLUCIÓN.”

RESUMEN

Esta investigación se enmarca dentro del campo de la Ingeniería de Software, particularmente en lo que hace a la aplicación de metodologías formales y semi-formales de evaluación de sistemas, especialmente de sitios web de Gobierno Electrónico (GE o e-gov) así como el de diseñar una metodología que conjugue las herramientas de gestión y los servicios al ciudadano, con conceptos de teoría de las organizaciones y a la utilización de los aspectos relacionados a la Calidad del Software (Ingeniería Web).

Palabras clave: Gobierno Electrónico. Métricas. Métricas de Software. Métricas de Productos. Métricas de sitios web. Tecnologías de Información y Comunicación (TIC).

1. INTRODUCCION

Tal como se cita en el documento de Consejo Federal de Inversiones (CFI), “Las Nuevas Tecnologías de Información y Comunicación y las Administraciones Públicas Provinciales “El gobierno electrónico (GE) o también denominado gobierno digital, concepto de reciente referencia en el lenguaje de lo público, no es más que una propuesta de la incorporación de las TIC al proceso de producción de bienes y servicios del estado, para sumar un nuevo espacio en la relación estado-ciudadano: el espacio virtual”.

El presente trabajo es realizado teniendo en cuenta la problemática que representa el no contar con un modelo de referencia para el diseño y la evaluación de sitios de Gobierno Electrónico (GE) referido a la característica de funcionalidad específica del dominio. Partiendo de esta necesidad, el trabajo tiene como objetivo principal definir un conjunto de atributos para dicha característica proponiendo, simultáneamente, un modelo que sirva de referencia para cualquier sitio Web de GE que se quiera desarrollar o que se encuentre en producción, es decir, una cantidad mínima de servicios que debería contener una sitio Web de alguna repartición pública perteneciente a cualquiera de los Poderes del Estado

Es importante señalar que el trabajo se enmarca en la metodología Web-site QEM [OLS00], [OLS01a], [OLS01b]; y en el método de evaluación de sistemas Logic Scoring of Preference (LSP) [DUJ87]. El modelo en desarrollo está, además, sustentado en el modelo de calidad ISO 9126 [ISO9126-91].

El trabajo es desarrollado y pensado en primer lugar en el contexto de GE, formando parte de la Metodología para la inserción de GE en el ámbito de la Administración Pública, que es el lugar en dónde se realizan los aportes más significativos del trabajo y luego en el de la Ingeniería Web, que sirve de herramienta para lograr el objetivo planteado anteriormente.

Entre otros de los trabajos relacionados que fueron consultados y tenidos en cuenta para la construcción del modelo, y en referencia al método de Logic Scoring Preference o LSP (Método de Puntaje Lógico Multicriterio) para evaluar y seleccionar hardware y software de sistemas complejos, fue el realizado por Michael Chun Long Yip y Emilia Mendes, “Web Usability Measurement: Comparing Logic Scoring Preference to Subjective Assessment” [YIPMEN05] en el cual se discuten los resultados de dos experimentos formales realizados para evaluar el grado en el que LSP incorpora la percepción subjetiva de usuarios, específicamente al concepto de usabilidad en la Web. En este trabajo se observa que los valores obtenidos mediante la utilización de LSP difieren significativamente de los valores obtenidos mediante sólo la opinión subjetiva de los usuarios. Más información sobre el método LSP se puede encontrar en [DUJ87].

En [DASS-FUN04], donde también se emplea el LSP, se discuten las diferentes características de los modelos de evaluación de Data Base Management Systems (DBMS) y el razonamiento detrás de las diferentes opciones realizadas, así como las funciones propuestas a utilizar para los criterios elementales y la estructura de agregación. Se explica como los modelos propuestos, pueden ser una manera de integrar los resultados obtenidos desde los benchmarks para conseguir una única imagen total y además una forma de integrar cuantitativamente en un número, esos resultados.

2. LINEAS DE INVESTIGACION y DESARROLLO

El tema de la evaluación del rendimiento de los sistemas de software es amplio y en muchos casos complejo.

En general, como se ha dicho antes, esta evaluación emplea métodos semi formales. Muchas veces de los llamados aditivos.

En este caso, además de emplear un método formal, que va más allá de los meramente aditivos, se construye un modelo para evaluar sitios web de e.government.

Finalmente, es importante señalar que este trabajo se encuentra enmarcado dentro del trabajo de investigación “TIC’s: automatización y estandarización del proceso de Gobierno electrónico” perteneciente al grupo de Investigación y Desarrollo en Gobierno Electrónico, perteneciente al Área de Ingeniería Informática de la Facultad de Ingeniería de la Universidad Nacional de Jujuy, Argentina,. dentro de cuyos objetivos está el de diseñar una metodología que conjugue las herramientas de gestión y los servicios al ciudadano, con conceptos de teoría de las organizaciones y a la utilización de los aspectos relacionados a la Calidad del Software (Ingeniería Web). En este marco, los trabajos “Formalización del proceso de gestión de las TI y su aplicación a GE” de Marcelo Castro, José Farfán, David Sánchez Rivero, Daniel Castro [CAS-FAR07a] y “TIC: una metodología para formalizar el proceso de gobierno electrónico” [CAS-FAR07b] plantean una guía metodológica básica que permite formalizar el proceso de gestión de las Tecnologías Informáticas (TI), específicamente en el proceso de implementación de Gobierno Electrónico (GE), para cualquier repartición pública independientemente de la plataforma informática con la que cuenta. En primer lugar, se plantea el inconveniente que supone incorporar las tecnologías de la información y de la comunicación (TIC) en el sector público, debido a la inexistencia de un método que guíe el proceso. Posteriormente, se plantea la necesidad de formalizar la incorporación de

GE a la administración pública, especificando los distintos aspectos a tener en cuenta. Finalmente, se realiza una primera propuesta de metodología para la formalización del proceso de GE de la cual el presente trabajo de tesis forma parte como una fase dentro de la metodología propuesta.

Además se encuentra enmarcado, como parte de la formación de recursos humanos, en el proyecto de incentivos de la Universidad Nacional de San Luis, código 22/F822: "INGENIERIA DE SOFTWARE: CONCEPTOS, METODOS Y HERRAMIENTAS EN UN CONTEXTO DE INGENIERÍA DE SOFTWARE EN EVOLUCIÓN."

3. RESULTADOS OBTENIDOS/ESPERADOS

Entre las contribuciones más importantes podemos decir que el presente trabajo ayudará, a partir del planteo de un modelo formal de referencia, a definir claramente los servicios a brindar al ciudadano a través de un sitio de GE, estableciendo atributos mensurables y métodos para computarlos.

Otra contribución que se intenta realizar en el trabajo, es mostrar las ventajas de la utilización de disciplinas, metodologías y estándares que ayuden a formalizar los procesos de GE, como es el caso de la utilización de la Ingeniería Web, metodologías como la Web-site QEM y estándares como ISO 9126.

Además, intentará aportar una ayuda a los diseñadores de sitios de GE, con el objeto de mejorar la calidad en el diseño y desarrollo de las aplicaciones.

Se está construyendo un modelo de evaluación de sitios de e-gov el que puede considerarse formal ya que emplean técnicas de evaluación de sistemas y sitios web que están basados en lógica continua.

El modelo, con el que ya se están llevando a cabo de algunas evaluaciones y comparaciones, puede servir para evaluar y comparar otros sitios de e-gov.

Este modelo se encuentra en permanente evolución, de la misma manera que lo hacen los sitios de e-gov. Debe tenerse en cuenta que muchos sitios de e-gov se encuentran incorporando tecnología orientada a la Web semántica, por lo que generar un modelo que permita evaluar este tipo de sitios es uno de los objetivos futuros del proyecto.

Asimismo se estudia la posibilidad de incorporar web crawlers, que basados en uno o varios modelos puedan realizar una evaluación cuasi permanente de varios sitios de e-gov.

4. FORMACION DE RECURSOS HUMANOS

El trabajo aquí presentado ha dado como resultado inicial una tesis de maestría que se encuentra en su última fase y que se presentará en los próximos meses.

Además este trabajo es una colaboración entre investigadores de dos grupos de investigación de dos universidades nacionales.

5. BIBLIOGRAFIA

- [AENOR, 1995] AENOR, UNE-EN-ISO 8402 Gestión de la calidad y aseguramiento de calidad. Vocabulario (ISO 8402:1994), AENOR, 1995
- [BAC01] Backus, M., "E-governance in Developing Countries", The International Institute for Communication and Development, IICD, Research Report N°3, Abril 2001.
- [BAS88] Basili, V.R. y Rombach, H.D., "The TAME Project: Towards Improvement-Oriented Software Environments", IEEE Transaction on Software Engineering, 14(6), 758-73 1988.
- [BAU05] Baum, C. y Di Maio. A. Gartner's Four Phases of E-Government Model », Gartner Group RAS Services Research Note TU-12-6113, Noviembre 2.000.
- [BEC99] Martín Becerra, "El proyecto de la Sociedad de la Información en su contexto" en, Análisis N°23- Universitat Autònoma de Barcelona-España.
- [CAL96] Calvo y Fernández, 1996 Calvo-Manzano, J.A. y Fernández, L., "Hacia la calidad del software a través de la mejora de procesos", Novatica, n° 123, 1996, pp. 35-42.
- [CAS-FAR07a] Castro, Marcelo, Farfán José, David Sánchez Rivero, Daniel Castro, "Formalización del proceso de gestión de las TI y su aplicación a GE", actas de 36 JAIIO 2007, 29 y 30 de agosto de 2007, Mar del Plata, Argentina.
- [CAS-FAR07b] Castro, Marcelo, Farfán José, David Sánchez Rivero, Daniel Castro, "Tic: una metodología para formalizar el proceso de gobierno electrónico",

- 3ras. Jornadas de Facultades de Ingeniería del NOA, 2007. 8 y 9 de Octubre de 2007, Tucumán, Argentina.
- [CAV05] Daniel Cravacuore, Programa Prioritario de Investigación “Aldea XXI” Universidad Nacional de Quilmes – Argentina.2005
- [CHI06]Gobierno de Chile, *Guía metodológica 2006. Sistema de Gobierno Electrónico Programa de Mejoramiento de la Gestión.* 8-9, 12-15, 2006.
- [CMM93] Capability Maturity Model (v1.1), Software Engineering Institute, Universidad Carnegie-Mellon.
- [DAR99] S. Dart, “Containing the Web Crisis Using Configuration Management,” Proc. 1st ICSE Workshop on Web Engineering, ACM, Los Angeles, May 1999.
- [DASS-FUN07] Debnath, Narayan , Dasso, Aristides , Funes, Ana, Uzal, Roberto, Paganini, José. “E-government Services offerings evaluation using Continuous Logic”, proceeding de. IEEE-AICCSA 2007. Jordania. 13 al 16 de mayo de 2007.
- [DASS-FUN04], Dasso, Arístides- Funes, Ana- Peralta, Mario- Salgado, Carlos User Oriented Evaluation Models for DBMS,actas de ASIS 2004 (JAIIO 2004), 20-24 de Septiembre, Córdoba, Argentina, 2004.
- [DUJ87] Dujmovic, J.J., D. Batory, S. Navathe, R. Elnicki.. Su, W., Stanley, Y. “A cost-Benefit Decisión Model: Analysis, Comparison and Selection of Data Managment Systems”. ACM Transactions on Database Systems, Vol. 12, No. 3, Septiembre 1987.
- [EEU,2001], ‘Common list of basic public services and eGovernment Indicators for Benchmarking eEurope’, the Stockholm European Council, Marzo 2001.
- [EST05] Esteves, J., “Análisis del Desarrollo del Gobierno Electrónico Municipal en España”, Madrid, Octubre, 2005.
- [FEN97] Fenton y Pfleeger, 1997 Fenton, N.E. y Pfleeger, S.L., Software Metrics, A rigorous approach; Chapman&Hall, 1997.
- [FOU01] Jane E. Fountain,” Building the Virtual State”, Brookings, 2001
- [FUN-DAS05] A. Funes, A. Dasso, C. Salgado, M. Peralta, “UML Tool Evaluation Requirements”, actas de ASIS 2005 (JAIIO 2005), 29 y 30 de agosto, Rosario, Argentina,2005.
- [GILB, 1987] Gilb, T, Principles of Software Engineering Management; Addison Wesley, 1987.
- [GILB76] Gilb, T. “ Software Metrics”, Chartwell-Bratt, Canbridge, Ma, 1976.
- [GILB88] Gilb, T. “Principles of Software Engineering Management” Adisson – Wesley, 1988.
- [GOV06] Governo do Brasil, *e-Ping- Padroes de interoperabilidade do Governo Eletronico-Documento do referencia.* 2006.
- [GRU03].Grupo de Gobierno Electrónico Área Sociedad de la Información - Fundación Creer y Crecer, *Modernización del Estado y Tecnología. Claves para el diseño de una política de Gobierno electrónico.* 2003
- [GUE02] Guerrero Omar “Principios de Administración Pública”,Instituto Latinoamericano de la Comunicación Educativa ” Mexico . Tomado de www.lwctura,ilce,edu.mx:3000
- [IEEE 1061, 1992] IEEE, IEEE Std 1061-1992. Standard for a software quality metrics methodology, IEEE, 1992.
- [IEEE 1074, 1995] IEEE, IEEE Std-1995, Standard for developing software life cycle processes, IEEE, 1995.
- [IEEE88] IEEE, IEEE Guide to the Use of the IEEE Standard Dictionary of Measures to Produce Reliable Software. IEEE Std 982.2-1988.
- [ISO 9000-3, 1997] ISO, Part 3: Guidelines for the application of ISO 9001:1994 to the development, supply, install and maintenance of computer software, ISO, 1997.
- [ISO9126-91] ISO, Software product evaluation. Quality characteristics and guidelines for their use, ISO, 1991.
- [ISO9001-94] ISO, ISO 9001:1994. Quality systems -- Model for quality assurance in design, development, production, installation and servicing, ISO, 1994.
- [ISO9002-94] ISO, ISO 9002:1994. Quality systems -- Model for quality assurance in production, installation and servicing, ISO, 1994.
- [ISO, 1998] ISO, ISO 15504-1: 1998 Software process assessment. Part-1: Concept and introductory guides, ISO, 1998.
- [KITCH-WAK 1989] [Kitchenham y Walker, 1989] Kitchenham, B.A. y Walker, J.G., "A quantitative approach to monitoring software development", Software Engineering Journal, vol. 4, nº 1, 1989, pp. 2-13.
- [McCall et al., 1977] McCall J A, Richards PK y Walters GF; Factors in software quality, Vols I,II,III; US Rome Air Development Center Reports NTIS AD/A-049 014, 015, 055, 1977.
- [MUR01] S. Murugesan, Y. Deshpande , S. Hansen, A. Ginige. “Web Engineering : A New Discipline for Development of Web- Based Systems.” Lecture Notes in Computer Science 2016 Springer 2001, pag 3 – 13.
- [OLS00] Olsina L.”Metodología cuantitativa para la evaluación y Comparación de Calidad de Sitios Web”, Tesis Doctoral, Facultad Ciencias Exactas, UNLP, La Plata, Argentina.
- [OLS01a] Olsina L., Papa, M.F., Souto, M.E., Rossi, G.; “Providing Automated Support for the Web Quality Evaluation Methodology”, Proceed. of the Fourth Workshop on Web Engineering, at the 10th International WWW Conference, Hong Kong, 2001, pp. 1-11.
- [OLS01b] L. Olsina, G. Lafuente, G. Rossi. “Specifying Quality Characteristics and Attributes for Websites.” Lecture Notes in Computer Science 2016 Springer 2001, pag 266 – 278.
- [OSZ78] Oszlak, Oscar, “Formación histórica del Estado en America latina:elementos teórico-metodológicos para su estudio”Estudios CEDES Colúmen 1 Nº 3- Buenos Aires, 1978
- [PAC02]Pacific Council of Internacional Policy, *Roadmap for e-government in the developing world,* 6-26, 2.002.

[PAE03] Páez A., “Iribarren C., Neuman M., “Gobierno electrónico y Administración Pública Local”, N° 35, Razón y Palabra. 2003.

[POC02] Pocoví, G. & G. Farabollini. Gobierno electrónico: un cambio estructural. la integración de la información como requisito. *XVI Concurso de Ensayos y Monografías del CLAD sobre Reforma del Estado y Modernización de la Administración Pública.*, 10-12, 2002.

[PRE98] R. Pressman, “Can Internet-Based Applications Be Engineered?” IEEE Software, September/October 1998, pag 104– 110.

[SEG01] Jorge Segura, “Gobierno electrónico: su aplicación en los gobiernos estatales y regionales”, Mimeo, 2001.

[YIPMEN05], Yip Michael Chun Long – Mendes, Emilia; Web Usability Measurement: Comparing Logic Scoring Preference to Subjective Assessment. ICWE’05 Conference, Springer LNCS 3579, 53-62, 2005.

STORYBOARD BASADOS EN ESCENARIOS FUTUROS

Gladys Kaplan¹, Jorge Doorn^{1,2}, Renata Guatelli¹, Nora Gigante¹, Graciela Hadad¹

¹Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de La Matanza

²Depto. de Computación y Sistemas, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Pcia. de Buenos Aires

gladyskaplan@gmail.com, jdoorn@exa.unicen.edu.ar, renata.guatelli@gmail.com, ngigante@unlam.edu.ar,

gracielahadad@gmail.com

CONTEXTO

La presente propuesta es parte de la línea de investigación “Validación de Requisitos” en el proyecto de investigación “Consolidación de Requisitos de Software” del grupo de investigación en “Ingeniería de Requisitos” de la UNLaM.

RESUMEN

La certeza de estar modelando el software correcto es una preocupación de los ingenieros desde las primeras etapas del proceso de construcción del software. Es así que garantizar la actividad de validación es un objetivo primordial y como tal debe asegurar que los involucrados comprendan, con total claridad, la propuesta del nuevo software y su contexto. Este proyecto aspira a presentar el conocimiento de los Escenarios Futuros desde una perspectiva gráfica mucho más cercana a la vida cotidiana; es así que se propone la generación de storyboards que faciliten la percepción, por parte del cliente o usuario, del impacto que el futuro sistema de software tendrá en el proceso del negocio. De esta manera se espera mejorar la interacción con los clientes o usuarios para que estos puedan aceptar, discutir o negociar el contexto del futuro sistema de software e indirectamente los requisitos del software empotrados en los Escenarios Futuros e identificar los requisitos tardíos.

Palabras clave: *validación, requisitos, storyboard.*

1. INTRODUCCIÓN

La complejidad de los productos software que hoy demandan las organizaciones alertan acerca de la imperiosa necesidad de asegurar la participación del usuario desde etapas muy tempranas. Esta complejidad de contextos y servicios conlleva a buscar alternativas que aseguren la comprensión de la propuesta que se le está presentando hoy pero que define cómo será el futuro del proceso del negocio [1] [2] [3]. Otros autores en cambio utilizan modelos de validación para asegurar la especificación de requisitos [7] [8] [9] [10]. El lenguaje natural (LN) ha sido un gran avance en este terreno. Los modelos en LN aseguran en gran medida la actividad de validación. Reforzado ampliamente cuando ese lenguaje es conocido por los usuarios con mucha precisión [11]. Aun así, no es suficiente. La complejidad que puede presentar el contexto futuro y la abstracción que se requiere por parte de los clientes o usuarios puede verse limitada al estar representada sólo en texto. Aparece entonces como muy ventajosa la posibilidad de ofrecer una atractiva visualización del proceso del negocio planificado. El sustento básico de esta

visualización consiste en construir arquetipos que representen los actores, las actividades y los objetos más importantes del proceso del negocio. Esta estrategia tiene límites en su capacidad expresiva. Son muy conocidas las dificultades que se presentan al exagerar el uso de íconos en aplicaciones computacionales como en la cantidad de otras representaciones, como por ejemplo las señales de tránsito. Pasado cierto límite, el destinatario de estas representaciones esquemáticas debe memorizar su significado para poder utilizarlas eficientemente. El núcleo de este proyecto está relacionado con expandir al máximo el poder expresivo de las storyboard a ser construidas, conservando hasta donde sea posible la facilidad de comprensión de las mismas. Esta visión acepta desde el comienzo la existencia de ambigüedad en la parte gráfica de la historia, por lo que el desambiguado de la misma es uno de los principales problemas a resolver.

Si bien la Ingeniería de Software ha utilizado abundantemente gráficos para representar el conocimiento adquirido, aquí no se está hablando de gráficos, sino de esquemas, dibujos o imágenes. Es por eso, que es necesario precisar la diferencia entre estos dos enfoques que en ciertos momentos se solapan. Según la RAE un gráfico es “la representación por medio de figuras o signos” y la imagen es una “figura, representación, semejanza y apariencia de algo”.

Por tal motivo, el presente trabajo intenta promover la comprensión por parte de los involucrados acerca del contexto donde correrá el futuro software y de los servicios que este proveerá generando una actividad de validación fácil y “amigable”. Se pretende acercar y comprometer al cliente o usuario en la actividad de validación esperando del mismo la aceptación de la propuesta, la generación de nuevas ideas (nuevos requisitos), de nuevas necesidades recién comprendidas (requisitos tardíos), la detección de conflictos del contexto ocultos en otros modelos o directamente no detectados. Es necesario estimular dicho proceso creativo y de razonamiento con modelos que permitan, sin un gran esfuerzo adicional, una activa participación. Para tal objetivo se tomará lo mejor del LN y de las imágenes para presentar una estrategia de validación [13] [14] más atractiva para el usuario. Se utilizará para ello storyboard donde se generarán imágenes contextuales para visualizar un Escenario Futuro, donde el texto complementa las limitaciones de la imagen y las imágenes las limitaciones del texto.

2. PERSPECTIVA DEL PROYECTO

Se estudiará cómo mejorar, si es posible, la actividad

Validar Escenarios Futuros de un proceso de requisitos específico [4] que se basa en modelos en lenguaje natural: Léxico Extendido del Lenguaje (glosario del universo de discurso), Escenarios Actuales (situaciones observables en el proceso del negocio), Escenarios Futuros (situaciones esperadas en el negocio con la incorporación del nuevo software) y Especificación de Requisitos (descripciones explícitas de cada requisito del software).

En el marco del proceso de requisitos este proyecto se enfoca dentro de la etapa Construcción de Escenarios Futuros, que se conforma de las siguientes actividades:

- Describir
- Organizar
- Verificar
- Validar

La actividad Describir permite definir la situación que el escenario representa. Mientras que Organizar es reunir los escenarios realizados por diferentes ingenieros de requisitos o en diferentes momentos con el objetivo de detectar superposiciones entre situaciones, de separar situaciones independientes, de jerarquizar los escenarios. Esta actividad tiene como salida la creación de los Escenarios Integradores. La actividad Verificar se realiza en un proceso formal de Inspección pero, cabe destacar que es una actividad que se lleva a cabo espontáneamente durante las otras actividades. Por último la actividad Validar, que también es transversal a todo el proceso, puede ser realizada según [12] o haciendo una revisión de los modelos con los clientes o usuarios.

Siendo el centro de este proyecto la *actividad Validar*, es importante destacar su complejidad ya que requiere de modelos que aseguren la mayor comprensión del cliente o usuario. Por tal motivo, la incorporación de storyboards para validar un escenario escrito en LN da un marco más seguro y completo, buscando condensar en una imagen el mensaje relevante de cada episodio a efecto de reforzar la idea principal.

Esta estrategia de validación separa al cliente o usuario de los detalles de interfases y le permite una visión desde el punto de vista en el que se encuentra más cómodo: su ámbito de trabajo. Se entiende entonces que los storyboards contextuales cumplen mejor los objetivos de la etapa de validación que, por ejemplo, los prototipos [5] [6] [10].

A continuación se detallan las principales características de los storyboards:

- ✓ Se preserva el punto de vista del proceso del negocio,
- ✓ Se puede validar un escenario,
- ✓ Se puede validar Escenarios Integradores logrando una visión global,
- ✓ Son más fácil de comprender por el usuario,
- ✓ No genera falsas expectativas,

- ✓ El usuario sigue trabajando con herramientas conocidas,
- ✓ Son fáciles de mantener o adaptar a los cambios,
- ✓ Permiten incorporar modificaciones durante la validación.

Para este proyecto el trabajo más importante consistirá en generar galerías de actores, fondos de la acción y situaciones en la que se encuentran los actores y que visualicen, lo más claramente posible, el mensaje del episodio en primer lugar y finalmente de los escenarios.

2.1. OBJETIVOS PROPUESTOS

Los objetivos propuestos para este proyecto son:

- Comprobar si los storyboard son adecuados para validar un Escenario Futuro
- Comprobar si los storyboards son adecuados para validar Escenarios Integradores
- Comprobar si es viable su construcción semiautomática.

3. ESTADO DE AVANCE

Se ha definido una primera versión de un storyboard (mostrado en la Figura 2), que representa un Escenario Futuro (mostrado en la Figura 1).

Se han desarrollado galerías con el objetivo de estereotipar los actores involucrados en cada episodio. Para ello, se ha generado una taxonomía que muestra un gran abanico de posibilidades. Se ha comprobado que muchos episodios similares extraídos de casos muy diferentes pueden representarse con el mismo esquema. Por ejemplo, un empleado administrativo en un escritorio frente a una pantalla es una buena representación de:

- una persona controlando las necesidades de compras.
- una persona leyendo un mail.
- una persona mirando la hoja de ruta de producción.
- una persona cargando un pedido de un cliente.
-

En cambio, el mismo administrativo en un escritorio frente a una pantalla y una impresora indica:

- una persona generando un comprobante de venta.
- una persona imprimiendo estadísticas de ventas mensuales.
- una persona confeccionando recibos de sueldo.
- ...

Se planifica estudiar en profundidad la mejor representación de cada episodio y de cada Escenario Futuro en un storyboard. También se espera representar

el conjunto de escenarios con el objetivo de validar los Escenarios Integradores, a través de la representación de un storyboard con mayor contexto que provea una percepción más abarcativa de la solución.

Escenarios Futuros” que está desarrollando la Lic. Gladys Kaplan en la UNLP y la tesis de maestría “Construcción semi-automática de storyboard” de la Lic. Renata Guatelli.

4. FORMACIÓN DE RECURSOS HUMANOS

El presente proyecto es parte de la tesis doctoral “Generación Semiautomática de Prototipos a partir de

Título: **Confirmar el perfil**
Objetivo: Registrar al [candidato](#) que fue seleccionado.
Contexto:
 Ubicación geográfica: Administración Central
 Ubicación temporal:
 Precondición: El [candidato](#) debe haber sido seleccionado de la [terna](#)

Recursos: [Preocupacional](#), plantilla de [terna](#), email del [candidato](#)
Actores: [selectora](#), sistema
Episodios:

- 1- La [selectora](#) marca en la plantilla de [terna](#) el [candidato](#) seleccionado.
- 2- El sistema registra los datos del [candidato](#).
- 3- El sistema registra el estado “confirmado” del [candidato](#).
- 4- El sistema le asigna al [candidato](#) el [Preocupacional](#)
- 5- El sistema envía un mail al [candidato](#) con la información del [Preocupacional](#).

Excepciones:

Figura 1 – Escenario Futuro “Confirmar el perfil”

Confirmar el perfil

Quando el candidato ya ha sido preseleccionado, en la oficina de *Administración Central...*

Se está tratando de registrar el candidato seleccionado ...

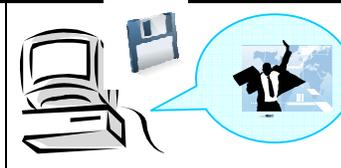
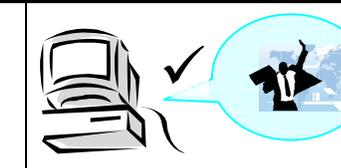
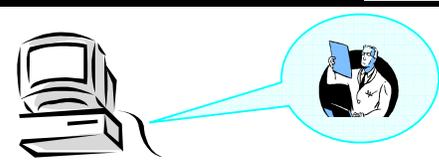
 <p>La selectora marca en la plantilla de terna el candidato seleccionado.</p>	 <p>El sistema registra los datos del candidato.</p>	 <p>El sistema registra el estado “confirmado” del candidato.</p>
 <p>El sistema le asigna al candidato el Preocupacional.</p>	 <p>El sistema envía un mail al candidato con la información del Preocupacional.</p>	

Figura 2 – Primera versión del Storyboard del Escenario Futuro “Confirmar Perfil”

5. REFERENCIAS

- [1] Leite, J.C.S.P., Hadad, G.D.S., Doorn, J.H., Kaplan, G.N.: A Scenario Construction Process, Requirements Engineering Journal, Vol.5, N° 1, pp. 38-61, 2000.
- [2] Sutcliffe, A.G., Maiden, N.A.M., Minocha, S., Manuel, D., "Supporting Scenario-Based Requirements Engineering", IEEE Transactions on Software Engineering, Vol. 24, N° 12, pp. 1072-1088, 1998.
- [3] Potts, C., "Using Schematic Scenarios to Understand User Needs", Proceedings of DIS'95 - Symposium on Designing Interactive Systems: Processes, Practices and Techniques, ACM Press, University of Michigan, pp. 247-256, 1995.
- [4] Leite, J.C.S.P., Doorn, J.H., Kaplan, G.N., Hadad, G.D.S., Ridao, M.N., "Defining System Context using Scenarios", en el libro "Perspectives on Software Requirements", Kluwer Academic Publishers, EEUU, ISBN: 1-4020-7625-8, capítulo 8, pp.169-199, 2004.
- [5] Elkoutbi M, Khriess I, Keller R. , "Generation User Interface Prototypes from Scenarios" , IEEE Software, 1999.
- [6] Logrippo D, Buhr R, Gray T, "Use Case Maps for the Capture and Validation of Distributed Systems Requirements" , RE'99: Fourth IEEE International Symposium on Requirements Engineering, 1999.
- [7] Seybold C., Glinz M., Meier S., "Simulation-based Validation and Defect Localization for Evolving, Semi-Formal Requirements Models" Proceedings of the 12th Asia-Pacific Software Engineering Conference (APSEC'05), IEEE, 2005.
- [8] He L., Carver J., Vaughn R., "Using Inspections to Teach Requirements Validation", The Journal of Defense Software Engineering, January 2008.
- [9] Somé S., "Use Cases based Requirements Validation with Scenarios", Proceedings of the 2005 13th IEEE International Conference on Requirements Engineering (RE'05), 2005
- [10] Aybüke Aurum · Claes Wohlin, "Engineering and Managing Software Requirements", Springer-Verlag Berlin Heidelberg, 2005.
- [11] Leite, J.C.S.P., "Eliciting Requirements Using a Natural Language Based Approach: The Case of the Meeting Scheduler Problem", March 1993.
- [12] Leite JCSP, Freeman PA "Requirements Validation Through Viewpoint Resolution", IEEE Transactions on Software Engineering, Vol. 17, No. 12, December 1991.
- [13] Grady, J. O., "System Validation and Verification", Boca Raton, FL: CRC Press, 1997.
- [14] Haumer P, Pohl K, Weidenhaupt K, "Requirements Elicitation and Validation with Real World Scenes." IEEE Transactions on Software Engineering, Cap.12 Vol 24, pp 1036-1054, 1998.

Desarrollo de un Framework para la Gestión de la Calidad en Ingeniería Reversa

Bruno, Juan Pablo
jp.bruno@gmail.com

Cuevas, Juan Carlos
juancarloscue@gmail.com

Gimenez Zens, Inés
igimenezzens@gmail.com

Mana, Franco
manafranco@gmail.com

Peralta, Roque Eduardo
roqueper@gmail.com

Universidad Tecnológica Nacional
Facultad Regional Córdoba
Departamento de Ingeniería de Sistemas de Información
Cátedra de Calidad en los Servicios del Software
y
Grupo de Investigación y Desarrollo en Ingeniería y Calidad del Software (GIDICALSO)

Maestro López Esq. Cruz Roja Argentina – Ciudad Universitaria
C.P 5016 – Córdoba – Argentina

CONTEXTO

Desde la cátedra de Gestión de la Calidad se dio origen, en el año 2006, al Grupo de Investigación y Desarrollo en Ingeniería y Calidad del Software (GIDICALSO). Inicialmente el objeto de investigación y desarrollo se focalizó en los aspectos de ingeniería y calidad en la producción de software. Este grupo, ante el requerimiento del Dpto. de Sistemas de esta Facultad Regional, y la Cátedra antes mencionados propusieron la creación de las asignaturas Métricas del Software y Calidad en los Servicios del Software.

Con respecto a esta última asignatura -cuyo foco principal de la asignatura está en el Mantenimiento de Software y dentro de éste en la Ingeniería Reversa- su propuesta se fundamentó en dos cuestiones principales: a) la reciente promulgación de la Ley de Promoción Industrial del Software a nivel nacional¹; y b) La carencia de contenidos curriculares específicos vinculados al Mantenimiento de software en la currícula de Ingeniería en

Sistemas de Información de la mencionada Facultad Regional Córdoba.

El planteamiento de este proyecto surge como consecuencia de los resultados obtenidos durante el dictado de la asignatura antes mencionada en el ciclo lectivo 2008. Estos resultados plantearon la necesidad de profundizar los conocimientos sobre el estado actual del arte en materia de mantenimiento de software e ingeniería reversa y la aplicación de prácticas de Gestión de la Calidad sobre estos servicios.

RESUMEN

En el contexto de la cátedra de Calidad en Los Servicios del Software y del GIDICALSO, se visualiza la necesidad y oportunidad de avanzar con programas de Investigación y Desarrollo en lo referente a la Gestión de la Calidad en la temática del Mantenimiento del

Software en General y la Ingeniería Reversa en particular.

Por lo expuesto, el presente proyecto de investigación tiene por objetivo el Desarrollo de un Marco de Trabajo destinado al desarrollo conjunto de Universidad-Empresa de un framework específico de Gestión de Calidad para ingeniería reversa orientado a incrementar la eficiencia y productividad en esta actividad tan específica.

PALABRAS CLAVES

Calidad, Mantenimiento de Software, Ingeniería reversa, Framework.

INTRODUCCION

El mantenimiento de software es la disciplina referida a los cambios relativos a un sistema software después de su puesta en producción. En los últimos años como consecuencia del gran crecimiento de la industria del software a nivel nacional e internacional, la producción de productos software –que han invadido nuestra vida cotidiana- se ha incrementado y por ende la necesidad de “mantener” dicho software operativo después de su puesta en producción, es decir que tales sistemas cumplan con sus objetivos, en forma eficaz cuando no eficientemente, es decir que sean útiles. En la actualidad, el funcionamiento y uso correcto de un sistema software puede ser una cuestión de vida o muerte. También es importante destacar que, en el ciclo de vida de un sistema software, el mantenimiento supera en mucho más de lo deseable el 70% de su ciclo de vida y sus costos varían de un 40% a más del 70% del costo total en dicho ciclo, lo cual implica la urgente necesidad de encontrar formas de

eliminar o reducir los problemas relativos a esta etapa de dicho ciclo.[Grubb-Takang:2005]

Por otro lado, cualquier tipo de cambio debe ser precedido por una comprensión del sistema software. Este proceso involucra una gran proporción del tiempo invertido en llevar a cabo el cambio. Entre las principales razones para ello se puede mencionar: documentación no existente, desactualizada o incorrecta, la complejidad del sistema, la indisponibilidad de los desarrolladores originales del sistema software o la carencia del conocimiento suficiente del dominio por parte del encargado de mantenimiento. Una forma de abordar y aliviar estos problemas es la de realizar abstracciones a partir de información relevante del código fuente acerca del sistema. Esto se denomina Ingeniería Reversa.

Existen diferentes abordajes sobre la Ingeniería Reversa [Grubb-Takang:2005], [Tonella-Potrich:2005] y .[Eilam:2005], sin profundizar demasiado en la Gestión de la Calidad de dicho proceso en forma integradora y completa, es decir en forma sistémica.

Otros trabajos abordan la calidad sobre algunos aspectos particulares del proceso la Ingeniería Reversa, como por ejemplo: atributos de la calidad –aplicabilidad, extensibilidad y escalabilidad- [CMU/SEI-98-TR-005], Métricas para Ingeniería Reversa [Zhou et al:1999], cuestiones de Calidad para Ingeniería Reversa en Base de Datos: el método, adquisición de la semántica del dominio en forma minuciosa, y criterio de evaluación de performance [Chiang-Barron:1995] e Ingeniería de Reversa para Requerimientos destacando aspectos vinculados a simplificar la complejidad y mejora de la calidad del software, entre otros.

Lo anteriormente puntualizado genera la necesidad de profundizar el conocimiento del estado del arte en lo referente a Ingeniería

Reversa en general y las prácticas de Gestión de la Calidad en dicho ámbito en particular. Seguidamente, se prevé abordar una investigación sobre la aplicación de dichas prácticas en la Industria del software, inicialmente, a nivel local. Posteriormente, se deberá, en forma conjunta con la industria, identificar y analizar las buenas prácticas utilizadas en la actualidad, su eventual mejoramiento e identificación de nuevas prácticas requeridas para completar un marco de trabajo orientado a la Gestión integral de la Calidad en la Ingeniería Reversa.

LINEAS DE INVESTIGACION Y DESARROLLO

Profundización del estudio del estado del Arte en Mantenimiento de Software

Profundización del estudio del Estado del arte en Ingeniería Reversa

Aplicación de los principios de Gestión de la Calidad a la Ingeniería Reversa

Gestión de Riesgos en la Ingeniería reversa, como herramienta de prevención, aspecto este de vital importancia en los sistemas de Gestión de la Calidad.

Investigación y desarrollo de buenas prácticas para Ingeniería Reversa

Desarrollo de un Marco de Trabajo para la gestión de la calidad en Ingeniería reversa

RESULTADOS OBTENIDOS/ESPERADOS

Profundización del conocimiento en el dominio del conocimiento vinculado a la Ingeniería reversa

Transferencia a la cátedra de Calidad en los Servicios del Software de los conocimientos adquiridos y profundizados.

Trabajo conjunto del GIDICALSO con empresas de la Industria del Software de la provincia de Córdoba a los fines de desarrollar y especificar buenas prácticas de aplicación en proyectos de Ingeniería Reversa.

Desarrollo de un marco de trabajo orientado a incrementar la eficiencia y la productividad en los proyectos de ingeniería reversa de productos software sobre la base de la aplicación de un Sistema de gestión de la Calidad.

Transferencia de los resultados obtenidos en materia de investigación y desarrollo a Empresas vinculadas a la industria del software a nivel local, nacional e internacional, mediante cursos, talleres, seminarios, y publicaciones en congresos y revistas especializadas.

FORMACION DE RECURSOS HUMANOS

Este aspecto es fundamental ya que los resultados de las investigaciones y desarrollos obtenidos redundarán en una mayor cualificación profesional para quienes participen en ellos, como así también los destinatarios de las eventuales transferencias que se realicen a estudiantes, docentes, profesionales, colegios profesionales, empresas y el sector gobierno. Además de los integrantes de la cátedra de Calidad en los Servicios del Software, se pretende incorporar –de hecho el GIDISCALSO permanentemente lo hace- a estudiantes, graduados, docentes y profesionales para participar en los proyectos del Programa de Investigación que se

encuentra en pleno desarrollo y vinculado a la Ingeniería Reversa.

Propender a la formación de nuevos investigadores en el GIDICALSO.

Proveer de temas y de guía u orientación para la realización de tesis para quienes cursan la Maestría en Sistemas de Información que se dicta en la Facultad Regional Córdoba y otras regionales de la Universidad Tecnológica Nacional.

Incorporar, mediante convenios de colaboración, a profesionales de la Industria a los fines de especificar buenas prácticas que sean la base para el desarrollo de un marco de trabajo referido a la aplicación de un Sistema de Gestión de la Calidad orientado a la ingeniería reversa.

BIBLIOGRAFIA

[Grubb-Takang:2005]: Penny Grubb, Armstrong A. Takang; Software Maintenance Concepts and Practice Second Edition; World Scientific 2005.

[Tonella-Potrich:2005]: Paolo Tonella, Alessandra Potrich; Reverse Engineering of Object Oriented Code; Springer Science 2005.

[Eilam:2005]: Eldad Eilam; Reversing Secrets of Reverse Engineering; Wiley Publishing 2005.

[CMU/SEI-98-TR-005]: Scott Tilley; A Reverse Engineering Environment Framework; CMU/SEI-98-TR-005 ESC-TR-98-005; Software Engineering Institute; 1998.

[Zhou et al:1999]: Shikun Zhou, Hongji Yang, Paul Luker, Xudong He; A Useful Approach to Developing Reverse Engineering Metrics; IEEE, 1999.

[Chiang-Barron:1995]: Roger H. L. Chiang, Terence M. Barron; Quality Issues in Database Reverse Engineering: An Overview; IEEE, 1995.

ESTANDARIZACIÓN COBIT PARA EL CONTROL DE TECNOLOGÍAS INFORMÁTICAS (TI) DEL GOBIERNO DE CATAMARCA

Maria A. Barrera⁽¹⁾, Carolina I. Chayle⁽¹⁾⁽²⁾, Claudia M. Herrera⁽¹⁾, Andrea Rosatto (1)

⁽¹⁾ *Facultad de tecnología y Ciencias Aplicadas. Maximio Victoria 55. Catamarca*

⁽²⁾ *Facultad de Ciencias de la Salud. Maestro Quiroga S/N Ira. C. Catamarca*

maritabarrera@arnet.com.ar & cchayle@salud.unca.edu.ar

CONTEXTO

Los integrantes del proyecto se encuentran en la etapa de desarrollo de sus tesis de posgrado en el marco de la carrera de maestría en Ingeniería del Software en temas específicamente relacionados al área del proyecto: Standares de Auditoría en ámbitos gubernamentales, Proyectos de Risk Management y Gobierno electrónico, todos aplicados en el gobierno de la provincia de Catamarca.

Además, la Lic. María Alejandra Barrera y la Lic. Claudia Mabel Herrera son docentes de la cátedra de Auditoría de la carrera Ingeniería en Informática de la Facultad de Tecnología y La Lic. Carolina Irene Chayle es docente de la cátedra de Calidad de Software de la carrera Ingeniería en Informática de la Facultad de Tecnología, ambas cátedras están ligadas estrechamente al tema en cuestión. Andrea Rosatto es alumna de la carrera de Ingeniería en Informática y se encuentran en etapa de desarrollo de su tesina de grado investigando sobre Auditoría Web. Un modelo de implementación para la evaluación de calidad del sitio Web de la Municipalidad de San Fernando del Valle de Catamarca.

Así mismo es prioritario para el Departamento de Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa, transferir soluciones concretas a las problemáticas que se plantean a los profesionales y a los organismos provinciales, en relación con el aseguramiento de la calidad y evaluación de los riesgos de las TI, para garantizar la seguridad de la información y prevenir las posibles contingencias en el uso de TI.

RESUMEN

El vertiginoso avance tecnológico ha incrementado cada vez más la dependencia de las tecnologías de información (TI) en las organizaciones, el crecimiento de este fenómeno aumenta también los riesgos informáticos a los que se enfrentan. La adopción de estándares de control de TI son actualmente utilizados a nivel mundial y proporcionan a las organizaciones que los aplican un mejoramiento en el aseguramiento de la información y de los activos informáticos, mediante la

actualización de sus procesos. Sin embargo, es confusa la implementación de buenos controles de TI por parte de entidades comerciales, sin fines de lucro o gubernamentales. En el caso de entidades públicas es menester comprender su enfoque global de funcionamiento para poder focalizar los aspectos sustantivos de una reforma aplicable a cualquier nivel de Estado. A nivel nacional, la información muestra que pocas provincias han establecido las bases para evaluar los riesgos de TI en ámbitos gubernamentales. En el presente proyecto se busca evidenciar esta necesidad, suministrando los lineamientos a seguir para la aplicación del Standard COBIT, como herramienta de control de TI en el ámbito del Gobierno de la Provincia de Catamarca.

Palabras clave: Tecnologías de Información; Riesgos informáticos; Auditoría Informática; Standard COBIT; Sistemas de Control de Información.

1. INTRODUCCION

Las empresas públicas y privadas están valorando cada día más la creciente importancia que representa mantener sistemas informáticos seguros, confiables y confidenciales, que eviten o prevengan la ocurrencia de errores u operaciones ilegales a partir de debilidades en los sistemas de control.

El aseguramiento de la información es la base sobre la que se construye la toma de decisiones de una organización. Sin aseguramiento, las empresas no tienen certidumbre de que la información sobre la que sustentan sus decisiones sea confiable, segura y esté disponible cuando se la necesita. Muchas organizaciones reconocen estos beneficios potenciales, y por eso, las organizaciones exitosas comprenden y administran los riesgos asociados con la implementación de esta tecnología. Por lo tanto, los administradores deben tener una apreciación y un entendimiento básico de los riesgos y limitantes del empleo de TI para proporcionar la dirección efectiva y los controles adecuados a los fines de decidir la inversión

razonable en seguridad y control, tratando de lograr un balance entre riesgos e inversiones en ambientes de este tipo, frecuentemente impredecibles.

El sector público no queda aislado de esta problemática, por lo que necesita definir una estrategia general de desarrollo informático, fundamentalmente para obtener información oportuna y veraz, tanto en su faz operativa como gerencial.

Tal estrategia implica uniformar criterios y estandarizar las actividades del Estado bajo un mismo enfoque. Este hecho es fundamental para planificar racionalmente las acciones tendientes a cumplir con la agenda de gobierno, para tomar decisiones al más alto nivel y a su vez, mejorar la gestión de las organizaciones haciéndolas más eficaces y eficientes en el cumplimiento de sus metas.

Existe una creciente necesidad de las organizaciones en cuanto a la seguridad en los servicios de TI a través del control y la auditoría. Sin embargo, es confusa la implementación de buenos controles de TI en sistemas de negocios por parte de entidades comerciales, entidades sin fines de lucro o entidades gubernamentales. Esta confusión proviene de los diferentes métodos de evaluación, tales como ITSEC, TCSEC, evaluaciones ISO9000, nuevas evaluaciones de control interno COSO, etc., y en consecuencia, los usuarios necesitan una base general a ser establecida como primer paso.

En este sentido, COBIT (Control Objectives for Information and related Technology) complementa a los modelos más generales como COSO (EEUU), CoCo (Canadá) o Cadbury (Inglaterra) y ayuda a salvar las brechas existentes entre riesgos de negocio, necesidades de control y aspectos técnicos. Proporciona "prácticas sanas" a través de un marco referencial de dominios y procesos y presenta actividades en una estructura manejable y lógica. Dichas prácticas aspiran a optimizar el uso de los recursos disponibles, es decir, personas, instalaciones, tecnologías, sistemas de aplicación y datos, definiendo el marco de trabajo y el entorno tecnológico adecuado. (1)

Es importante mencionar aquí lo que sostiene el COBIT para organismos gubernamentales (2). El mismo define cuatro dominios para agrupar procesos de TI, concebir responsabilidades en una estructura organizacional y encuadrar los mismos según su ciclo de vida aplicable o su ciclo de administración. Tales dominios incluyen:

- **Planificación y Organización:** Vincula la identificación de la forma en que la Tecnología de Información puede contribuir más adecuadamente con el logro de los objetivos del gobierno. Precisa, planifica, comunica y administra la realización de la

visión estratégica de Tecnologías de Información. Define una correcta organización e infraestructura tecnológica.

- **Adquisición e Implementación:** Identifica, desarrolla o adquiere soluciones de Tecnologías de Información, y luego las implementa e integra a los procesos del gobierno. Incluye cambios y mantenimiento de sistemas existentes para garantizar su ciclo de vida.
- **Entrega y Soporte:** Cumple con la prestación efectiva de los servicios requeridos, que comprenden desde operaciones tradicionales, aspectos de seguridad y continuidad de servicios hasta la capacitación. Incluye el procesamiento real de los datos por los sistemas de aplicación.
- **Monitoreo:** Evalúa los procesos de Tecnologías de Información a medida que transcurre el tiempo para determinar su calidad y el cumplimiento de los requerimientos de control.

El monitoreo incluye la descripción de la regulación de las mejores prácticas de Auditoría en Informática, como así también, la definición de cómo administrar los riesgos de TI en el sector público en base a los estándares establecidos por organismos nacionales e internacionales.

En la República Argentina, puede señalarse que muy pocas provincias han intentado establecer las bases para evaluar los riesgos de TI en ámbitos gubernamentales. En la Provincia de Catamarca, el órgano rector de la seguridad y control de TI, es la recientemente creada Dirección Provincial de Gestión de la Información, que aún no ha establecido las normas necesarias para definir los estándares comunes a todas aquellas tareas que implican el uso de las mismas.

A los fines de garantizar la seguridad de la información y prevenir las posibles contingencias en el uso de TI, se propone asegurar la confidencialidad, confiabilidad y disponibilidad de los datos en el gobierno de la Provincia de Catamarca, a través de la aplicación del Standard COBIT. Para ello, se analizarán los distintos parámetros rectores y su mejor adaptación al manejo de TI en el ámbito específico del gobierno provincial, a los efectos de proporcionar los lineamientos a seguir para una aplicación exitosa de dicho estándar.

1) COBIT Marco, Comité de Dirección COBIT y la Information Systems Audit and Control Foundation, Buenos Aires, 1998.)

2) COBIT, Gobernabilidad, Control y Auditoría de Información y Tecnologías Relacionadas, Information Systems Audit and Control Foundation,

2. LINEAS DE INVESTIGACION Y DESARROLLO

Ingeniería del Software: Auditoría y Control de Tecnologías informáticas

3. RESULTADOS ESPERADOS y OBTENIDOS

En la Provincia de Catamarca, al igual que en la mayoría de las provincias argentinas, no se han establecido aún las bases para evaluar los riesgos de TI en ámbitos gubernamentales.

Sin información segura, confidencial, confiable y accesible, difícilmente el Gobierno Provincial pueda planificar racionalmente sus acciones. La aplicación de una adaptación adecuada del Standard COBIT para garantizar la seguridad de la información y prevenir las posibles contingencias en el uso de TI, permitirá mejorar la gestión de las organizaciones gubernamentales de Catamarca, haciéndolas más eficaces y eficientes en el cumplimiento de sus metas. La definición y aplicación de normas de Riesgos y Auditoría de TI en el ámbito del Gobierno de la Provincia de Catamarca, producirá un impacto positivo y tangible en los procesos de toma de decisión que requieren de información, tal como diversos antecedentes en el mundo y en la región lo demuestran.

El control de riesgos y seguridad en el manejo de TI, proporcionará al Gobierno de la Provincia de Catamarca la herramienta necesaria para planificar racionalmente sus acciones sobre bases más confiables. Dicho control, que incluye políticas, estructuras, prácticas y procedimientos organizacionales, es responsabilidad de la administración pública provincial.

El control interno de calidad de información permitirá optimizar el empleo de los recursos disponibles, los cuales incluyen: personal, instalaciones, tecnología, sistemas de aplicación y datos.

La adaptación de tales estándares para el manejo de TI en el ámbito específico del gobierno provincial, no sólo permitirá minimizar los riesgos y aumentar la seguridad en el manejo de TI, sino además, contribuirá a fijar los lineamientos a seguir en distintas provincias argentinas para garantizar mejores prácticas de manejo de datos en las esferas gubernamentales.

Resultados obtenidos

- Definición del Proceso de Administración de los Riesgos y Auditoría de TI en el ámbito del

Gobierno de la Provincia de Catamarca.

- Establecimiento de las pautas a seguir en la aplicación del Standard COBIT para garantizar la seguridad de la información y prevenir las posibles contingencias en el uso de TI por parte del Gobierno de la Provincia de Catamarca.
- Selección de criterios adecuados de evaluación de riesgos de TI.
- Definición de procedimientos a seguir en cada fase del proceso de administración de los riesgos de TI.
- Proporcionar normas de Riesgos y Auditoría de aplicación en el manejo de la información del Gobierno de la Provincia de Catamarca.
- Formación de recursos humanos en administración de riesgos y auditoría en TI.

4. FORMACION DE RECURSOS HUMANOS

El programa de capacitación y formación de recursos humanos, contempla las siguientes actividades:

- Incorporación de alumnos de los últimos años de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa., en calidad de auxiliares de investigación.
- Dirección de tesinas de grado de la carrera de Ingeniería en Informática de la Facultad de Tecnología y Ciencias Aplicadas de la UNCa. En este sentido la alumna e integrante del equipo del proyecto: Andrea Rosatto, se encuentra en etapa de elaboración del Trabajo Final de la carrera de Ingeniería en Informática denominado: Auditoría de Artefactos Web. Caso de Estudio: Pagina Web de la Municipalidad de San Fernando del Valle de Catamarca
- Participación de los integrantes del proyecto en cursos de actualización y posgrado en el área de estudio.
- Participación en talleres o workshops de herramientas informáticas relacionadas con el control y auditoría de TI.
- Celebración de convenios con entes estatales, como la Dirección Provincial de Gestión de la Información de la Provincia de Catamarca, la Municipalidad de la Capital y otras universidades, para capacitación de los recursos humanos en las herramientas normalizadas.
- Dictado de cursos de capacitación diseñados según demanda de la propia

Administración, con el objeto de transmitir las mejores prácticas a los agentes públicos asistentes.

Para garantizar la capacitación y actualización del equipo de investigación, así como la difusión de los avances y resultados logrados, se propuso la participación en eventos nacionales e internacionales de la especialidad, como congresos, simposios, seminarios y cursos.

5. BIBLIOGRAFIA

- COBIT Marco, Comité de Dirección COBIT y la Information Systems Audit and Control Foundation, Buenos Aires, 1998.
- COBIT, Gobernabilidad, Control y Auditoría de Información y Tecnologías Relacionadas, Information Systems Audit and Control Foundation, Edición Especial para Organismos Gubernamentales, Buenos Aires, 1998.
- COSO [Committee of Sponsoring Organisations of the Treadway Commission Internal Control-Integrated Framework, 1992]
- Prince, A. *Gobierno Digital, Primer Foro Gobierno Digital en Argentina*, Buenos Aires, Julio 2000.
- Herrera Cognetta, A. & M. A., Castro, Las TIC en la Administración Pública de Jujuy, *Investigaciones Docentes en Ingeniería*, Vol II, 987-9170-60 1, 580-585, 2006.
- <http://www.isaca.org>, 2007.
- ISO/IEC TR 13335 - Tecnologías de la información (TI) - Guía para la gestión de la seguridad de TI; Parte 4; capítulo 8.1.6.
- Piattini, M & E. del Peso, *Auditoría Informática. Un enfoque práctico*, Ed Alfaomega, Mexico 2005.
- Echenique Garcia, J. A., *Auditoría en Informática*, Ed. McGraw-Hill, Mexico 2004.
- Alvarez, B. R. & A. R., Garnacho, *Avances en Criptología y Seguridad de la Información*, Ed. Diaz de Santos, España, 2004.
- Izquierdo Duarte, F., Administración de Riesgos de TI, IX Encuentro Nacional y IV Internacional de Control Interno 2003.
- Herrera Cognetta, A. & M. A., Castro, Las TIC en la Administración Pública de Jujuy, *Investigaciones Docentes en Ingeniería*, Vol II, 2006.
- <http://www.isaca.org>, 2007.
- The Australian/New Zealand Joint Standards Committee AS/NZS 4360 Risk Management

Comprensión de Programas

Mario M. Berón, Pedro Rangel Henriques, Roberto Uzal y Maria João Varanda Pereira
Ingeniería del Software: Conceptos Métodos Herramientas en un Contexto de Ingeniería de Software en Evolución

Universidad Nacional de San Luis - Argentina
Program Comprehension by Visual Inspection and Animation
Universidade do Minho – Braga – Portugal
Departamento de Informática
Universidad Nacional de San Luis

CONTEXTO

Las líneas de investigación descritas en este artículo se encuentran enmarcadas en el contexto del proyecto: *Ingeniería del Software: Conceptos Métodos Herramientas en un Contexto de Ingeniería de Software en Evolución* de la Universidad Nacional de San Luis y del Proyecto *Program Comprehension by Visual Inspection and Animation* desarrollado en la Universidade do Minho, Portugal. Las líneas aquí presentadas surgen a partir de la realización de una tesis doctoral auspiciada por ambos proyectos y de los vínculos de investigación establecidos entre la Universidad Nacional de San Luis y la Universidade do Minho, Portugal.

RESUMEN

La *Comprensión de Programas* (CP) es una disciplina de la Ingeniería de Software cuyo objetivo es proveer *Modelos, Métodos, Técnicas y Herramientas* para facilitar el estudio y entendimiento de programas.

La construcción de estos productos de comprensión implica el estudio de disciplinas tales como *Modelos Cognitivos, Visualización de Software, Estrategias de Interrelación de Dominios y Métodos de Extracción de la Información*.

En este artículo se presentan líneas de investigación cuyos objetivos consisten en el análisis los productos de comprensión existentes y en la construcción de otros nuevos basados en los conceptos comunes a las grandes áreas mencionadas en el párrafo anterior.

Palabras Claves: Comprensión de Programas, Modelos, Métodos, Técnicas, Herramientas.

1. INTRODUCCION

La Comprensión de Programas es un área de la Ingeniería del Software destinada a elaborar

Modelos, Métodos, Técnicas y Herramientas, basados en un proceso cognitivo y de ingeniería con el objetivo de facilitar el entendimiento de software. El proceso cognitivo implica el estudio y análisis de las fases y pasos seguidos por los programadores para entender programas. Este tema es abordado a través de la investigación de los *Modelos Cognitivos* de CP.

El proceso de ingeniería incluye investigaciones sobre *Visualización de Programas, Estrategias de Interrelación de Dominios y Métodos de Extracción de la Información*. En este contexto interdisciplinario, el desarrollo de productos de comprensión se basa en encontrar el común denominador a esas tres grandes disciplinas .

Actualmente existen muchos sistemas destinados a facilitar el entendimiento de software. Sin embargo, en muchas situaciones, no es claro como las teorías cognitivas, estrategias de visualización y extracción de la información se instancian en esas herramientas. Además, esas aplicaciones están centradas en analizar y presentar el código fuente del programa limitando CP a la Inspección de código. Como se verá en el desarrollo de este artículo, la Comprensión de Programas implica: i) plasmar claramente los conceptos de sus principales áreas en los productos de comprensión y ii) encontrar relaciones entre el *Dominio del Problema* y el *Dominio del Programa*, es decir, detectar las componentes de software utilizadas por el sistema para producir su salida [Bro78].

Este artículo está organizado de la siguiente manera. La sección 2 describe sintéticamente las líneas de investigación desarrolladas en nuestro proyecto. La sección 3 expone brevemente los resultados obtenidos en cada una de las temáticas después de un año de trabajo. Finalmente, la sección 4 menciona las estrategias planificadas para la formación de recursos humanos.

2. LINEAS DE INVESTIGACION

A continuación se exponen las líneas de investigación desarrolladas en nuestro proyecto.

2.1 Modelos Cognitivos

El término Modelo Cognitivo hace referencia a las estructuras de la información y estrategias de estudio usadas por los programadores para entender programas [BVDB93]. Los modelos cognitivos constan de diferentes componentes. Ellas son: el *Conocimiento*, un *Modelo Mental* y un *Proceso de Asimilación*.

Existen dos tipos de Conocimiento, el *Interno*, compuesto por el conjunto de conceptos y relaciones que conforman la estructura de conocimiento del programador; y el *externo* cuyos componentes son los nuevos conceptos proporcionados por el sistema de estudio.

El *Modelo Mental* se define como la representación mental que tiene el programador del sistema. El *Grafo de Funciones*, *Comunicaciones de Módulos*, etc. son posibles modelos mentales. Finalmente, el *Proceso de Asimilación* describe la estrategia utilizada por el programador para entender programas. Esta puede ser *top-down*, *bottom-up* o *híbrida*. Teniendo en cuenta esos elementos y sus relaciones muchos autores sostienen que: *un programador entiende un programa cuando puede encontrar las componentes de software usadas para producir la salida del sistema* [Bro82]. El camino adecuado para alcanzar este objetivo consiste en: i) Proveer representaciones para los dominios del problema y programa; ii) Definir un proceso que permita unir ambas representaciones. Los pasos mencionados previamente conforman la base para construir verdaderas aplicaciones de comprensión de programas.

Esta línea de investigación estudia el interjuego entre las diferentes componentes de los modelos cognitivos con el objetivo de elaborar estrategias que faciliten el entendimiento del software.

2.2 Visualización de Software

La Visualización de Software [BkK01a,Che06, Hen01] es una disciplina de la Ingeniería del Software cuyo objetivo es mapear ciertos aspectos de software en una o mas representaciones multimediales. Para alcanzar este objetivo es necesario la interacción con otras áreas del

conocimiento tales como: *Diseño Gráfico*, *Psicología Cognitiva* y otras disciplinas directamente relacionadas con la elaboración de efectos multimediales [Ext02]. Si la visualización esta orientada a la comprensión de programas, el principal desafío consiste en construir vistas que permitan relacionar el Dominio del Problema con el Dominio del Programa.

Existen innumerables herramientas de visualización de programas que, según sus autores tienen como finalidad facilitar la comprensión de programas [BkK01b]. Sin embargo, la gran mayoría propone visualizaciones concernientes con el Dominio del Programa (por ejemplo *Funciones*, *Módulos*, *Variables*, etc.) dejando de lado dos importantes componentes como lo son el Dominio del Problema y su relación con el Dominio del Programa.

Esta línea de investigación estudia las visualizaciones que relacionan el Dominio del Problema con el Dominio del Programa y la integración de este tipo de vista con las visualizaciones tradicionales [LELP06].

2.3 Métodos de Extracción de la Información

El desarrollo de técnicas de extracción de la información estática es importante porque permite recuperar todos los atributos de los objetos definidos en el programa.

El desarrollo de técnicas de extracción de la información dinámica es relevante porque posibilita conocer cuales son las componentes del programa utilizadas para una ejecución específica del sistema. Ambos tipos de información son necesarias para la elaboración de técnicas de comprensión que faciliten el estudio de sistemas grandes. Esta tarea es uno de los principales objetivos de nuestro grupo de investigación.

Esta línea de investigación tiene por objetivo el estudio y elaboración de técnicas de extracción de información estática y dinámica desde los sistemas de software [CCMT93,CHZ+07,KGG05].

2.4 Estrategias de Interrelación de Dominios

Una de las formas de facilitar la comprensión de grandes sistemas consiste en relacionar el Dominio del Problema (es decir el comportamiento del programa) con el Dominio del Programa (o sea la operación del programa). Esto se debe a que dicha relación permite que el programador pueda localizar rápidamente los objetos del programa que se

utilizaron para una funcionalidad específica del sistema de estudio.

Si el sistema que se está analizando es pequeño (5 Kloc o menos) entonces las bondades de este tipo de relación no se pueden apreciar claramente. Pero si el sistema es de gran envergadura la relación antes mencionada disminuye el esfuerzo del programador porque él puede identificar rápidamente los objetos sobre los cuales debe concentrar su estudio. Por otra parte, los costos implicados en la modificación, mantenimiento o evolución del sistema se disminuyen debido a que se decreta el tiempo requerido por el programador para la realización de la tarea. Es importante mencionar que son casos excepcionales las herramientas de comprensión profesionales que incluyen este tipo de relación. Esta característica hace que esta temática sea de suma importancia en el contexto de CP debido a la posibilidad que se presenta para realizar contribuciones. Además del desarrollo de estrategias para interconectar los Dominios del Problema y Programa se estudian formas de integrar esta importante relación con vistas tradicionales como por ejemplo: *Grafo de Funciones*, *Grafo de Módulos*, *Grafo de Tipos*, etc.

2.5 Evaluación de Herramientas de Comprensión

En nuestro grupo de investigación se ha realizado un estudio integral de las herramientas de comprensión abarcando todas las temáticas involucradas en su creación, las cuales han sido sintéticamente descritas en las secciones previas. Cada temática introduce una serie de criterios que permiten evaluar la calidad y pertinencia de cada uno de los subsistemas que componen a una herramienta de comprensión [GC99]. En esta línea de investigación se estudia la forma más adecuada de integrar todos esos criterios en un marco de evaluación que posibilite un análisis acabado de las herramientas de comprensión.

3. RESULTADOS OBTENIDOS/ESPERADOS

En las siguientes sub-secciones se describen los resultados obtenidos en cada una de las líneas de investigación descritas en la sección 2.

3.1 Modelos Cognitivos

Los resultados obtenidos en este contexto están relacionados con la formalización de los conceptos utilizados para describir las distintas formas que el

programador utiliza para comprender programas. Estos conceptos son ambiguos y difíciles de entender por parte de la comunidad de Ciencias de la Computación porque los mismos son descritos por profesionales de otras áreas, como por ejemplo la Psicología Cognitiva. Por esta razón se realizó un esfuerzo en tratar de formalizar los conceptos dándoles un significado más preciso. Estas formalizaciones tienen como objetivo permitir el desarrollo de técnicas que posibiliten traducir el código fuente del sistema en una especificación basada en modelos con alto poder descriptivo. Está última tarea (el desarrollo de técnicas) es un resultado esperado para los futuros pasos de la investigación en esta área. El lector interesado en ver las formalizaciones de los conceptos puede leer [BHVU08].

3.2 Visualización de Software

Los *Sistemas de Visualización de Software Orientados a la Comprensión de Programas* son una nueva clase de sistemas que surge a partir de la investigación de los *Sistemas de Visualización* en general. La principal diferencia entre este nuevo tipo de sistemas y los tradicionales radica en que los *Sistemas de Visualización Orientados a la Comprensión de Programas* contemplan el Dominio del Problema y su relación con el Dominio del Programa. Esta peculiaridad permitió detectar la ausencia, en las clasificaciones actuales, de criterios para caracterizar el Dominio del Problema y la relación existente con el Dominio del Programa [Mye90,PSB92]. Esta debilidad permitió desarrollar una nueva clasificación de los *Sistemas de Visualización Orientados a la Comprensión* que define nuevos criterios que solucionan el inconveniente presentado previamente. Es importante notar que este resultado es relevante porque es original y por consiguiente es una contribución en el contexto de CP. El lector interesado en conocer esta nueva clasificación puede leer [BHVU08].

Los trabajos futuros en este ámbito están relacionados con la creación de estrategias de visualización que cumplan con la mayoría de los criterios declarados en la nueva taxonomía elaborada por nuestro grupo de investigación. Además se pretende que dichos esquemas de visualización sean incorporados en las herramientas de comprensión de nuestra propiedad con el objetivo de realizar estudios experimentales que permitan

analizar el grado de mejora alcanzada por los programadores en el proceso de comprensión.

3.3 Extracción de la Información

Para la extracción de la información dinámica se definió un esquema de *Instrumentación de Código*. Esta técnica consiste en insertar sentencias dentro del código fuente del sistema de estudio con la finalidad de recuperar las componentes del programa que se utilizaron para producir la salida. Para implementar una estrategia de estas características es necesario responder a los siguientes interrogantes: i) ¿Cuáles son los puntos del programa candidatos instrumentar? y ii) ¿Qué información debe ser recuperada? Teniendo en mente esas preguntas, se seleccionaron como puntos de inspección el inicio y fin de cada función del sistema. La razón de esta decisión se basa en que en esos lugares del programa se puede obtener información resumida acerca de las componentes del programa. Por ejemplo, se pueden conocer las funciones utilizadas, sus parámetros y si se desea ser más preciso los datos (valores de las variables globales y de los parámetros) que son utilizados por la función. Esta aproximación, aún recuperando parte de las operaciones y datos utilizados por el programa, tiene el inconveniente de extraer una enorme cantidad de información. Por este motivo, es necesario el empleo de técnicas de control de las iteraciones. Una de las estrategias elaboradas por nuestro grupo de investigación consistió en insertar sentencias antes, dentro y después de las iteraciones. Las sentencias previas a la iteración colocan en una pila de control el número de veces que las funciones invocadas dentro de la iteración pueden ser recuperadas. Las sentencias dentro del loop decrementan ese valor en uno. Cuando el valor del tope de la pila es cero las sentencias insertadas no recuperan mas información. Finalmente, las instrucciones insertadas después de la iteración suprimen el valor del tope de la pila. El lector interesado en conocer todos los pormenores de la estrategia de instrumentación puede leer [BHVU08]. Los trabajos futuros en esta temática se relacionan con el desarrollo de nuevos esquemas de instrumentación de código y técnicas de análisis de la información dinámica.

3.4 Estrategias de Interrelación de Dominios

Se desarrollaron dos técnicas para la interrelación de dominios, que utilizan los conceptos extraídos de las

investigaciones presentadas en las secciones previas. Una de ellas denominada SVS (*Simultaneous Visualization Strategy*) se basa en la ejecución paralela del sistema instrumentado y del administrador de funciones de inspección (un programa que implementa las acciones de las sentencias incorporadas en el código fuente del sistema). Esta característica permite que las componentes de software usadas sean mostradas cuando el sistema está en ejecución.

La otra estrategia es BORS (*Behavioral-Operational Relation Strategy*), este procedimiento, al igual que SVS, utiliza la información reportada por el esquema de instrumentación pero de una manera diferente. BORS requiere que el sistema sea ejecutado, después de eso se procesa la información y se construyen algunas estructuras de datos útiles para construir explicaciones, como por ejemplo el árbol de ejecución de funciones. Luego se realizan algunas consultas sobre dichas estructuras para recuperar alguna información relacionada con los objetos del dominio del problema. Como trabajo futuro se pretende el desarrollo de otras estrategias de interrelación de dominios que permitan extraer más información de tiempo de ejecución.

3.5 Evaluación de Herramientas de Comprensión

Los resultados obtenidos en esta línea de investigación no son tan avanzados como los obtenidos en las otras líneas descriptas en las subsecciones precedentes. No obstante se puede mencionar que se optó por adaptar una metodología de evaluación utilizada para sitios web que usa una serie de criterios definidos para este tipo de software y define un proceso de ranking que posibilita establecer que sitio web es el mejor [OGLR99,OR02]. La elección de este camino se debió a que el proceso de ranking de esta metodología hace un fuerte uso de los atributos del objeto de estudio los cuales son difíciles de definir. Como la definición de los atributos de las herramientas de comprensión ha sido realizada para cada uno de sus componentes en las otras líneas de investigación que conforman nuestro proyecto, entonces los mismos sirven como entrada al proceso de ranking definido por la metodología de evaluación seleccionada. Se espera poder aplicar este proceso de evaluación a varias herramientas de comprensión para establecer un ranking y analizar

los resultados obtenidos con los nuevos criterios definidos por nuestro grupo investigación.

4. FORMACIÓN DE RECURSOS HUMANOS

Los resultados alcanzados en este momento están relacionados con la realización de una tesis doctoral denominada: “Inspección de Programas para Interconectar las Vistas Operacional y Comportamental para la Comprensión de Programas”, desarrollada con entre la Universidad Nacional de San Luis y la Universidad de Minho, Portugal. Además de la realización de la tesis doctoral se proyectan la elaboración de diferentes publicaciones nacionales e internacionales como así también el desarrollo de tesis de licenciatura, maestría y doctorado. Todas las tareas mencionadas previamente serán realizadas con el auspicio de la UNSL y la Universidade do Minho.

5. BIBLIOGRAFIA

- [BkK01a] Sarita Bassil and Rudolf k. Keller. A Qualitative and Quantitative Evaluation of Software Visualization Tools. Proc. of the IEEE Symposium on Information Visualization, pages 69–75, 2001.
- [BkK01b] Sarita Bassil and Rudolf k. Keller. Software Visualization Tools: Survey and Analysis. Proc. of Program Comprehension, 2001. IWPC 2001, pages 7–17, 2001.
- [Bro78] R. Brook. Using a behavioral theory of program comprehension in software engineering. Proceedings of the 3rd international conference on Software engineering, pages 196–201, 1978.
- [Bro82] R. Brook. A theoretical analysis of the role of documentation in the comprehension of computer programs. Proceedings of the 1982 conference on Human factors in computing systems, pages 125–129, 1982.
- [BVDB93] K. Bertels, P. Vanneste, and C. De Backer. A cognitive approach to program understanding. Reverse Engineering, 1993., Proceedings of Working Conference on, pages 1–7, 1993.
- [CCMT93] G. Canfora, A. Cimble, M. Munro, and C. Taylor. Extracting abstract data types from C programs: A case study. Proceedings, 12th IEEE International Conference on Software Maintenance, 27:100–209, 1993.
- [Che06] Chaomei Chen. Information Visualization. Springer Verlag, 2006.
- [CHZ+07] Bas Cornelisse, Danny Holten, Andy Zaidman, Leon Moonen, Jarke Wijk, and Arie Deursen. Understanding execution traces using massive sequence and circular bundle views. In Delft University of Technology Software Engineering Research Group Technical Report Series, pages 2–9. Delft University, 2007.
- [Ext02] C. Exton. Constructivism and program comprehension strategies. Program Comprehension, 2002. Proceedings. 10th International Workshop on, pages 281–284, 2002.
- [GC99] Gerald C. Gannod and Betty H. C. Cheng. A framework for classifying and comparing software reverse engineering and design recovery techniques. In WCRE '99: Proceedings of the Sixth Working Conference on Reverse Engineering, page 77, Washington, DC, USA, 1999. IEEE Computer Society.
- [BHVU08] Mario Berón, Henriques Pedro, Maria João Varanda, Roberto Uzal. Inspección de Programas para Interconectar las Vistas Comportamental y Operacional para la Comprensión de Programas. Reporte de Tesis Doctoral. UNSL.
- [Hen01] Gómez Henriques. Software Visualization: an Overview. Informatik, 2:4–7, 2001.
- [KGG05] A. Kuhn, O. Greevy, and T. Girba. Applying Semantic Analysis to Feature Execution Traces. Program Comprehension through Dynamic Analysis, 1:48–53, 2005.
- [LELP06] W. Lowe, M. Ericsson, J. Lundber, and T. Panas. Software Comprehension-Integrating Program Analysis and Software Visualization. Technical Report, 2006.
- [Mye90] B. Myers. Taxonomies of Visual Programming and Program Visualization. Journal of Visual Languages and Computing, 1(1):97–123, 1990.
- [OGLR99] Luis Olsina, Daniela Godoy, Guillermo Lafuente, and Gustavo Rossi. Assessing the Quality of Academic Web Sites: a Case Study. New Review Hypermedia Journal, 05:81– 103, 1999.
- [OR02] Luis Olsina and Gustavo Rossi. Measuring Web Application Quality with WebQEM. IEEE MultiMedia, 2002, 09(4):20–29, 2002.
- [PSB92] BA Price, IS Small, and RM Baecker. A taxonomy of software visualization. System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on, 2, 1992.

PROCESO DE AGREGACIÓN PARA ESTUDIOS EXPERIMENTALES DE INFORMATICA

Enrique Fernández^{1,2}, Patricia Pesado^{1,3}, Ramón García-Martínez^{1,4}, Oscar Dieste^{1,5}

¹ Programa de Doctorado en Ciencias Informáticas. Facultad de Informática. UNLP

² Centro de Ingeniería de Software Ingeniería del Conocimiento. Escuela de Postgrado. ITBA

³ Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP - CIC

⁴ Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA

⁵ Grupo de Ingeniería de Software Experimental. Facultad de Informática. UPM

enfernan@itba.edu.ar, ppesado@lidi.info.unlp.edu.ar, rgarciamar@fi.uba.ar

CONTEXTO

El Proyecto articula líneas de trabajo de los siguientes Proyectos: [a] “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaria de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires; y [b] “Caracterización de Comportamientos Sociales Mediante Métodos de Explotación de Información” con financiamiento solicitado al Programa Ibero Americano de Ciencia y Tecnología para el Desarrollo. CYTED 2008-2011 / P608RT0404 a radicar en el Grupo de Ingeniería de Software Experimental de la Facultad de Informática de la Universidad Politécnica de Madrid.

RESUMEN

El incremento en la cantidad de estudios experimentales que se desarrollan en Informática ha traído aparejado un nuevo problema, que hacer cuando dos o más estudios que analizan el mismo tema no dan el mismo resultado?, este hecho puede solucionarse mediante el desarrollo de una síntesis cuantitativa de resultados. Este proyecto propone desarrollar un procedimiento de síntesis cuantitativa acorde a las características que actualmente presenta la Informática (pocos estudios, en general, hechos con pocos sujetos experimentales), el cual dista bastante de los propuestos hasta la fecha que han sido tomados de la Medicina.

Palabras clave: *Síntesis Cuantitativa, Agregación, Meta-análisis*

1. INTRODUCCION

Desde hace varios años, la cantidad de estudios experimentales en Informática se ha incrementado significativamente. Estos experimentos abarcan los más variados temas, como ser: el desempeño de las técnicas de prueba [Sabaliauskaitė, *et al*, 2004] o la educación de requisitos [Browne y Rogich, 2001]

entre otros. Si bien los experimentos aportan conocimientos interesantes en cada caso, en general son pequeños [Guerini, 2008], por ello para que la información que aportan sea valiosa los resultados deben agregarse para poder obtener conclusiones avaladas con la mayor evidencia empírica posibles.

La agregación de experimentos consiste en combinar los resultados de varios experimentos que analizan el comportamiento de un par de tratamientos específico, en un contexto determinado, para obtener un *único resultado final*. El nuevo resultado será *más general y fiable que los resultados individuales*, por que el mismo estará sustentado por un mayor nivel de evidencia empírica [Cochrane, 2008].

Es importante destacar que para que los resultados de un proceso de agregación sea realmente fiable, el mismo debe evitar caer en problemas de sesgos vinculados a la búsqueda y selección de experimentos. Para ello, en general, los procesos de agregación de resultados se incluyen dentro de una Revisión Sistemática (RS). Una SR es un procedimiento que aplica estrategias científicas para aumentar la fiabilidad del proceso de recopilación, valoración crítica y agregación de los estudios experimentales relevantes sobre un tema [Goodman, 1996].

En la Informática las RS fueron propuestas por [Basili, *et al*, 1996], desde entonteces varios autores abordaron el tema. A modo de ejemplo podemos citar al procedimiento de revisión sistemática desarrollado por [Kitchenham, 2004], el cual surge de una adaptación de los procesos de revisión sistemática desarrollados en medicina, donde se contemplan, entre otras cosas, diversos niveles de calidad de experimento acorde al contexto experimental de la Informática. En dicho procedimiento, se recomienda el uso del método estadístico Diferencia Medias Ponderadas (DMP) [Hedges y Olkin, 1985] para la agregación de los resultados de los experimentos.

Además de Basili y Kitchenham, existen otros autores que han trabajado en la aplicación de las revisiones sistemáticas en Informática, a modo de ejemplo podemos citar la revisión sistemática desarrollada recientemente por [Dyba, *et al*, 2007], donde se identifican 11 estudios experimentales

vinculados al desempeño de los programadores cuando trabajan de a pares, los cuales se Agregaron aplicando el método de Diferencias Medias Ponderadas como se sugiere en [Kitchenham, 2004]. Ahora bien, no todas las RS hechas dentro del ámbito de la Informática tuvieron éxito, a modo de ejemplo podemos citar a: [Banker y Keremer, 1989; Shull, *et al*, 2003; Hu, 1997; Wohlin, *et al*, 2003; Juristo, *et al*, 2004; Jørgensen, 2004] donde, si bien los autores pudieron desarrollar el procedimiento de búsqueda y selección de experimentos, la combinación de los mismos mediante Diferencias Medias Ponderadas resultó impracticable. Los principales motivos fueron:

- Escasez de experimentos, replicaciones y homogeneidad entre los mismos. [Davis, *et al*, 2006; Miller, 2000].
- Carencia de estándares para reportes de experimentos. Por ejemplo, [Burton, *et al*, 1990] no publican varianzas y [Crandall Klein, *et al*, 1989] ni siquiera reporta las medias de los resultados experimentales.
- Falta de estandarización de las variables respuesta. Por ejemplo, los trabajos de [Agarwal, *et al*, 1990] y [Woody, *et al*, 1996] utilizan diferentes variables respuesta para analizar un mismo aspecto, lo cual hace que estos experimentos no puedan ser agregado.

Por lo expuesto, surge la necesidad de desarrollar un procedimiento de agregación de estudios experimentales acorde a las características del actual contexto experimental de la Informática.

2. LINEAS DE INVESTIGACION y DESARROLLO

En el marco de este proyecto se investigará:

- [a] Las características de los estudios experimentales hechos en Informática.
- [b] Diferentes técnicas y estrategias de agregación usadas en otras áreas de la ciencia.
- [c] La fiabilidad de las distintas técnicas de agregación identificada dentro de los parámetros estándar de la Informática.

La aplicación de una estrategia de agregación para el dominio Informático que considere el uso de las técnicas de agregación de forma conjunta y combinada en lugar de aplicarlas de forma excluyente y aislada.

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto tiene como objetivo general definir una estrategia de agregación de estudios experimentales en Informática que permita combinar los resultados de un conjunto estudios experimentales, para obtener piezas de conocimientos representativas a nivel general.

El proyecto se enfocará a construir una estrategia de agregación que complemente a los protocolos de Revisiones Sistemáticas para Informática (por ejemplo: [Kitchenham, *et al*, 2004]), el cual contemplará el uso combinado de varias técnicas de agregación. Esta combinación de técnicas permitirá generar conclusiones con diversos niveles de fiabilidad, los cuales estarán relacionados con la fiabilidad de la técnica y las características de los estudios combinados mediante las mismas. Se considera que el proyecto implica investigación básica en la definición del proceso de agregación para Informática.

Entre las cuestiones particulares que se abordarán en el proyecto se encuentran: la caracterización de los estudios experimentales hechos en el ámbito de la Informática, la identificación de los problema que habitualmente suceden cuando se desarrollan revisiones sistemáticas dentro del ámbito de la Informática, la identificación de técnicas de agregación aplicadas en otras áreas de la ciencia, la determinación de la fiabilidad de las distintas técnicas de agregación en el contexto de la experimentación en Informática, y el establecimiento de un proceso de agregación con varios niveles de evidencia acorde a las características de los experimentos hechos en Informática.

Se espera que este proyecto contribuya a solucionar el problema de agregación de estudios experimentales que actualmente sufren las revisiones sistemáticas en informática, permitiendo de esta forma que las piezas de conocimientos generadas empíricamente puedan ser consideradas de aplicación general como sucede en las ciencias mas maduras.

A la fecha se ha validado la necesidad de desarrollar un procedimiento [Fernández; 2007], se ha presentado una versión preliminar del procedimiento en [Fernández; et al; 2007], se han desarrollado trabajos vinculado al análisis de la calidad experimental de los trabajos hechos en Informática en [Guerini; *et al.*, 2006; Guerini, 2008], se está evaluando la fiabilidad de diferentes métodos de síntesis cuantitativa [Pollo Cattaneo; 2009] para determinar en que condiciones aplicar cada uno, se ha desarrollado un trabajo de validación del procedimiento de agregación [Malcrida; 2009] mediante el agregado de 25 experimentos vinculados al análisis de métodos de inspección de código. Se está trabajando en el desarrollo de un aplicativo que permita seleccionar de forma automática estudios experimentales [Caimer, 2009] el cual será validado mediante el análisis de estudios vinculados a patrones de diseño, los cuales serán agregados en [González, 2008]. Por último, se está trabajando en una tesis doctoral con el objeto de describir en detalle las pautas para el agregado de experimentos en Informática [Fernández; 2008].

4. FORMACION DE RECURSOS HUMANOS

En el marco de este proyecto se esta desarrollando una tesis doctoral y cuatro tesis de maestria (una de ellas defendida y otra a la espera de defensa). Esta línea vincula al LIDI de la Facultad de Informática de la Universidad Nacional de La Plata, al Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires y al Grupo de Ingeniería de Software Experimental de la Facultad de Informática de la Universidad Politécnica de Madrid.

5. BIBLIOGRAFIA

- Agarwal, R.; Tanniru, M.; 1990; *Knowledge Acquisition Using Structured Interviewing: An Empirical Investigation*; Journal of Management Information System, M.E. Sharpe; Vol. 7 N. 1
- Banker and Keremer; 1989; *Scale economies in new software development*. IEEE Transactions on Software Engineering. (15): 10, pp. 1199-1205.
- Basili, V.; Green, S.; Laitenberger, O.; Lanubile, F.; Shull, F.; Sorumgaard, S.; Zelkowitz, M.; 1996; Packaging researcher experience to assist replication of experiments ISERN Meeting
- Browne, G.; Rogich, M.; *An empirical Investigation of User Requirements Elicitation: Comparing the Effectiveness of Prompting Techniques*; Journal of management Information System; Spring 2001; Vol. 17 N. 4
- Burton, A., Shadbolt, N., Rugg, G. y Hedgecock, A.; 1990. *The Efficacy of Knowledge Elicitation Techniques: A Comparison Across Domains and Level of Expertise*. Knowledge Acquisition 2(2): 167-178.
- Caimer, P; 2009. *Uso de minería de texto para identificación de trabajos experimentales*. Proyecto de tesis de Maestría, ITBA.
- Cochrane; 2008; *Curso Avanzado de Revisiones Sistemáticas*; www.cochrane.es/?q=es/node/198
- Crandall Klein, B. y Asociados; 1989. *A Comparative Study Of Think-Aloud And Critical Decision Knowledge Elicitation Method*. SIGAR Newsletter, April 1989, Number 108, Knowledge Acquisition Special Issue, páginas 144-146.
- Davis, A.; Dieste o.; Hickey, A.; Juristo, N.; Moreno, A.; 2006; *Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review*; 14th IEEE International Requirements Engineering Conference (RE'06) pp. 179-188
- Dyba, T., Aricholm, E.; Sjoberg, D.; Hannay J.; Shull, F.; 2007; *Are two heads better than one? On the effectiveness of pair programming*. IEEE Software; 12-15.
- Fernández, E. 2007. *Aggregation Process with Multiple Evidence Levels for Experimental Studies in Software Engineering*. Proceedings 2nd International Doctoral Symposium on Empirical Software Engineering. Pag. 75-81. ISBN 978-84-690-7340-7.
- Fernández, E., Britos, P., Dieste, O., García-Martínez, R. 2007. *Proceso de agregación con múltiples niveles de evidencia para estudios experimentales en Informática*. Proceedings IX Workshop de Investigadores en Ciencias de la Computación. Pag. 487-491. ISBN 978-950-763-075-0.
- Fernández, E. 2008. *Desarrollo de un proceso de agregación para estudios experimentales de informatica*. Proyecto de Tesis Doctoral en Ciencias Informáticas, Universidad Nacional de La Plata.
- Glass, G; 1976; *Primary, secondary, and meta-analysis of research*. Educational Researcher 5: 3-8
- González, S; 2008. *Revisión y agregación de estudios experimentales vinculados a patrones de diseño*. Proyecto de tesis de Maestría, ITBA.
- Goodman C.; 1996; *Literature Searching and Evidence Interpretation for Assessing Health Care Practices*; SBU; Stockholm.
- Guerini, M. 2008. *Revisión de Resultados Experimentales en Técnicas de Prueba y de Educción de Conocimientos*. Tesis de Magister en Ingeniería del Software. ITBA
- Guerini, M., Fernández, E., Britos, P., García-Martínez, R. 2008. *Revisión de Resultados Experimentales en Técnicas de Prueba y de Educción de Conocimientos*. Proceedings X Workshop de Investigadores en Ciencias de la Computación. Pág. 471-475. ISBN 978-950-863-101-5.
- Guerini, M., Fernández, E., Ochoa, M., Merlino, H., Diez, E., Britos, P. y García-Martínez, R. 2006. *Revisión de Resultados Experimentales sobre Performance de Técnicas Pruebas de Software*. XII Congreso Argentino de Ciencias de la Computación. San Luis. Argentina
- Hedges, L.; Olkin, I.; 1985; *Statistical methods for meta-analysis*. Academic Press.
- Hu, Q.; 1997; *Evaluating Alternative Software Production Function*. IEEE Transactions on Software Engineering. (23): 6, pp. 379-387.
- Jørgensen, M.; 2004; *A Review of Studies on Expert Estimation of Software Development Effort*. Journal of Systems and Software. (70): 1-2, pp. 37-60.
- Juristo, N.; Moreno, A. M., and Vegas, S.; 2004; *Reviewing 25 Years of Testing Technique Experiments*. Journal of Empirical Software Engineering; 9(1 - 2):7-44.
- Kitchenham, B. A.; 2004; *Procedures for performing systematic reviews*. Keele University; TR/SE-0401. Keele University Technical Report.
- Malacrida, J; 2009. *Revisión y agregación de estudios experimentales vinculados Técnicas de Inspección*. Tesis de Maestría pendiente de defensa, ITBA.

- Miller, J.; 2000; *Applying Meta-analytical Procedures to Software Engineering Experiments*. Journal of Systems and Software. (54): 1, pp. 29-39.
- Pollo Cattaneo, F; 2009. *Análisis de precisión de técnicas de agregación en contextos experimentales poco maduros*. Tesis de Maestría pendiente de defensa, ITBA.
- Sabaliauskaite, G., Kusumoto, S., Inoue, K.; 2004; *Assessing defect detection performance of interacting teams in object-oriented design inspection*; Information and Software Technology 46 (2004) 875–886; Available online at: www.sciencedirect.com
- Shull, F.; Carver, J.; Travassos, G. H.; Maldonado, J. C.; Conradi, R., and Basili, V. R.; 2003; *Replicated Studies: Building a Body of Knowledge about Software Reading Techniques*. Lecture Notes on Empirical Software Engineering. Chapter 2, pp. 39-84. World Scientific.
- Thelin, T., Andersson, C., Runeson, P., Dzamashvili-Fogelström, N.; 2004; *A Replicated Experiment of Usage-Based and Checklist-Based Reading*; Proceedings of the 10th International Symposium on Software Metrics (METRICS'04)
- Wohlin, C., Petersson, H., & Aurum, A.; 2003; *Combining data from reading experiments in software inspections: a feasibility study*. (pp. 85-132). World Scientific Publishing Co., Inc.
- Woody, J.; Will, R.; Blanton, J.; 1996; *Enhancing Knowledge Elicitation using the Cognitive Interview*; Expert system with application; Vol. 10 N. 1

METODOLOGÍA Y HERRAMIENTAS DE COSTEO EN INGENIERÍA DEL CONOCIMIENTO

F. Pollo-Cattaneo, H. Amatriain, D. Rodríguez, P. Britos, R. García-Martínez

Grupo de Investigación Aplicada en Sistemas Inteligentes. UTN
Centro de Ingeniería de Software Ingeniería del Conocimiento. Escuela de Postgrado. ITBA
Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA
Área Ingeniería del Software. Licenciatura Sistemas. UNLa
Área Ingeniería del Software. Unidad Académica Río Gallegos. UNPA

hamatriain@frlp.utn.edu.ar, {fcattaneo,drodrigu,pbritos}@itba.edu.ar, rgarciamar@fi.uba.ar

CONTEXTO

El Proyecto articula líneas de trabajo del Grupo de Investigación Aplicada en Sistemas Inteligentes de la Universidad Tecnológica Nacional y del Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires.

RESUMEN

La finalidad de este proyecto es la de generar una metodología, con las técnicas y herramientas correspondientes; para brindar información relativa al costo de desarrollo de un Sistema Experto. En base a la propuesta, los CIO's (Chief Information Officer) podrán abordar proyectos de estas características en forma comparable a lo que hacen, en la actualidad, con el desarrollo de software tradicional. Estimar permite anticiparse y proyectarse de mejor manera. El desarrollo de herramientas de costeo contribuye notoriamente en este proceso. El resultado también podrá ser utilizado por una nueva posición gerencial: los Chief Knowledge Officer (CKO), encargados de gestionar en forma eficiente los activos intangibles de la organización.

Palabras Clave: Ingeniería del Conocimiento, Herramientas de Costeo, Metodologías.

1. INTRODUCCION

Los Sistemas Inteligentes tienen como propósito reproducir las acciones y el razonamiento de los seres vivos inteligentes en dispositivos artificiales, cuyo objetivo es conseguir una teoría comprensiva de la inteligencia tal y como aparece en animales y máquinas. Tal vez, el ejemplo más paradigmático de máquinas inteligentes, por el éxito que están alcanzando, sean los sistemas basados en conocimientos (SSBBCC en plural y SBC, en singular), en general, y los sistemas expertos (SSEE en plural y SE, en singular) como puede verse en [Britos, 2001; Rossi, 2001; Rizzi, 2001; Bermejo, 2002; Ierache, 2001; Cao, 2003; Diez, 2003; Gómez, 2003; Hossian, 2003] A la actividad de construir estos sistemas se le denomina Ingeniería del Conocimiento [Hayes-Roth et al, 1983], en adelante INCO, cuya misión es adquirir, formalizar, representar y usar grandes cantidades de

conocimientos de la más alta calidad y específicos de una tarea [García Martínez, 1990a; 1990c; García Martínez y Marsiglio, 1991b; García Martínez et al, 1990; 1991; 1996; Cucatto et al; 1990; 1991; Gómez et al, 2001a; 2001b; Bermejo et al, 2002; Diez, 2003]. Los Sistemas Inteligentes buscan identificar procedimientos, métodos y técnicas que se asocian a la capacidad de pensar y razonar hasta de manera inteligente. En este proceso pretende automatizar las tareas del pensamiento y razonamiento humano, proveyendo un modelo cognitivo de su funcionamiento [García Martínez 1999a, 1999b].

Las metodologías de construcción de SE, desde su viabilidad hasta el testeado, se encuentran maduras desde el punto de vista metodológico. Sin embargo, se ha señalado la necesidad de herramientas complementarias para identificar costos asociados y generar un ambiente metodológico que permita gestionar completamente a estos proyectos.

Por otro lado, el área de sistemas inteligentes en general y, en particular, la Ingeniería en Conocimientos han sido declaradas áreas de vacancia por la Secretaría de Ciencia y Tecnología del Ministerio de Cultura y Educación de la República Argentina [Secretaría de Ciencia y Tecnología, 1999] dentro de las áreas de desarrollo insuficiente.

Los servicios que brindará el sistema MERCOINCO (Metodología y herramientas de Costeo en Ingeniería en Conocimiento) tienen como objetivo primordial completar la falta de herramientas de estimación que existen actualmente dentro de la Ingeniería en Conocimiento para así permitir a empresas de todo tipo o envergadura encarar proyectos de construcción de agentes inteligentes dentro del marco del negocio. Los proyectos de construcción de sistemas expertos no será para unos pocos sino que las empresas podrán utilizarlos como un elemento más que brinda soporte y mejora al negocio.

Dentro del ámbito universitario se ha conformado el Grupo GIASI (Grupo de Investigación Aplicada en Sistemas Inteligentes). Un equipo de docentes y alumnos con actividad en el área específica.

En línea con el proyecto MERCOINCO - el GIASI ha desarrollado los siguientes trabajos: posee como base empírica la implementación de sistemas expertos (SE) en el área Legal [Gómez S, et al, 2001], control del espacio aéreo [Ierache et al, 2004] y control de edificados inteligentes [Sierra E, et al, 2005].

En sus bases conceptuales ha trabajado en verificación de SE a partir de transformaciones algebraicas de grafos [Dramis, L, et al.,2000] y su correspondiente validación [Rizzi, F et, al, 2000]. Sin descuidar las cuestiones pedagógicas de su enseñanza [Britos, P, et al, 1999]. El grupo ha abordado el tema de métricas desde diversos enfoques. En una primera etapa se trabajó con el tema de Bases de Conocimientos [García Martínez, R.,1996] y [García Martínez, R,1997] para continuar recientemente con métricas de madurez [Pollo Cattaneo M, et.al 2007] y [Pollo Cattaneo M, et al, 2008].

Por lo antedicho, para GIASI, la finalidad de este proyecto es desarrollar un ambiente metodológico para el costeo de proyectos de Ingeniería en Sistemas Expertos (generando además las herramientas necesarias que den apoyo a dicho entorno) brindando soporte a la gestión del proyecto y generando valor agregado al proceso de desarrollo de sistemas inteligentes.

2. LINEAS DE INVESTIGACION y DESARROLLO

El sistema MERCOINCO - Metodología y herramientas de Costeo en Ingeniería en Conocimiento - integra dentro del proceso troncal, entidades o ramas de investigación que colaboran e integran en objetivo principal del proyecto.

Las entidades que lo conforman son:

- MetSe - Métricas de Sistemas Expertos: mediante el estudio de métricas existentes dentro de la Ingeniería en Software, MetSe, estudia la aplicación y adaptación de éstas en el campo de los sistemas inteligentes.

- HerCo - Herramientas de Costeo: Las herramientas de Costeo permiten estimar el uso de recursos en un proceso de construcción. La Ingeniería en Conocimiento no se encuentra ajena a esta realidad, para lo cual, embebiéndose de la propuesta de MetSe deberá, de igual forma, estudiar la adaptación y aplicación de herramientas de costeo de la Ingeniería en Software dentro del campo de los sistemas inteligentes.

- HerComEdu - Herramienta en pruebas dentro de la Comunidad Educativa: esta entidad implementa todo lo propuesto por los agentes que participan del proyecto en elementos tangibles producidos dentro del ámbito educativo. Los alumnos con sus desarrollos permitirán "ver" resultados concretos. Esto conlleva a la reflexión y enriquece el ambiente de aprendizaje.

3. RESULTADOS OBTENIDOS/ESPERADOS

La finalidad central de este proyecto es la de generar una metodología con sus herramientas para brindar información relativa al costo de desarrollo de un Sistema Experto. De esta manera, los CIO's (Chief Information Officer) podrán abordar proyectos de estas características en forma comparable a lo que hacen en la actualidad con el desarrollo de software tradicional.

En lo que respecta al avance del conocimiento científico y/o tecnológico, generar un ambiente metodológico que facilite el costeo dentro del proceso de construcción de Sistemas Expertos promueve el desarrollo y la validación de métodos, técnicas y herramientas. Esto conlleva a una mejora sustancial en la calidad de los proyectos de construcción. La aplicación de una metodología permite además dotar al proceso de construcción de: objetividad, sistematicidad, racionalidad, generalidad y fiabilidad, contribuyendo al avance del conocimiento científico mediante el uso de técnicas consistentes.

Tal como se ha descrito anteriormente, el costeo es uno de los aspectos a tener en cuenta al encarar un proyecto de construcción desde una perspectiva ingenieril. El uso eficiente de los recursos constituye un factor clave de éxito para todos aquellos proyectos en el nuevo siglo. Asimismo, estos ambientes conformarán el entramado socio-técnico en el cual deberá desenvolverse el futuro profesional. Este proyecto contribuye sustancialmente a incorporar e implicar a los alumnos de esta carrera en el uso de metodologías de costeo para la gestión de proyectos de Ingeniería en Sistemas Expertos en particular, y de Proyectos Software, en general. El uso cotidiano de las mismas seguramente los conformará en profesionales de mayor calidad, a la vez que les brindará las herramientas necesarias para adquirir una mayor flexibilidad y adaptarse a entornos de formación continua y cambiante.

De esta forma, se espera que el proyecto logre la motivación de los implicados, movilizándolos y brindándoles así la posibilidad de constituirse ellos mismos en agentes de cambio. Consecuentemente, en gran medida se convertirán en generadores de herramientas que mejoren y expandan las actuales técnicas de costeo en los proyectos vigentes. Asimismo, aquellos profesionales y alumnos avanzados que diseñen y desarrollen proyectos con estas características serán pioneros en el área de la gestión de proyectos de sistemas inteligentes, proveyendo innovación y retroalimentación continua al proceso.

Por lo expuesto anteriormente, el proyecto Mercoinco, plantea la integración a esta línea de Investigación Aplicada, de alumnos avanzados en la carrera de Ingeniería en Sistemas de Información, con posibilidades de articular Tesis de Trabajos Finales de Carrera de Grado y Tesis de Magister.

Los recursos humanos deben poseer una firme vocación de trabajar en el área de los sistemas inteligentes, de manera de canalizar y proveer una base sustentable de aporte al proyecto. De esta manera, se logra un doble beneficio, el proyecto obtiene e incorpora el conocimiento tecnológico de los recursos humanos en el área de la especialidad, a la vez que plantea un esquema de formación de especialistas de punta en el procesos de gestión.

4. FORMACION DE RECURSOS HUMANOS

En lo que respecta al Proyecto, el Grupo GIASI pretende incorporar diez trabajos Finales de Carrera y dos Tesis de Maestría. Por otro lado, se busca lograr que los participantes sean promovidos dentro del escalafón de la carrera de investigadores.

5. BIBLIOGRAFIA

- Britos, P., Rossi, B. y García Martínez, R. *Notas sobre Didáctica de las Etapas de Formalización y Análisis de Resultados de la Técnica de Emparrillado. Un Ejemplo*. Proceedings del V Congreso Internacional de Ingeniería Informática. Páginas 200-209. Editado por Departamento de Publicaciones de la Facultad de Ingeniería. Agosto 1999.
- Dramis, L., Britos, P., Rossi, B. y García Martínez, R. *Transformación Algebraica de Grafos en Verificación de Bases de Conocimiento*. Proceedings del VI Congreso Internacional de Ingeniería Informática. Páginas 729-738. Editado por Departamento de Publicaciones de la Facultad de Ingeniería. ISBN 987-98197-0-5. 2000.
- Ferreira, J., Takai, O., Pu, C. (2005). *Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services*. Proceedings Seventh IEEE International Conference on E-Commerce Technology. Pág. 471-474.
- García Martínez, R. *Una Medida para Evaluar la Calidad de una Base de Conocimiento*. Resúmenes de las III Jornadas de Informática e Investigación Operativa. Páginas 48-49. Facultad de Ingeniería. Universidad de la República. Montevideo. Uruguay. Septiembre 1996.
- García Martínez, R. *Calidad de Bases de Conocimiento. Una Medida de Evaluación*. Proceedings del III Congreso Internacional de Informática y Telecomunicaciones.. Páginas 191-198. Buenos Aires. 1997.
- Gómez, S., Perichinsky, G. y García Martínez, R. *Un Sistema Experto Legal para la Individualización y Acuerdos para Penas*. Proceedings del Simposio Argentino de Informática y Derecho. Pág. 23-33. Septiembre. Editado por la Sociedad Argentina de Informática e Investigación Operativa. 2001.
- Ierache, J. y García-Martínez, R. *Sistema Experto Aplicado al Control del Espacio Aéreo*. Proceedings del IX Congreso Argentino de Ciencias de la Computación. 2004.
- Lönnqvist, A., Pirttimäki, V. (2006). *The Measurement of Business Intelligence*. Information Systems Management, 23(1): 32-40.
- Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2007. *Aplicación de métricas de madurez en conceptualización de sistemas expertos*. Proceedings IX Workshop de Investigadores en Ciencias de la Computación. Pag. 327-331. ISBN 978-950-763-075-0.
- Pollo-Cattaneo, F., Fernández, E., Merlino, H., Rodríguez, D., Britos, P., García-Martínez, R. 2008. *Métricas de Madurez en Conceptualización de Sistemas Expertos. Casos de Estudio*. Proceedings VII Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento. Pág. 107-115. ISSN 1390-292X.
- Rizzi, F., Britos, P., Dramis, L., Rossi, B. y García Martínez, R. *Validación de Sistemas Basados en Conocimiento*. Proceedings del VI Congreso Internacional de Ingeniería Informática. Páginas 739-749. Editado por Departamento de Publicaciones de la Facultad de Ingeniería. Abril 2000. ISBN 987-98197-0-5. 2000.
- Secretaría de Ciencia y Tecnología. Ministerio de Cultura y Educación. *La investigación Científica y Tecnológica en Argentina. Diagnóstico e Identificación de Áreas de Vacancia*. Editado por Estudio Estudio Sigma SRL. Octubre 1999.
- Sierra, E., Hossian, A., García-Martínez, R. y Marino, P. *Sistema Experto para Control Inteligente de las Variables Ambientales de un Edificio Energéticamente Eficiente*. Proceedings de la XI Reunión de Trabajo en Procesamiento de la Información y Control. Universidad Nacional de Río Cuarto. Pág. 446-452. 2005.

INTEGRACIÓN DE SISTEMAS BASADOS EN CONOCIMIENTO Y DE DESCUBRIMIENTO. APLICACIÓN A LA TOMA DE DECISIONES ESTRATÉGICAS OPERACIONALES

Claudio Rancan^{1,2}, Patricia Pesado^{1,3}, Ramón García-Martínez^{1,4}

¹ Programa de Doctorado en Ciencias Informáticas. Facultad de Informática. UNLP

² Centro de Ingeniería de Software Ingeniería del Conocimiento. Escuela de Postgrado. ITBA

³ Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP - CIC

⁴ Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA

crancan@itba.edu.ar, ppesado@lidi.info.unlp.edu.ar, rgarciamar@fi.uba.ar

CONTEXTO

El Proyecto articula líneas de trabajo de los siguientes Proyectos: [a] “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaria de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires.

RESUMEN

Este proyecto se centra en definir una arquitectura de integración entre sistemas de descubrimiento de conocimiento (adquisición automática) y sistemas basados en conocimiento (sistemas expertos). Se busca estimar la viabilidad de implementación de dicha arquitectura en función de la madurez de las tecnologías involucradas.

Palabras clave: Descubrimiento de conocimiento. Sistemas basados en conocimiento. Arquitecturas de Sistemas

1. INTRODUCCION

Los sistemas basados en conocimientos (SBC) ó sistemas expertos emulan el comportamiento humano experto en un área de conocimiento determinada. Constituyen sistemas de ayuda a la toma de decisiones en áreas tan diversas como la selección de estrategias instruccionales [Sierra *et al.*, 2004], el control de variables ambientales [Sierra *et al.*, 2005], la configuración de ventiladores en neonatología [Bermejo *et al.*, 2002] la individualización y acuerdos para penas en procesos judiciales [Gómez *et al.*, 2001] o la generación asistida del mapa de actividades de proyectos de desarrollo de software [Diez *et al.*, 2003].

La base de conocimiento de un sistema experto encapsula en algún formalismo de representación (reglas, marcos, redes semánticas entre otras), el conocimiento del dominio que debe ser puesto en juego por el sistema para resolver un problema dado.

Las metodologías de construcción de bases de conocimiento se han consolidado en los últimos 15 años [Debenham, 1990; 1998; Gomez *et al.*, 1997; García-Martínez y Britos, 2004].

Los sistemas inteligentes constituyen el campo de la Informática en el que se estudian y desarrollan algoritmos que implementan los distintos modelos de aprendizaje y su aplicación a la resolución de problemas prácticos [Michalski, 1991; Dejong & Money 1986]. Entre los problemas abordados en este campo, está el de descubrir conocimientos [Michalski *et al.*, 1983; 1986; 1998; Michalski y Tecuci, 1994; Mitchel, 1996; Michie, 1988].

El descubrimiento de conocimiento (KD Knowledge Discovery) es la búsqueda de patrones interesantes y de regularidades importantes en grandes bases de información [Fayad *et al.*, 1996; Grossman *et al.*, 1999]. Al hablar de descubrimiento de conocimiento basado en sistemas inteligentes o Minería de Datos/Información Inteligente [Evangelos *et al.*, 1996; Michalski *et al.*, 1998] nos referimos específicamente a la aplicación de métodos de aprendizaje automático u otros métodos similares, para descubrir y enumerar patrones presentes en dicha información.

En este contexto, trabajos recientes sobre sistemas de ayuda a la toma de decisiones estratégicas operacionales (SATD-EO) basados en SBC [Sierra *et al.*, 2006] en áreas como el control aéreo [Ierache y García-Martínez, 2004] ó el alistamiento de unidades navales [Rancán, 2004] han puesto de manifiesto que es un problema abierto la definición de cómo se puede integrar a este tipo de sistemas, procesos de descubrimiento de conocimiento basados en aprendizaje automático [García-Martínez y Borrajo, 1997; 2000; Grosser *et al.*, 2005; Felgaer *et al.*, 2006, Cogliati *et al.*, 2006] que les permitan mejorar “en línea” la calidad de la base de conocimiento utilizada para la ayuda a la toma de decisiones.

2. LINEAS DE INVESTIGACION y DESARROLLO

Las líneas de investigación de este proyecto son:

- [a] El paradigma de arquitecturas integradas de planificación y aprendizaje basado en construcción de teorías y su potencial aplicación al descubrimiento de conocimiento estratégico operacional [Fritz, *et al.*, 1989; García-Martínez y Borrajo, 1996;1997; 1998; 2000; Sierra *et al.*, 2006].
- [b] La arquitectura subyacente a sistemas basados en conocimiento de ayuda a la toma de decisiones conocidos [García Martínez y Britos, 2004; Britos, 2001; Rizzi, 2001; Ierache y García-Martínez, 2004; Bermejo *et al.*, 2002; Gomez *et al.*, 2001; Hostian 2003].
- [c] Los paradigmas de descubrimiento de conocimiento con énfasis en la evaluación de conocimiento [Jensen, 2002], sus estructuras [Utgof *et al.*, 2000; Jensen y Neville, 2002; Neville y Jensen, 2002], los procesos de adquisición distribuidos [Jensen *et al.*, 1999] y las tecnologías de sistemas inteligentes asociadas al descubrimiento de conocimiento [Britos *et al.*, 2005].

3. RESULTADOS OBTENIDOS/ESPERADOS

A la fecha se ha definido la arquitectura de integración de sistemas de descubrimiento de conocimiento y sistemas basados en conocimiento [Rancan *et al.*, 2007a] y se han realizado experimentos exploratorios sobre la calidad de los resultados obtenibles [Rancan *et al.*, 2007b]. El año pasado (2008) se han identificado casos reales de aplicación año se prevé la prueba de conceptos sobre casos reales.

4. FORMACION DE RECURSOS HUMANOS

En el marco de este proyecto se esta desarrollando una tesis doctoral y tres tesis de ingeniería informática (dos de ellas defendidas y otra en curso). Esta línea vincula al LIDI de la Facultad de Informática de la Universidad Nacional de La Plata y al Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires.

5. BIBLIOGRAFIA

Bermejo, F., Britos, P., Rossi, B y García Martínez, R. 2002. *Sistema de Asistencia para la Configuración de Ventiladores OAF en Neonatología*. Revista del Instituto Tecnológico de Buenos Aires. 28: 24-68. ISSN: 0326-1840. 2002.

Britos, P. 2001. *Sistema de Ayuda sobre Legislación Argentina en Riesgos de Trabajo*. Tesis de Magister en Ingeniería del Conocimiento. Facultad de Informática de la Universidad Politécnica de Madrid.

Britos, P., Hossian, A., García Martínez, R. y Sierra, E. 2005. *Minería de Datos Basada en Sistemas Inteligentes*. 876 páginas. Editorial Nueva Librería. ISBN 987-1104-30-8.

Cogliati, M., Britos, P. y García-Martínez, R. (2006). *Patterns in Temporal Series of Meteorological Variables Using SOM & TDIDT*. Springer IFIP Series. Aceptado.

De Giusti E., Mollo Brisco G., La Battaglia, J., Pasini, A. Dirección: Pesado, P. 2004. *Sistema de Simulación de Escenarios y Decisiones Empresarias*. Proceedings de las XII Jornadas de Jóvenes Investigadores AUGM 2004. Universidad Federal Do Paraná - Curitiba, Brasil.

Debenham, J. 1990. *Knowledge Systems Design*. Prentice Hall.

Debenham, J. 1998. *Knowledge Engineering: Unifying Knowledge Base and Database Design*. Springer-Verlag.

DeJong, G.F., Mooney, R.J. 1986. *Explanation-Based Learning: An Alternative View*, Machine Learning, 1: 145-176.

Diez, E., Britos, P., Rossi, By García-Martínez, R. 2003. *Generación Asistida del Mapa de Actividades de Proyectos de Desarrollo de Software*. Reportes Técnicos en Ingeniería del Software. 5(1):13-18. ISSN 1667-5002.

Evangelos, S., Han, J, (eds). 1996. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, EE.UU.

Fayad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uhturudsamy, R. (eds). 1996. *Advances in Knowledge Discovery and Data Mining*, San Mateo, AAAI Press.

Felgaer, P., Britos, P. and García-Martínez, R. 2006. *Prediction in Health Domain Using Bayesian Network Optimization Based on Induction Learning Techniques*. International Journal of Modern Physics C 17(3): 447-455.

Fritz, W., García Martínez, R., Rama, A., Blanqué, J., Adobatti, R, y Sarno, M. 1989. *The Autonomous Intelligent System*. Robotics and Autonomous Systems, 5(2): 109-125.

García Martínez, R. & Borrajo Millán, D. 1996. *Unsupervised Machine Learning Embedded in Autonomous Intelligent Systems*. Proceedings of the XIV International Conference on Applied Informatics. Páginas 71-73. Innsbruck. Austria.

García Martínez, R. & Borrajo Millán, D. 1998. *Learning in Unknown Environments by Knowledge Sharing*. Proceedings of the Seventh European Workshop on Learning Robots. Páginas 22-32. Editado University of Edinburg Press.

García Martínez, R. y Borrajo Millán, D. 1997. *Planning, Learning and Executing in Autonomous Systems*. Lecture Notes in Artificial Intelligence. 1348:208-210.

García Martínez, R. y Borrajo Millán, D. 2000. *An Integrated Approach of Learning, Planning and Executing*. Journal of Intelligent and Robotic Systems 29(1):47-78.

García-Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería.

- Gomez, A., Juristo, N., Montes, C. y Pazos, J. 1997. *Ingeniería del Conocimiento*. Editorial R. Areces. Madrid.
- Gómez, S., Perichinsky, G. y García Martínez, R. 2001. *Un Sistema Experto Legal para la Individualización y Acuerdos para Penas*. Proceedings del Simposio Argentino de Informática y Derecho. Pág. 23-33. Septiembre. Editado por la SADIO.
- Grosser, H., Britos, P. y García-Martínez, R. 2005. *Detecting Fraud in Mobile Telephony Using Neural Networks*. Lecture Notes in Artificial Intelligence 3533: 613-615.
- Grossman, R., Kasif, S., Moore, R., Roche, D. and Ullman, J. 1999. *Data Mining Research: Opportunities and Challenges*, A Report of three NSF Workshops on Mining Large, Massive, and Distributed Data, January 1999, Chicago
- Hossian, A. 2003. *Sistema de Asistencia para la Selección de Estrategias y Actividades Instruccionales*. Tesis de Magister en Ingeniería del Software. ITBA.
- Ierache, J. y García-Martínez, R. 2004. *Sistema Experto Aplicado al Control del Espacio Aéreo*. Proceedings del IX Congreso Argentino de Ciencias de la Computación.
- Jensen D. 2002. *Knowledge Evaluation*. Handbook of Data Mining and Knowledge Discovery. Kloesgen, W. and J. Zytkow (Eds.). Oxford: Oxford University Press
- Jensen D. and J. Neville 2002. *Schemas and models*. Proceedings of the Multi-Relational Data Mining Workshop, 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Jensen D., Y. Dong, B. Lerner, E. McCall, L. Osterweil, S. Sutton Jr., and A. Wise 1999. *Coordinating agent activities in knowledge discovery processes*. In Proceedings of the International Joint Conference on Work Activities Coordination and Collaboration. pp. 137-146.
- La Battaglia, J., Rodriguez, I., Thomas, P., Bertone, R. Dirección: Pesao P. 2003. *Tecnología aplicada a gestión distribuida*. Proceedings de las XI Jornadas de Jóvenes Investigadores AUGM 2003. Universidad Nacional de La Plata – Argentina.
- Miatón, I., Pesado, P., Bertone, R. y De Giusti. 2003. *Agentes Basados en Sistemas Distribuidos*. Proceedings V WICC.
- Michalski, R. 1993. *A Theory and Methodology of Inductive Learning*, Artificial Intelligence, 20:111-161.
- Michalski, R. Bratko, I. Kubat, M (eds.) 1998. *Machine Learning and Data Mining, Methods and Applications*, John Wiley & Sons Ltd, West Sussex, England
- Michalski, R. S. 1991. *Toward an Unified Theory of Learning: An Outline of Basic Ideas*, Proceedings of the 3rd World Conference on the Fundamentals of Artificial Intelligence, Paris.
- Michalski, R. S., Carbonell, J. G., Mitchell, T. M. (eds.), 1986. *Machine Learning: An Artificial Intelligence Approach, Vol. II*, Morgan-Kauffman
- Michalski, R. S., Carbonell, J. G., Mitchell, T. M. (eds.). 1983. *Machine Learning: An Artificial Intelligence Approach, Vol. I*. Morgan-Kauffman
- Michalski, R. S., Tecuci, G. (eds) 1994. *Machine Learning: A Multistrategy Approach, Vol. III*, Morgan Kauffman
- Michie, D. 1988. *Machine Learning in the next five years*, EWSL-88, 3rd European Working Session on Learning, Glasgow, Londres, Pitman.
- Mitchell, T. M. 1996. *Machine Learning*, McGraw-Hill.
- Neville J. and D. Jensen 2002. *Supporting relational knowledge discovery: Lessons in architecture and algorithm design*. Proceedings of the Data Mining Lessons Learned Workshop, 19th International Conference on Machine Learning.
- Pesado, P. Feierherd G. y Pasini, A. 2005. *Requirement Specifications for Electronic Voting System*. Journal of Computer Science & Technology, 5(4): 312-319 . ISSN: 1666-6038
- Rancán, C. 2004. *Arquitectura de Sistema Híbrido de Evaluación del Alistamiento de Unidades Navales Auxiliares*. Reportes Técnicos en Ingeniería del Software. 6(1): 45-54. ISSN 1667-5002.
- Rancan, C., Kogan, A., Pesado, P. and García-Martínez, R. (2007b). *Knowledge Discovery for Knowledge Based Systems. Some Experimental Results*. Research in Computing Science Journal, 27: 3-13.
- Rancán, C., Pesado, P. y García-Martínez, R. (2007a). *Toward Integration of Knowledge Based Systems and Knowledge Discovery Systems*. Journal of Computer Science & Technology, 7(1): 91-97.
- Rizzi, M. 2001. *Sistema Experto Asistente de Requerimientos*. Tesis de Magister en Ingeniería del Software. ITBA.
- Sierra, E., García-Martínez, R., Hossian, A., Britos, P. y Balbuena, E. 2006. *Providing Intelligent User-Adapted Control Strategies in Building Environments*. Research in Computing Science Journal, 19: 235-241.
- Sierra, E., Hossian, A. y García-Martínez, R. 2003. *Sistemas Expertos que Recomiendan Estrategias de Instrucción. Un Modelo para su Desarrollo*. Revista Latinoamericana de Tecnología Educativa. 1(1): 19-30. ISSN: 1695-288X. 2003.
- Sierra, E., Hossian, A., García-Martínez, R. y Marino, P. 2005. *Sistema Experto para Control Inteligente de las Variables Ambientales de un Edificio Energéticamente Eficiente*. Proc. XI RTPIC. Pág. 446-452.
- Utgoff P., V. Lesser, and D. Jensen 2000. *Inferring task structure from data*. University of Massachusetts, Department of Computer Science. Technical Report UM-CS-2000-054.

CLASIFICACION AUTOMATICA PARA LA PREVENCION DEL ESTRÉS DE LOS SUELOS Y LA FATIGA DE SOJA EN EL NOROESTE ARGENTINO

Sanson, E., Britos. P., Rodriguez, D., García-Martínez, R.

Centro de Ingeniería de Software Ingeniería del Conocimiento. Escuela de Postgrado. ITBA

Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA

Área Ingeniería del Software. Licenciatura Sistemas. UNLa

Área Ingeniería del Software. Unidad Académica Río Gallegos. UNPA

{pbritos,drodrigu}@itba.edu.ar, rgarciamar@fi.uba.ar

CONTEXTO

El Proyecto articula líneas de trabajo del Proyectos “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaria de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires

RESUMEN

El crecimiento de la soja y sus derivados en la Argentina en los últimos 30 años se ha constituido en un fenómeno trascendente y fundamental para el país, tanto desde el punto de vista económico como social. Para la campaña 2002/03, la soja representó más del 50% del total de los granos producidos, considerando los cinco cultivos mas difundidos en la pampa húmeda. Lo anterior implica que el sector agropecuario, especialmente en la pampa húmeda, ha tornado hacia un modelo de "especialización sojera", que se expande rápidamente hacia regiones extrapampeanas, al que muchos definen como el proceso de "sojificación", y el nuevo paradigma de la agricultura argentina basado en los materiales transgénicos y la siembra directa. [Begenisic, F; *et al*; 2004; Formento, N; de Souza, J.; 2007; Graciela, E; 2006; Sinavimo, 2008; Rearte, D; 2007].

Ante este presente las tecnologías informáticas han dado soporte desde lo las funciones esenciales del negocio y sin llegar a profundizar en la forma analítica-predictiva masivamente.

El presente trabajo intenta acortar dicha brecha y de esa forma por ayudar a la industria agrícola-ganadera del país, por el medio de ayudar en la prevención de la Roya de Soja y la Fatiga de los suelos. Para lograr este objetivo se utilizaran las mediciones meteorológicas en la zona noreste de la provincia de Córdoba.

Palabras clave: Minería de datos, agronomía, sistemas inteligentes.

1. INTRODUCCION

La soja ha significado en el mundo una gran revolución productiva-comercial en la cadena agroalimentaria mundial a partir de los años setenta por ser a la vez principal fuente de proteína vegetal para la alimentación animal en sistemas de cría intensiva así como proveedora de aceites para alimentación humana. Adicionalmente, en estos últimos años es también uno los cultivos claves para la obtención de biocombustibles.

Argentina es el tercer exportador de grano (dependiendo del semestre del año), luego de Estados Unidos y Brasil pero es el primer exportador mundial de productos procesados: Aceites –tanto crudo como envasados - y harinas proteicas en un nivel de oferta muy superior al de Brasil que es el segundo exportador mundial.

Si bien Argentina exporta productos procesados oleaginosos a más de 100 países en los cinco continentes, los mercados principales de aceite de soja son China e India –países de altísima demanda debido al intenso crecimiento de sus ingresos a partir de este milenio.

A continuación se describen dos situaciones que afectan a la industria agrícola-ganadera del país, la primera es la Roya de Soja y la Fatiga de los suelos.

La roya de soja es un hongo que afecta a varios procesos de crecimiento de la planta tales como la capacidad fotosintética y del crecimiento del tallo, aborto de las vainas e interrupción del llenado de granos. Esta considerada como la enfermedad más importante para este tipo de sembrado que reducen los rendimientos y que pueden incluso la pérdida total de la producción de un lote.

Por otra parte, el término fatiga del suelo o estrés de los suelos engloba un amplio concepto que podría definirse como la pérdida de vigor y rendimiento productivo de las plantas cuando se efectúa un cultivo reiterado sobre un mismo suelo. Evidentemente, existe un gran número de factores que pueden, de forma más o menos conjunta, desencadenarla [[Begenisic, F; *et al*; 2004; Formento, N; de Souza, J.; 2007; Graciela, E; 2006; Sinavimo, 2008; Rearte, D; 2007].

Para este fin se tomo, en particular, el estudio de las medidas meteorológicas y como estas inciden en los suelos, para de esa forma obtener indicadores sobre

alguno de los problemas agrícolas en la zona. Puntualmente se buscara generar patrones a partir de los datos meteorológicos de esa región del país, para inferir en enfermedades referentes a los cultivos de soja, como es la roya de soja y la fatiga de los suelos. Al establecer el uso de patrones en los datos para inferir en dichas problemáticas, es donde se hará uso de las herramientas de Explotación de la Información.

2. LINEAS DE INVESTIGACION y DESARROLLO

Se denomina Explotación de la Información (data mining) al análisis de archivos y bitácoras de transacciones, trabaja a nivel del conocimiento con el fin de descubrir patrones, relaciones, reglas, asociaciones o incluso excepciones útiles para la toma de decisiones.

Las técnicas asociadas a la Explotación de la Información son diversas y se pueden nombrar las pertenecientes al aprendizaje automático, el análisis estadístico de los datos, la visualización de datos y las redes neuronales artificiales (RNA). La Explotación de la Información se refiere a la aplicación de técnica de aprendizaje automático, entre otros métodos, para encontrar importantes patrones en los datos. [García Martínez, R., *et al*, 2003].

Por lo tanto podemos establecer que un proceso de descubrimiento de conocimientos consiste de una repetición iterativa de los siguientes pasos:

- Limpieza de datos (Data Cleaning) procesamiento de los datos ruidosos, erróneos, faltantes o irrelevantes
- Integración de datos (Data integration) integración de múltiples y heterogéneas fuentes de datos en una única fuente.
- Selección de datos (Data selection) extracción de los datos relevantes al área de análisis del almacenamiento de datos.
- Transformación de datos (Data transformation) transformación o consolidación de los datos en formas apropiadas para la minería mediante procedimientos de agregación.
- Explotación de la Información: proceso esencial donde se aplican diversos métodos para extraer patrones de los datos.
- Evaluación de patrones (Pattern evaluation) identificación de patrones interesantes basándose en algún parámetro de comparación impuesto por el usuario.
- Presentación de los conocimientos (Knowledge presentation) técnicas de visualización y representación de los conocimiento

A si mismo, para este proyecto se empleo la metodología CRISP-DM [Chapman, P; *et al*, 1999], para el análisis y diseño ya que este tipo de herramienta son especificas para los distintos problemas de la Explotación de la Información con

una estructura o fases, bien definidas. En el presente trabajo se emplearon dichas fases.

La metodología CRISP-DM [Chapman, 1999], la cual es un estándar para la minería de datos y el análisis predictivo, presenta un ciclo de vida en orden jerárquico, que consiste en un conjunto de tareas que se describen en cuatro niveles de abstracción:

- Fases: por cada una de estas fases, existen un número determinado de tareas. Forman parte del Modelo Genérico de CRISP-DM.
- Tareas Genéricas: estas tareas son llamadas genéricas por que intentan cubrir todas las posibles situaciones de la minería de datos, como por ejemplo extracción de los datos.
- Tareas Específicas: estas tareas describen como algunas de las tareas genéricas van a ser efectuadas en específicas situaciones, por ejemplo extracción de los datos específicamente de humedad de los suelos en la provincia de Córdoba.
- Instancia de Procesos: es decir la instancia en que las tareas son efectuadas, las decisiones que son necesarias para dicha tarea y los resultados para esa tarea específica de la minería de datos.

Para la metodología CRISP DM se puede distinguir entre la referencia del modelo y la guía del usuario. La primera presenta una rápida vista general de las fases del modelo, tareas y las salidas esperadas en el modelo de minería de datos.

En el presente trabajo y ante la problemática antes descrita, en una primera instancia se puede observar como las mediciones meteorológicas que son generadas por la consola de medición climática Groweather, estas se exportan en un archivo de texto (archivo plano) hacia el servidor. Por medio de un básico proceso de extracción, transformación y carga de los datos, estos son llevados a la base de datos para de esa forma ser depurados y relacionados con otros datos.

En este punto, vale destacar que entre los datos con los que se cuentan provenientes de la consola de Groweather son fecha y hora de la muestra, temperatura máxima, temperatura media, humedad relativa del ambiente, radiación solar, velocidad del viento, sensación térmica entre otros.

Con los datos ya procesados y relacionados entre si, se puede emplear cualquiera de las herramientas de Explotación de la Información para de esta forma poder inferir en la problemática vigente. Dicho de otra forma, se podría generar por ejemplo, un árbol de decisiones a partir del modelo ID3 por el cual obtendríamos la precisión para saber si ante determinado rango de temperatura media, humedad relativa del ambiente y en determinada temporada del año hay un alto riesgo de que los suelos de esa zona geográfica puedan tener un índice de fatiga alto o no.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se utiliza una muestra de 731 casos, a los cuales se le aplicó en primera instancia dos métodos inductivos C4.5 y MC4 para determinar el stress del suelo. Pudiéndose obtener los siguientes resultados: Reglas que se desprenden del árbol aplicando C4.5:

Estado del suelo: Estrés

1. Si la Temperatura Media es mayor o igual a 26,05 entonces los suelos están en el estado de Estrés
2. Si la Temperatura Media es mayor o igual a 25,55 y menor a 26,05 y la Humedad Relativa es menor o igual a 69,9 entonces los suelos están en el estado de Estrés.

Estado del suelo: Normal

3. Si la Temperatura Media es menor a 22,25 entonces los suelos están en el estado Normal.
4. Si la Temperatura Media es mayor o igual a 22,25 y menor a 22,85 y la Humedad Relativa es menor a 66,45 entonces los suelos están en el estado Normal.

Estado del suelo: Fatiga

5. Si la Temperatura Media es mayor o igual a 22,85 y menor a 25,55 entonces los suelos están en el estado de Fatiga.
6. Si la Temperatura Media es mayor o igual a 22,25 y menor a 22,85 y la Humedad Relativa es mayor o igual a 66,45 entonces los suelos están en el estado de Fatiga.
7. Si la Temperatura Media es mayor o igual a 25,55 y menor a 26,05 y la Humedad Relativa es mayor a 69,9 entonces los suelos están en el estado de Fatiga.

Los resultados del modelo son representativos de la realidad de la muestra. Se puede ver como en el estudio de los datos, para los del tipo Normal, el modelo solo acertó en el 99,57% de los casos. En el caso de los del tipo Fatiga, acertó en el 98,9% de los casos y en el tipo Estrés acertó el 92,7% de los casos.

A continuación se aplica el método MC4 (este método es similar al C4.5 pero se distingue por ser más sensitivo a los costos que el anterior. Emplea una generalización de la estimación de Laplace, con lo que minimiza las pérdidas en la clasificación).

Estado del suelo: Estrés

1. Si la Temperatura Media es mayor o igual a 26,05 entonces los suelos están en el estado de Estrés
2. Si la Temperatura Media es mayor o igual a 25,55 y menor a 26,05 y la Humedad Relativa es menor o igual a 69,9 entonces los suelos están en el estado de Estrés.

Estado del suelo: Normal

3. Si la Temperatura Media es menor a 22,25 entonces los suelos están en el estado Normal.
4. Si la Temperatura Media es mayor o igual a 22,25 y menor a 22,85 y la Humedad Relativa es menor a 70,9 entonces los suelos están en el estado Normal.

Estado del suelo: Fatiga

5. Si la Temperatura Media es mayor o igual a 22,85 y menor a 25,55 entonces los suelos están en el estado de Fatiga.
6. Si la Temperatura Media es mayor o igual a 22,25 y menor a 22,85 y la Humedad Relativa es mayor o igual a 70,9 entonces los suelos están en el estado de Fatiga.
7. Si la Temperatura Media es mayor o igual a 25,55 y menor a 26,05 y la Humedad Relativa es mayor a 69,9 entonces los suelos están en el estado de Fatiga.

Se puede ver como al comparar las reglas generadas por el método C4.5 y el MC4 las mismas son iguales en general. La diferencia se puede ver en el ajuste de valores de las reglas 4 y 6 en donde el método MC4 discrimina por valores de la Humedad Relativa en 70,9 en lugar de 66,45 del C4.5.

Seguidamente a este estudio se utilizarán redes neuronales SOM para agrupamiento y posteriormente inducción para determinar reglas de comportamiento en grupo de acuerdo a los procesos planteados en [Britos, P; 2008]

4. FORMACION DE RECURSOS HUMANOS

En la línea de investigación cuyos resultados parciales se reportan en esta comunicación, se encuentran trabajando: dos tesis de maestría, un tesis de especialización y tres investigadores formados.

5. BIBLIOGRAFIA

- Begenisic, F., Ploper, D., Ivancovich, A. 2004. Plan Nacional de Soja de la Soja. www.sinavimo.gov.ar/?q=node/3123. Vigente a 03/09.
- Britos, P. 2008. Procesos de Explotación de Información Basados en Sistemas Inteligentes. Tesis Doctoral. Facultad de Informática UNLP
- Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 1999. CRISP-DM 1.0 Step by step BGuide. www.crispdm.org/CRISPWP-0800.pdf. Vigente a 03/09
- [Formento, N., de Souza, J. 2007. Folleto Soja De La Soja – INTA EEA Paraná – SAGPyA. 2007. www.sagpya.mec.gov.ar/new/0-0/nuevositio/agricultura/sanidad/index.php?fondo_agri_01=Sanidad&fondo_agri_sanidad=soja. Vigente a 03/09.
- García Martínez, R., Servente, M. y Pasquini, D. 2003. Sistemas Inteligentes. Editorial Nueva Librería. Buenos Aires.
- Graciela, E; 2006. Grupo Economía Agrícola. 2006. Márgenes Brutos de los Principales Cultivos. www.inta.gov.ar/saenzpe/info/documentos/economia/2006/2008-12s-1.pdf. Vigente a 03/09
- Rearte, D. 2007. Situación de la Ganadería Argentina en el contexto Mundial. www.inta.gov.ar/balcarce/carnes/sitgan.htm. Vigente a Marzo 2009.
- Sinavimo. P. 2008. Soja de la soja en Argentina. www.planetasoja.com/trabajos/trabajos800.php?id1=27914&publi=&idSec=107&id2=27915. Vigente a 03/09.

Clasificación de Componentes SIG

Gabriela Gaetán, Viviana Saldaño
Grupo UNPA-UACO
Universidad Nacional de la Patagonia Austral
Unidad Académica Caleta Olivia
Santa Cruz, Argentina

Alejandra Cechich y Agustina Buccella
Grupo de Investigación GIISCo
Departamento de Ciencias de la Computación
Universidad Nacional del Comahue
Neuquén, Argentina

Resumen

Hoy en día existen muchos enfoques que proponen metodologías para el desarrollo de software. Entre ellos, el Desarrollo de Software basado en Componentes (DSBC) propone una manera más rápida y segura de desarrollar aplicaciones. Por otro lado, en cuanto a la información que se modela en las mismas, la información geográfica esta ganando más espacio. Es sabido que cada vez es mayor la cantidad de entidades u organizaciones en el mundo que están utilizando la información geográfica como base para sus sistemas de información. Así, tanto la creación de nuevos componentes geográficos como la necesidad de los mismos para incluirlos en Sistemas de Información Geográficos (SIG) han generado una explosión en el campo del DSBC para SIG. En este trabajo, nos centramos en ambas áreas creando nuevas herramientas que faciliten la búsqueda, publicación y selección de componentes tanto para los proveedores como para los consumidores de los mismos.

1. Motivación

En nuestro trabajo se combinan dos áreas de la Ciencia de la Computación para dar soluciones a la tarea de desarrollo de software. Por un lado, nos interesa el *Desarrollo de Software basado en Componentes (DS-BC)* ya que intenta minimizar tiempos de desarrollo y asegurar productos más confiables utilizando el *reuso* como herramienta fundamental de modelado [13, 14]. Principalmente está basado en la utilización de piezas prefabricadas reusando así tareas de análisis, diseño, e implementación de los servicios que proveen. A su vez, además de la reducción del tiempo de de-

sarrollo y costos, la calidad del producto final mejorará debido a que estas piezas ya han sido probadas y validadas.

Por otro lado, el área de los *Sistemas de Información Geográficos (SIG)* incluye a todos aquellos productos que poseen tanto información geográfica como herramientas para manipularla. Una definición completa de los SIG ha sido propuesta en [9] en donde “*un SIG es un sistema basado en computadora diseñado para modelar, capturar, almacenar, manipular, consultar, recuperar, analizar y visualizar información eficientemente, donde parte de la misma es de naturaleza geográfica*”.

Con esta definición se ve claramente que los SIG son más que herramientas para producir mapas. Mientras que en la cartografía tradicional el mapa es la base de datos, en los SIG el mapa solo es una proyección de una vista particular de una base de datos geográfica en un momento dado. De esta manera el usuario de un SIG posee un número ilimitado de alternativas de análisis y de alternativas para realizar mapas desde puntos de vista diferentes de acuerdo a diferentes aspectos de la información [3]. Las características exclusivas de la información geográfica generan requerimientos funcionales especiales en términos de modelos conceptuales y lógicos, estructura de datos, acceso a métodos, técnicas de análisis y procedimientos de visualización. Los mismos deben considerarse a la hora de iniciar un desarrollo de software para sistemas geográficos.

La combinación de estas dos áreas, DSBC y SIG, trae consigo varios aspectos a tener en cuenta tanto para la clasificación y selección de componentes como para los servicios requeridos de los mismos. Así hemos definido y trabajado en dos modelos, *de demanda* y *de oferta*, definiendo requerimientos especiales y soluciones de acuerdo a las restricciones y problemas que involucran cada uno de ellos.

El *modelo de demanda* se refiere a los requerimientos o servicios que necesitan los clientes de SIG (empresas que desarrollan SIG), los cuales, en un desarrollo basado en componentes, buscan en la Web componentes que brinden servicios que se adecúen a las necesidades de sus aplicaciones. Para esto, nuestro trabajo se centra en definir las herramientas necesarias para consultar en un sistema integrado los componentes que brindan esos servicios. Aquí, es crucial el nivel de especificidad y la forma en que los clientes solicitan dicha información para encontrar los componentes deseados.

El *modelo de oferta* se refiere a los requerimientos o servicios que brindan los desarrolladores de componentes para SIG. En general los componentes ofrecidos se publican en la Web con información asociada indicando aspectos técnicos y funcionales. Esto es llamado la documentación del componente la cual debe ser lo suficientemente clara, compacta y específica para ser bien comprendida por clientes.

De esta manera, para que los que demandan componentes se entiendan con aquellos que los ofertan, es fundamental la forma en que los componentes son publicados, llamado *proceso de publicación*, y la forma en que

los mismos son seleccionados de acuerdo a las necesidades de los clientes, llamado *proceso de selección* [10]. A su vez, para que ambos procesos se comuniquen surge la necesidad de utilizar un *modelo de mediación* que facilite tanto la recuperación de la información como la selección automatizada de componentes candidatos [6].

2. Contribuciones a la Fecha

Durante el año 2008 se ha trabajado en ambos modelos, de oferta y demanda, proponiendo metodologías que mejoren la forma de llevar a cabo las tareas incluídas en ellos.

Dentro del modelo de demanda hemos definido una taxonomía de servicios para SIG que brinde mejoras al proceso de selección y búsqueda de componentes en ese dominio. En este trabajo, publicado en [12], se estudiaron varias propuestas en la literatura que definen diferentes taxonomías y hasta incluso ontologías para clasificar la gama de servicios geográficos que pueden solicitarse [8, 11, 15]. Además se analizó la norma ISO 15115/OGC (Services) [2] y la Arquitectura de Servicios de OpenGIS (Service Architecture) [1] las cuales proveen un framework para crear software que permite a los usuarios acceder y procesar datos geográficos a partir de diferentes fuentes de datos a través de una interfaz genérica. Esta arquitectura se basa en el Modelo de Referencia para Procesamiento Distribuído Abierto (RM-ODP) y está conformada por un conjunto de componentes, conexiones y topologías. Utilizando toda esta información más la información obtenida de relevamientos efectuados a clientes SIG pertenecientes a organizaciones de la zona, se definió una taxonomía de servicios SIG, la cual cuenta con cuatro capas que incluyen un conjunto de servicios clasificados en cada una de ellas.

Dentro del modelo de oferta hemos definido otra taxonomía conteniendo la información necesaria para describir un componente y así ayudar al proceso de publicación. Este trabajo fue publicado en [5] y se basó principalmente en analizar cincuenta componentes SIG disponibles en la Web¹²³ y clasificarlos según la taxonomía propuesta. La misma se compone de 21 categorías distribuidas en tres grupos. El objetivo principal de este trabajo fue analizar qué información está realmente presente en la documentación de los catálogos de componentes para saber con que datos contar a la hora de implementar el modelo de mediación.

¹<http://www.componentsource.com/>

²<http://freegis.org/>

³<http://freshmeat.net/>

3. Trabajo Futuro

Para ambos modelos, oferta y demanda, se está actualmente trabajando en la creación de metodologías que extraigan la información necesaria para completar las dos taxonomías definidas en los trabajos previos. La idea principal es conocer qué información está disponible y qué información se necesita para poder implementar luego el modelo de mediación.

Para esto, se están analizando técnicas de lenguaje natural [7] que extraigan información de los catálogos de componentes en la Web y de los casos de uso (en UML [4]) para aplicarlas a ambos modelos.

4. Impacto y Tranferencia

El mayor impacto del proyecto se centra en la formación de recursos humanos. A la fecha algunos integrantes del grupo están iniciando sus maestrías en la Universidad Nacional de La Plata bajo la supervisión de investigadores de UNC. Al mismo tiempo, se transfieren técnicas y métodos de investigación que permitan el crecimiento sostenido de la actividad, tanto individualmente como grupalmente. La futura consolidación de estas actividades y relaciones permitiría pensar en futuras transferencias al medio y en mayor impacto desde el punto de vista de contribución al conocimiento científico.

5. Investigadores

Este proyecto reúne aproximadamente 6 (seis) investigadores, entre los que se cuentan docentes de la Universidad Nacional del Comahue (UNC) y de la Universidad de la Patagonia Austral (UNPA) y alumnos de la UNPA. Algunos de los docentes-investigadores de UNC-UNPA han comenzado sus carreras de postgrado y otros están terminado. Durante el año 2009 se avanzará en las 2 tesis de maestrías iniciadas con maestrandos en UNPA supervisados por investigadores de UNC.

Contacto: Ing. Gabriela Gaetán (ggaetan@uaco.unpa.edu.ar, gabrielagaetan@yahoo.com.ar)

Referencias

- [1] Open gis consortium. The OpenGIS Abstract Specification: Service Architecture, 2002.
- [2] Geographic information. Services, International Standard 19119, ISO/IEC, 2005.
- [3] P. Burrough and R. McDonnell. *Principles of Geographical Information Systems*. Oxford University Press, 1998.

- [4] Martin Fowler and Kendall Scott. *UML distilled (2nd ed.): a brief guide to the standard object modeling language*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2000.
- [5] G. Gaetán, A. Buccella, and A. Cechich. Un esquema de clasificación facetado para publicación de catálogos de componentes sig. In *Proceedings of the CACIC'08: XIV Congreso Argentino en Ciencias de la Computación*, Chilecito, La Rioja, Argentina, 2008.
- [6] L. Iribarne. *Un Modelo de Mediación para el Desarrollo de Software Basado en Componentes COTS*. Phd thesis, Univerdad de Almería, 2003.
- [7] Z. Kedad, N. Lammari, E. Métais, F. Meziane, and Y. Rezgui, editors. *Natural Language Processing and Information Systems*. Lecture Notes in Computer Science 4592. Springer-Verlag, 2007.
- [8] W. Li, S. Zhao, H. Sun, and X. Zhang. Ontology-based qos driven gis grid service discovery. In *Proceedings of the SKG'06: Second International Conference on Semantics, Knowledge, and Grid*, 2006.
- [9] M.A. Rodríguez Luaces. *A Generic Architecture for Geographic Information Systems*. PhD thesis, Univerdade da Coruña, 2004.
- [10] V. F. Lucena. *Flexible Web-based Management of Components for Industrial Automation*. Phd thesis, Stuttgart University, 2002.
- [11] M. Lutz. Ontology-based service discovery in spatial data infrastructures. In *Proceedings of the GIR'05: Proceedings of the 2005 workshop on Geographic information retrieval*, pages 45–54, New York, NY, USA, 2005. ACM.
- [12] V. Salda no, A. Buccella, and A. Cechich. Una taxonomía de servicios geográficos para facilitar la identificación de componentes. In *Proceedings of the CACIC'08: XIV Congreso Argentino en Ciencias de la Computación*, Chilecito, La Rioja, Argentina, 2008.
- [13] C. Szyperski. *Component Software-Beyond Object-Oriented Programming*. Addison-Wesley, 1998.
- [14] K. Wallnau, S. Hissam, and R. Seacord. *Building Systems from Commercial Components*. Addison-Wesley, 2002.
- [15] P. Yue, L. Di, P. Zhao, W. Yang, G. Yu, and Y. Wei. Semantic augmentations for geospatial catalogue service. In *Proceedings of the IGARSS'06: Proceedings of the IEEE International Conference on Geoscience and Remote Sensing Symposium*, Denver, USA, 2006.

Las Ubicuas Dependencias de Comparación de Conjuntos de Valores: Conceptualización, Análisis y Modelado

Viviana E. Ferraggine¹ Laura C. Rivero^{1,2}

¹ INTIA, Facultad de Ciencias Exactas, U.N.C.P.B.A., Tandil, Buenos Aires, Argentina

² LINTI, Facultad de Informática, U.N.L.P. La Plata, Buenos Aires, Argentina
{vferra, lrivero} at exa.unicen.edu.ar

CONTEXTO

Este trabajo forma parte de las actividades científicas de la Línea "Integridad en Base de Datos guiada por Requisitos", del Proyecto "Base de Datos y Procesamiento de Señales". Este proyecto es continuación de proyectos previos en la temática y está actualmente en desarrollo en el Instituto de Investigación en Tecnología Informática Avanzada (INTIA) de la Facultad de Ciencias Exactas de la U.N.C.P.B.A.

El objetivo general de la línea es estudiar los problemas de la elicitación de conocimiento del Universo del Discurso (UdeD), así como el impacto de la realidad estudiada sobre los problemas de modelado e integridad en el contexto de bases de datos relacionales y post-relacionales.

RESUMEN

Como respuesta a la incesante necesidad de incorporar todo el conocimiento que encierran las reglas del negocio en los sistemas de información han surgido diversas propuestas para representar dichos conceptos en los modelos de datos desde los primeros estadios de su concepción. Sin embargo, la captura de estas reglas en el modelo conceptual no es una tarea simple, dado que son de naturaleza compleja, tienen una estructura arbitraria, son volátiles y difíciles de aplicar.

Dado que la existencia de conjuntos de atributos compatibles y semánticamente vinculados conduce frecuentemente a comparaciones entre ellos, éste trabajo se enfoca en un tipo especial de reglas del negocio que pueden representarse por medio de *Dependencias de Comparación de Conjuntos de Valores*. Ellas limitan el modo en el cual un conjunto de datos se relaciona con otro; es así que se pueden establecer cuatro modos: mediante relaciones de inclusión, de exclusión, de igualdad o de superposición parcial o solapamiento. Esta parte de nuestro proyecto está dedicada a estudiar dichas dependencias desde un punto de vista con centro en la estructura sintáctica de los atributos que están sujetos a comparación, haciendo uso del marco teórico que ha brindado previamente el análisis de las dependencias de inclusión.

Palabras clave: Diseño conceptual, dependencias de comparación de conjuntos de valores, dependencias de inclusión, dependencias de igualdad, dependencias de exclusión, dependencias de superposición parcial, reglas del negocio, modelo de datos.

1. INTRODUCCIÓN

El desarrollo de la tecnología de base de datos ha sido estimulado por la necesidad de garantizar la consistencia de los datos almacenados y la demanda de procesar datos más complejos con mayor contenido semántico. Estos factores han promovido un tratamiento más comprehensivo de las propiedades de los datos motivando la evolución de los Sistemas de Gestión de Base de Datos (Database Management Systems, DBMS) que paulatinamente han ido incorporando nuevas facilidades para manejar estos aspectos.

Se pueden identificar dos dificultades principales para manejar las propiedades de los datos dentro del contexto de un paradigma de base de datos: 1) Son difíciles de categorizar, en especial las que son evidentes en el mundo real; 2) Su representación varía durante el proceso de diseño de software. La primera cae en el campo de la Ingeniería de Requisitos [18]. La correcta elicitación de requisitos de un producto de software es un factor importante en el éxito del proceso de desarrollo de software pero no es fácil el tratamiento de estos requisitos. Ellos tienen diferente naturaleza intrínseca y

pueden aparecer mostrando distintos perfiles. En muchos casos, algunos requisitos están total o parcialmente ocultos en la información recogida por los diseñadores del software [17].

La segunda cae en el campo del Modelado Conceptual y Modelado Lógico Temprano o genérico. Durante estas actividades, las particularidades del modelo de datos y los paradigmas de bases de datos generan una diversidad de formas en las que ellos pueden ser considerados. Características pragmáticas tales como el grado de adhesión de un DBMS u otras relacionadas con el desempeño pueden introducir, como consecuencia, mutaciones adicionales.

No todas las propiedades de los datos deben ser modeladas. Ellas deben ser cuidadosamente estudiadas para elucidar si son necesarias dentro del alcance y el objetivo del artefacto de software.

Cuando una propiedad describe los valores permitidos de un dato, se incluye entre las Propiedades de los Datos. Otros aspectos semánticos describen cualidades de las asociaciones entre diferentes datos; éstas se conocen como Propiedades de las Conexiones entre Datos (o Relaciones). Cuando una propiedad de los datos conlleva una semántica que es específica del UdeD, se denomina Propiedad General. Este concepto también es conocido como regla de negocio, y es una declaración que define o restringe algunos aspectos del negocio. Éstas tienen la función de mantener la estructura comercial o controlar o influir en la conducta de negocio [8, 10, 22, 4]. En el mundo real, las propiedades de los datos se perciben desde un punto de vista semántico, permitiendo su agrupación según la complejidad semántica intrínseca y su clasificación de acuerdo a su esencia. Pero en el mundo de las bases de datos se perciben básicamente desde un punto de vista sintáctico.

A medida que el desarrollo del artefacto del software avanza, debe encontrarse una representación de estas propiedades según una estructura lícita del modelo. Esta representación será ciertamente diferente en el modelo de datos conceptual, en el modelo de datos lógico y en el modelo de datos del motor de la base de datos [19, 20, 21].

Como consecuencia de lo antedicho, han surgido variadas propuestas para representar estas reglas del UdeD en modelos de datos desde los primeros estadios de su concepción [22, 8, 4, 15], sin perder de vista que son generalmente complejas, tienen una estructura arbitraria y son volátiles.

Aunque los modelos conceptuales actuales han contribuido en gran medida a la captura de los aspectos relevantes del UdeD, no tienen el poder expresivo suficiente como para representar integralmente la diversidad natural de las reglas del negocio. Los medios formales o semiformales que ofrecen los modelos permiten representar parte de su semántica, sobre todo cuando se trata de reglas estáticas. Muchas de ellas pueden ser representadas en algunas de las variantes del Modelo de Entidades y Relaciones (MER) [9, 24], por ejemplo por medio de restricciones de cardinalidad y dependencias de existencia. Otras reglas que requieren la formulación explícita de eventos, condiciones y acciones específicos, no pueden ser representadas de un modo equivalente. Esta limitación deriva al menos en dos consecuencias. Por un lado, ha promovido la materialización de numerosas propuestas de ampliación a este modelo (devenido en el modelo semántico estándar "de facto" [2, 5, 6, 15, 16]) por ejemplo sugiriendo el uso de OCL (Object Constraint Language), de sintaxis ad-hoc y aún de lenguaje natural para enriquecer los diagramas representativos del esquema conceptual [13]. Por otro lado, algunas restricciones de integridad que podrían haberse representado adecuadamente con los mecanismos de abstracción provistos por el modelo, se difieren infundadamente hasta etapas más tardías del proceso de diseño y frecuentemente su materialización no utiliza los recursos expresivos más apropiados para el caso.

Desde otro punto de vista, existen enfoques en defensa del MER original proponiendo una rigurosa y amplia utilización de las capacidades que éste ofrece, sin las extensiones que dificultarían su uso, interpretación, validación y formalización [14]. Adhiriendo a este punto de vista, este proyecto se desarrolla en dos direcciones complementarias: 1) ofreciendo siempre que sea posible un modo de representar cada regla utilizando las construcciones básicas del MER (de no ser posible, se deberían diseñar patrones en lenguaje declarativo o en código de programación); 2) sugiriendo estrategias de transformación en función de lo obtenido en 1), cuando se ha plasmado una regla de forma incorrecta o ambigua.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Nuestras actividades de investigación están centradas actualmente en un tipo de reglas de sencilla interpretación en el mundo real pero con dificultades manifiestas de representación en el MER: las "*Dependencias de Comparación de Conjuntos de Valores*" (*dccv*), nombre genérico otorgado a las expresiones que permiten establecer comparaciones entre los valores de conjuntos de atributos estructuralmente compatibles y semánticamente relacionados. Esta clase de reglas incluye dependencias de inclusión (*di*) [7, 10]; dependencias de exclusión (*de*) [1, 12]; dependencias de igualdad (*dig*) y dependencias de superposición parcial (*dsp*) [23].

Las *dccv* se analizan desde una base sintáctica, a fin de establecer: a) la posibilidad de su representación en un MER utilizando las estructuras estandarizadas; b) la necesidad de ampliar el modelo para capturar la semántica de esas reglas, cuando las estructuras estandarizadas no sean suficientes, y c) la posibilidad de realizar transformaciones estructurales que podrían proporcionar algún beneficio en relación con la semántica o la calidad de los aspectos operativos, d) proponer otros medios de representación cuando los medios analíticos disponibles sean insuficientes, [19, 20, 21].

Formalmente se definen de la siguiente manera:

Sean R_i y R_d relaciones, denominadas término izquierdo y derecho respectivamente; W_i y W_d conjuntos de atributos compatibles de R_i y R_d . r_i y r_d indican las extensiones de R_i y R_d respectivamente. Siendo t una tupla, una subtupla de t correspondiente al conjunto de atributos W se indica como $t.W$. El Cuadro 1 contiene las definiciones de cada uno de los cuatro tipos de dependencias.

Una *dependencia de inclusión (di)* se define como la existencia de atributos en una relación cuyos valores deben ser un subconjunto de los valores de atributos compatibles en otra relación (o la misma). Formalmente una *di* es una expresión $R_i[W_i] \subseteq R_d[W_d]$ y se satisface si para $t \in r_i$ existe una tupla $t' \in r_d$, tal que $t.W_i = t'.W_d$. Si W_d es la clave primaria (K_d) de R_d , la *di* es basada en clave y usualmente se denomina restricción de integridad referencial (*rir*) o simplemente referencia. W_i es una clave extranjera (FK_i) de R_i , y la *rir* correspondiente se indica $R_i[FK_i] \ll R_d[K_d]$.

Una *dependencia de igualdad (dig)* se define como la existencia de un conjunto de atributos en una relación cuyos valores deben ser los mismos que los correspondientes a un conjunto de atributos compatibles en otra relación (o la misma). Esto significa que cada miembro del primer conjunto debe ser un miembro del segundo conjunto y viceversa. Formalmente una *dig* se indica $R_i[W_i] = R_d[W_d]$, esto se satisface si para $t \in r_i$, existe una tupla $t' \in r_d$ tal que $t.W_i = t'.W_d$ y $t' \in r_d$, existe una tupla $t \in r_i$ tal que $t.W_i = t'.W_d$. Una *dig* es equivalente a dos dependencias de inclusión inversas $R_i[W_i] \subseteq R_d[W_d] \wedge R_d[W_d] \subseteq R_i[W_i]$.

Una *dependencia de exclusión (de)* se define como la existencia de un conjunto de atributos en una relación cuyos valores deben ser mutuamente excluyentes respecto de los correspondientes a un conjunto de atributos compatibles en otra relación (o en la misma). Esto significa que ambos conjuntos de atributos no tienen miembros en común. Formalmente una *de* es una expresión $R_i[W_i] \parallel R_d[W_d]$, y se satisface si no existe $t \in r_d$ y $t' \in r_i$, tal que $t.W_d = t'.W_i$.

Por último, una *dependencia de superposición parcial (dsp)* se define como la existencia de un conjunto de atributos en una relación cuyos valores se solapan parcialmente con los correspondientes a un conjunto de atributos compatibles en otra relación (o en la misma). Se denotará $R_i[W_i] \cap R_d[W_d]$ y tiene un significado equivalente a: $R_d.W_d \not\subseteq R_i.W_i \wedge R_i.W_i \not\subseteq R_d.W_d \wedge R_d.W_d \not\parallel R_i.W_i$, siendo W_d y W_i conjuntos de atributos compatibles. Esto es $r_d.W_d \cap r_i.W_i \neq \emptyset \wedge r_d.W_d - r_i.W_i \neq \emptyset \wedge r_i.W_i - r_d.W_d \neq \emptyset$.

CUADRO 1: Tipos de Dependencias de Comparación de Conjuntos de Valores

En los últimos años la investigación de las *dis* ha categorizado muchos de sus aspectos [5, 6, 16]. Un marco de referencia para su estudio puede encontrarse en [10], como así también el estudio de su caso particular, las *rir*s [11, 19, 20, 21]. Respecto de las *dis*, un análisis de su posible origen desde una perspectiva semántica puede leerse en [19], y [9] es un estudio preliminar de las *dis*.

Para el análisis de las *dccvs* se ha seguido lo expuesto en [20, 21], examinando la estructura sintáctica de los conjuntos de atributos involucrados, independientemente del dominio de aplicación. Esto permite inferir el posible tipo de vínculo semántico que intenta plasmar la dependencia.

La estructura sintáctica de los términos de una dependencia puede definirse en función de la posición del término respecto de la clave de la relación (correlación). Siendo W un conjunto de atributos de una cierta relación R , K la clave primaria de R y Z un subconjunto de atributos no pertenecientes a la clave, la Figura 1 ilustra las cinco posiciones posibles de los términos.

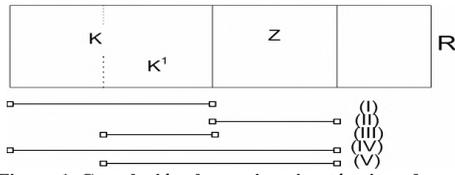


Figura 1. Correlación clave primaria - términos de una dependencia

I) $W \equiv K$; II) $W \equiv Z$; III) $W \equiv K^1$; IV) $W \equiv K \cup Z$ V) $W \equiv K^1 \cup Z$.
 K^1 es un subconjunto estricto de K , $K^1 \neq \emptyset$ y $Z \neq \emptyset$ para todos los casos. Teniendo en cuenta las correlaciones los 25 posibles pares de $\langle R_i.W_i, R_d.W_d \rangle$ quedan detallados en la Tabla 1.

Dado que las *dis* no son simétricas, la Tabla 1 muestra todas las posibles combinaciones de pares $\langle R_i.W_i, R_d.W_d \rangle$ y para las *des*, *digs* y *dsps* en las celdas sombreadas en gris se ha aclarado el caso para el cual resulta equivalente su análisis, por ejemplo el caso 2- $Z_i \theta K_d$ tiene un análisis equivalente (\equiv) al del caso 6- $K_i \theta Z_d$. θ es el tipo de *dcv*.

Tipo de Derecho	Término	Izquierdo	I	II	III	IV	V
		W_i	K_i	Z_i	K_i^1	$K_i \cup Z_i$	$K_i^1 \cup Z_i$
		W_d					
I	K_d		1- $K_i \theta K_d$	2- $Z_i \theta K_d$	3- $K_i^1 \theta K_d$	4- $K_i \cup Z_i \theta K_d$	5- $K_i^1 \cup Z_i \theta K_d$
II	Z_d		6- $K_i \theta Z_d$ ($\equiv 2$)	7- $Z_i \theta Z_d$	8- $K_i^1 \theta Z_d$	9- $K_i \cup Z_i \theta Z_d$	10- $K_i^1 \cup Z_i \theta Z_d$
III	K_d^1		11- $K_i \theta K_d^1$ ($\equiv 3$)	12- $Z_i \theta K_d^1$ ($\equiv 8$)	13- $K_i^1 \theta K_d^1$	14- $K_i \cup Z_i \theta K_d^1$	15- $K_i^1 \cup Z_i \theta K_d^1$
IV	$K_d \cup Z_d$		16- $K_i \theta K_d \cup Z_d$ ($\equiv 4$)	17- $Z_i \theta K_d \cup Z_d$ ($\equiv 9$)	18- $K_i^1 \theta K_d \cup Z_d$ ($\equiv 14$)	19- $K_i \cup Z_i \theta K_d \cup Z_d$	20- $K_i^1 \cup Z_i \theta K_d \cup Z_d$
V	$K_d^1 \cup Z_d$		21- $K_i \theta K_d^1 \cup Z_d$ ($\equiv 5$)	22- $Z_i \theta K_d^1 \cup Z_d$ ($\equiv 10$)	23- $K_i^1 \theta K_d^1 \cup Z_d$ ($\equiv 15$)	24- $K_i \cup Z_i \theta K_d^1 \cup Z_d$ ($\equiv 20$)	25- $K_i^1 \cup Z_i \theta K_d^1 \cup Z_d$

TABLA 1. Combinaciones de pares $\langle R_i.W_i, R_d.W_d \rangle$

3. RESULTADOS OBTENIDOS Y ESPERADOS

Siguiendo la primera dirección de desarrollo, se han determinado los patrones de representación de *dis* y *digs* en función de *rirs*, y se encuentra avanzado el mismo trabajo respecto de las *des* y *dsps*. En función de los logros anteriores, se han evaluado las ventajas y desventajas de aplicar las transformaciones según el nivel semántico del esquema original y la implementación de las operaciones de actualización básicas. Se pueden resumir las siguientes conclusiones parciales:

- Las operaciones de actualización que es factible llevar a cabo en el esquema original, también son factibles en el esquema transformado. Del mismo modo las inaplicables en uno y otro esquema son las mismas.
- Varias de las transformaciones estudiadas generan caminos cíclicos de integridad referencial que requieren una especificación por medio de aserciones o código ad hoc, siendo el "rechazo" la acción en caso de violación.
- Existen casos que pueden resolverse aplicando algunos axiomas de transformación básicos del modelo relacional (específicamente regla de pullback y axioma de transitividad). Otros mejoran la calidad de la representación mediante sencillas transformaciones fundamentadas en atributos y relaciones inaplicables. Se ha determinado el conjunto de casos para los cuales la calidad semántica no mejora porque las transformaciones aún siendo aplicables no permiten revelar objetos y relaciones mal representados, o bien porque no hay transformación que sea aplicable.
- Siempre que sea conveniente la transformación, el nuevo esquema adhiere al MER convencional o al MER extendido con agregaciones.
- Con respecto a razones de diseño, los esquemas transformados generalmente introducen nuevas tablas. Esto introduce un nuevo contexto de desempeño que debería ser analizado teniendo en cuenta las características del negocio (estabilidad de la información, frecuencia de operaciones de actualización, etc.)

- Finalmente, el esquema resultante debe re-analizarse basado en las expresiones de multiplicidades asociadas a los roles de las entidades participantes para determinar la validez de los caminos cíclicos y eliminar relaciones redundantes. Debido al agregado de tablas y relaciones se hace necesario un nuevo análisis de la coherencia global del esquema resultante. El procedimiento de validación del esquema conceptual permite la detección de multiplicidades y participaciones incorrectas [3].

La aplicación de transformaciones (en conjunción con el análisis de vínculos semánticos redundantes y el uso de mecanismos de evaluación de esquemas) ha resultado un marco metodológico apropiado para lograr una parcial pero útil reingeniería de esquemas de base de datos relacionales de pobre calidad, desactualizados o erróneos. La fase de transformación contribuye en el mejoramiento del nivel semántico del esquema, poniendo en evidencia objetos mal modelados y relaciones que los unen con otros objetos. Las transformaciones aplicables a las dependencias de inclusión y concretadas a través de una heurística fueron apropiadas también para las dependencias de igualdad. Algunos casos no pudieron mejorarse con este enfoque debido a que persisten dependencias intratables. No es posible establecer una regla estricta a seguir, pero generalmente, si una relación es conceptualmente relevante o las operaciones de actualización son poco frecuentes, la transformación es conveniente aunque aumentará el número de entidades y relaciones. Por otro lado, cuando el esquema resultante comience a ser poblado por entidades y relaciones antinaturales, el procedimiento de transformación debería detenerse.

Las *des* y *dsps* serán enfocadas desde otra perspectiva ya que ninguna de ambas está en el espíritu del MER, con la excepción de las jerarquías exclusivas y compartidas respectivamente.

Dentro de los futuros trabajos de investigación se incluye además el análisis de las acciones que podrían desencadenar las *dccvs* haciendo una analogía con las acciones referenciales de las *rirs* y el estudio de patrones de código (triggers) para soportar las *dis* remanentes en cada caso.

4. FORMACION DE RECURSOS HUMANOS

La investigación desarrollada ha permitido elaborar un programa para materias optativas centradas en tópicos avanzados de modelado conceptual de datos. En el mismo sentido varios grupos de alumnos han desarrollado o han profundizado en aspectos conceptuales relativos a este tema durante el desarrollo de su tesis de grado para aspirar al título de Ingeniero de Sistemas-UNCPBA.

5. REFERENCIAS

- [1] Albrecht, M., Buchholz, E., Düsterhöft, A., Thalheim, B., "An Informal and Efficient Approach for Obtaining Semantic Constraints using Sample Data and Natural Language Processing", Lecture Notes In Computer Science, 1358, pp. 1-28. Selected Papers from a 1st. Workshop on Semantics in Databases, Czech Republic. Thalheim, B.&Libkin, L.(Eds.), 1998.
- [2] Badía, A., "Entity-Relationship Modeling Revisited", SIGMOD Record, 33(1), 2004.
- [3] Boufarés, F., Bennaceur, H., "Consistency Problems in ER-Schemas for Database Systems", Information Sciences 163, pp. 263-274, 2004.
- [4] Business Rule Group, "Defining Business Rules ~ What Are They Really?", 1995, http://www.businessrulesgroup.org/first_paper/br01c0.htm.
- [5] Camps, R., "From Ternary Relationship to Relational Tables: a Case Against Common Beliefs", SIGMOD Record, 31(2), pp. 46-49, 2002.
- [6] Camps, R., "Transforming N-Ary Relationships to Database Schemas: an Old and Forgotten Problem", RR LSI-02-5-R, Univ.Politécnica de Catalunya, Spain, 2002.
- [7] Casanova, M., Fagin, R., Papadimitriou, C., "Inclusion Dependencies and their Interaction with Functional Dependencies", J. Computer and System Sciences, 28(1), 1984.
- [8] Ceri, S., Fraternali, P., "Designing Database Applications with Objects and Rules. The IDEA Methodology", Addison Wesley Publ. Co, 1997.
- [9] Chen, P., "The Entity-relationship Model: Toward a Unified View of Data", ACM TODS, 1(1), pp. 9-36, 1976.
- [10] Codd, E., "The Relational Model for Database Management. Version 2", Addison Wesley Publ. Co., 1990.
- [11] Cuadra, D., Nieto, C., Martínez, P., Castro, E., Velasco, M., "Preserving Relationship Cardinality Constraints in Relational Schemata", In Database Integrity: Challenges and Solutions, Idea Group Inc., Jorge Doorn and Laura Rivero Editors, pp. 66-112, 2002.

- [12] De Miguel, A., Piattini, M., Marcos, E., "Diseño de Bases de Datos Relacionales", Alfaomega Grupo Editor, 2000.
- [13] Goelman, D., Hussmann, H., "Using UML/OCL Constraints for Relational Database Design", In UML'99: The Unified Modeling Language - Beyond the Standard, Lecture Notes in Computer Science, Springer, 1723, France, R. & Rumpe, B. Editors, pp. 598-613, 1999.
- [14] Goelman, D., Song, I-Y., "Entity-Relationship Modeling Re-revisited", ER 2004, Proceedings 23rd International Conference on Conceptual Modeling, Shanghai, China. Lecture Notes in Computer Science, Springer, 3288, Paolo Atzeni, Wesley W. Chu, Hongjun Lu, Shuigeng Zhou, Tok Wang Ling Editors, pp. 43-54, 2004.
- [15] Halpin, T., "Information Modeling and Relational Databases. From Conceptual Analysis to Logical Design", Morgan Kaufmann Publishers, 2001.
- [16] Jones, T., Song, I-Y., "Analysis of Binary/Ternary Cardinality Combinations in Entity-Relationship Modeling", Data & Knowledge Engineering, 19 (1), pp. 39-64, 1996.
- [17] Jackson, M. (1995). Software Requirements & Specifications. A Lexicon of Practice, principles and prejudices. Workingham: Addison Wesley Publishers, ACM Press.
- [18] Loucopoulos, P & Karakostas. (1995). System Requirements Engineering. McGraw Hill International Series in Software Engineering.
- [19] Rivero, L., "Inclusion Dependencies", In Encyclopedia of Information Science and Technology. Idea Group. Inc., Mehdi Khosrow-Pour Editor, pp. 1425-1430, 2005.
- [20] Rivero, L., Doorn, J., Ferraggine, V., "Inclusion Dependencies". In Developing Quality Complex Database Systems: Practices, Techniques and Technologies, Idea Group Inc., Shirley A. Becker Editor, pp 261-278, 2001.
- [21] Rivero, L., Doorn, J., Ferraggine, V., "Enhancing Relational Schemas through the Analysis of Inclusion Dependencies", Int'l Journal of Computer Research, 12(4), Nova Publishers, pp. 489-511, 2004.
- [22] Ross, Ronald G., "The Business Rule Book. Classifying, Defining and Modeling Rules", Database Research Group, R. Ross Editor/Publisher, 1997.
- [23] Tari, Z., Bukhres O., Stokes J., Hammoudi S., "The Reengineering of Relational Databases Based on Key and Data Correlations", In Searching for Semantics: Datamining, Reverse Engineering, etc. Proc.7th IFIP 2.6 Working Conference on Database Semantics (DS-7), Chapman & Hall, Lausanne, S. Scappapietra and F. Maryanski Editors, pp. 183-214, 1998.
- [24] Teorey, T.J., "Database Modeling and Design. The Entity-Relationship Approach", Morgan Kaufmann Publishers, 1990.

Especificación Formal en RSL de una Infraestructura abierta y estándar de Servicios Web para Sistemas de Información Geográfica

Oscar Testa²; Daniel Riesco^{1,2}; Germán Montejano^{1,2}

¹Departamento de Informática

Universidad Nacional de San Luis

Ejército de los Andes 950 – (5700) San Luis – San Luis – Argentina

Tel.: +54-2652-424027 – Int. 251

[driesco, gmonte]@unsl.edu.ar – web: <http://www.sel.unsl.edu.ar>

²Departamento de Matemática

Universidad Nacional de La Pampa

Av. Uruguay 151 – (6300) Santa Rosa – La Pampa – Argentina

Tel.: +54-2954-425166– Int. 28

otesta@exactas.unlpam.edu.ar

CONTEXTO

El presente trabajo se enmarca en el Proyecto de Investigación: Ingeniería de Software, Conceptos, Métodos y Herramientas en un Contexto de “Ingeniería de Software en Evolución” – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Las líneas aquí presentadas surgen a partir de la realización de una Tesis de Posgrado en la Maestría en Ingeniería de Software auspiciada por el proyecto citado y una beca de la Universidad Nacional de La Pampa.

RESUMEN

En este trabajo se realizan aportes tanto a la tecnología de Servicios Web como a la de sistemas de información geográfica. Se presenta una infraestructura o framework de servicios Web de GIS para brindar solución a la problemática de diversidad de formatos e información. La especificación de la infraestructura se realiza a través de métodos formales, utilizando la herramienta RAISE – RSL. Esta infraestructura surge de la necesidad de contar en las organizaciones con herramientas que les permitan desarrollar sistemas de información geográfica en forma rápida y sencilla, siguiendo modelos establecidos y probados. Esta infraestructura permitirá la construcción de sistemas de información geográfica a partir de modelos probados y de utilización directa.

La utilización de RAISE RSL como lenguaje de especificación formal se basa en el hecho de su reconocimiento en la industria del

software para especificaciones formales de desarrollos reales.

Palabras clave: RAISE, RSL, Servicios Web, Framework, Infraestructura, GIS, Métodos formales, Sistemas de Información Geográfica.

1. INTRODUCCION

Los sistemas de información geográfica (GIS) han surgido hace ya más de 20 años y su inserción y uso dentro de las empresas y los organismos públicos ha sido muy importante. Cabe mencionar que más del 80% de los negocios requieren de servicios o prestaciones de localización geográfica. Los beneficios que aporta para la toma de decisiones, inteligencia que adiciona a los negocios, operaciones e información pública son innumerables, sin embargo muchos factores han mitigado estos beneficios. Algunos de los factores se mencionan a continuación (solo por mencionar los más importantes) [1], [2], [3]:

- La cartografía no se encuentra en el formato deseado.
- Los mecanismos de localización son particulares o propietarios del software adquirido y no se condicen con nuestra base de datos de clientes.
- La cartografía del área de cobertura de la aplicación no es brindado por los mismos proveedores ni en los mismos formatos.
- La forma de comunicación entre el servidor de mapas y el dispositivo

que realiza la petición no se encuentra estandarizada.

- El costo de licenciamiento del software de base es muy alto.
- Duplicación de información geográfica dentro de la misma región (empresa, País, etc.) sin posibilidad de interoperabilidad entre ellos.
- Alto costo de la cartografía de base.

Estos problemas que a menudo sucedían dentro de las organizaciones, luego sucedieron a nivel de regiones/países y se fueron convirtiendo en necesidades de los propios organismos de gobierno. Es decir que la necesidad de: compartir la cartografía de base, no tener que hacer enormes inversiones en licenciamiento de las herramientas, tener un repositorio común de la información, con el único propósito de mejorar el rendimiento de los SIG (aumentando la utilización de los mismos), mejorar la confianza en el formato y actualización de la información y en la reducción de costos, se fueron convirtiendo en las necesidades básicas para poder llevar adelante exitosamente un plan de implementación de una solución con sistemas de información geográfica.

Hoy, con el advenimiento en primer lugar de la tecnología de servicios Web y luego con los servicios Web orientados a SIG estos problemas han encontrado una solución.

Los servicios Web son servicios que están disponibles en Internet (intranet), utilizan un sistema de mensajería estandarizada y no están ligados a sistema operativo alguno o lenguaje de programación. A su vez los servicios Web pueden (es deseable que así sea) tener dos propiedades adicionales: se describen a sí mismos y se deben publicar para que se conozca [4].

Los servicios Web de SIG son una solución ideal para las organizaciones que desean hacer uso de información espacial y no encontrarse con los problemas mencionados anteriormente. Al poder hacer un “outsourcing” de los servicios de datos espaciales, se encuentra la

organización con una serie de beneficios: menor costo, menores riesgos, menores tiempos de desarrollo y disponibilidad de la información, menor cantidad de recursos especializados para poder llevar adelante la solución [5].

Como se menciona en el apartado anterior, los negocios de las organizaciones manejan un 80% de información geográfica o espacial, esto nos muestra lo importante que es contar en las empresas y organizaciones con un GIS. Es decir la utilización de un sistema de esta naturaleza mejoraría el desempeño y los resultados estratégicos de la organización, sobre todo si se piensa que existen constantemente presiones para que se tomen mejores decisiones y de una manera menos costosa y más eficiente.

El éxito de la implementación en las organizaciones de un sistema de estas características depende de muchos factores, pero como se introdujo anteriormente, básicamente los costos y la diversidad de información y formatos figuran entre los más destacados.

Estamos convencidos que si las organizaciones tuvieran herramientas que les permitan desarrollar sistemas de información geográfica en forma rápida y sencilla, siguiendo modelos establecidos y utilizando cartografía y datos espaciales existentes y probados podrían tener ventajas competitivas respecto del resto (o al menos podrían tener una visión de la realidad diferenciada que les permita analizar los eventos a través de una dimensión extra: la geográfica).

Debido a que no existen frameworks que permitan la generación automatizada de sistemas de información geográfica, se propone como objetivo construir, a través de la utilización de especificaciones formales, una infraestructura abierta y estándar de servicios Web, de forma tal que se pueda abordar la construcción de sistemas de información geográfica a partir de modelos probados y de utilización directa.

Compartir información geográfica y servicios asociados es una tarea que se viene investigando y haciendo ensayos desde hace varios años, su resurgimiento se debe principalmente a los organismos gubernamentales, principalmente de Estados Unidos.

Un primer proyecto, en el año 1994, fue iniciado por el entonces presidente William J. Clinton, a través de la orden presidencial 12906 para poner en marcha la Infraestructura Nacional de Datos Espaciales de EEUU (NSDI) [3], cuyo propósito fundamental fue el razonamiento de que “compartir conocimiento es fuente del crecimiento económico”.

Recién en Agosto de 2002, cuando es reafirmada la orden del año 1994, es que se comienza con la construcción del framework, el cual fue encomendado al Federal Geographic Data Committee para su coordinación. De todos modos esta infraestructura no contempla aún servicios para la utilización de los datos, sino simplemente el hecho de compartir la información espacial, tanto entre organismos gubernamentales o privados.

Otro de los proyectos iniciados para formar una Infraestructura de Datos Espaciales fue el de la Comunidad Económica Europea (INSPIRE), a través de la Directiva 2007/2CE del Parlamento Europeo y del Consejo, con fecha 14 de Marzo de 2007 [6]. En dicha Directiva se hace mención a los problemas relativos a la disponibilidad, calidad, organización, y puesta en común de información espacial y servicios asociados a las mismas.

A su vez define la infraestructura de información espacial como los metadatos, conjuntos de datos espaciales y los servicios de datos espaciales; los servicios y tecnologías de red; los acuerdos de acceso de la información. Y define servicios de datos espaciales como las operaciones que se pueden efectuar, a través de aplicaciones, sobre los

datos espaciales.

En Argentina en el año 2004, se crea el Proyecto Sistema de Información Geográfica Nacional de la República Argentina (PROSIGA), donde se plantea como objetivo vincular los generadores y usuarios de información espacial mediante una estructura nodal de intercambio de datos a través de redes públicas y privadas, permitiendo a la comunidad en general contar con una herramienta de base para la información general y la toma de decisiones basadas en criterios espaciales homogéneos [7].

2. LINEAS DE INVESTIGACION y DESARROLLO

El objetivo principal es construir, a través de la utilización de especificaciones formales, una infraestructura abierta y estándar de servicios Web, de forma tal que se pueda abordar la construcción de sistemas de información geográfica a partir de modelos probados y de utilización directa. La infraestructura que se propone especificar formalmente, se basa fundamentalmente en los trabajos de investigación [8] [9] [10] [11], en las infraestructuras desarrolladas de los distintos países del mundo [12] [13] [14], en la especificación de servicios OpenLS [15] realizada por el OpenGis Consortium [16] y en el aporte personal de los autores del proyecto que brinda la experiencia en la utilización de sistemas de información geográfica por más de 10 años.

El lenguaje de especificación formal usado en este caso es el RAISE Specification Language (RSL) dado que es reconocido en la industria del software para especificaciones formales de desarrollos reales.

RSL es un lenguaje formal, basado en el formalismo de la matemática usando conceptos tales como la teoría de conjuntos, lógica de primer orden, lógica de orden superior, destacándose entre otros conceptos matemáticos, y, que están netamente orientados a construir modelos, ya sea describiendo un dominio de la realidad o

describiendo una herramienta a desarrollar y sus requerimientos

3. RESULTADOS OBTENIDOS/ESPERADOS

Se evidencia entonces que de existir un framework con esas características, más aún si se encuentra basado en especificaciones formales, el desarrollo de aplicaciones GIS con interface Web se vería beneficiado ampliamente tanto en el ámbito privado como en el de organismos públicos o gubernamentales. Algunas de las áreas (sólo por nombrar algunas y sin ser una lista exhaustiva) que podrían utilizar este tipo de arquitectura o framework serían las de turismo, servicios públicos, emergencias (catástrofes naturales, provocadas por el hombre), prevención de delitos, incendios.

De acuerdo a lo planteado se desprende que existe un campo de trabajo interesante (y con áreas no cubiertas) en el ámbito de servicios Web de mapas, y más aún en su especificación formal.

4. FORMACION DE RECURSOS HUMANOS

Además de los resultados obtenidos / esperados en el punto 3, se espera como resultado en la formación de recursos humanos, la continuación de esta misma línea de proyecto como tesis doctoral del tesista y la mayor interrelación con la Universidad Federal de Minas Gerais con la que contamos con un convenio con tal objetivo como parte de él. Adicionalmente, se espera que otras tesis de Maestría, así como tesinas de Licenciatura surjan a partir de los logros obtenidos en la presente línea investigativa.

5. BIBLIOGRAFIA

- [1] Kolodziej, Kris, "OpenGIS Web Map Server Cookbook", OGC Document, 2003.
- [2] Harmon, John, Anderson, Steven, "The design and implementation of Geographic Information Systems", Wiley, 2003.
- [3] Somers, Rebecca, Claire, Somers-St., "Framework - Introduction and Guide", Federal Geographic Data Committee, 1997.
- [4] Ethan, Cerami, "Web Services Essentials - Distributed Applications with XML-RPC, SOAP, UDDI & WSDL", O'Reilly, 1th ed, 2002.
- [5] "Gis Web Services - The changing GIS Landscape", White paper, GisFactory, 2003.
- [6] "Directiva 2007/2/CE", Parlamento Europeo, 2007.
- [7] Anexo I - Términos de Referencia del Proyecto Sistema de Información Geográfica Nacional de la República Argentina, 2004.
- [8] Luo Yingwei, Wang Xiaolin and Xu Zhuoqun, "Design of a Framework for Multi-User/Application Oriented WebGIS Services", Computer Networks and Mobile Computing, 2001. Proceedings. 2001 International Conference on, IEEE, 2001.
- [9] Zhiming GUI, Kai SONG, "Building Improved GIS Service Based on WSRF", International Conference on Internet Computing in Science and Engineering, IEEE, 2008.
- [10] Yue Li, Zhong Xie, and Zhiyong Huang, "Design and Implementation of a WAP GIS Framework", Web Information Systems Engineering (Workshops), 2002. Proceedings of the Third International Conference on, IEEE, 2002.
- [11] Yuan Ying, Bian Fuling, "An Open Sharing and Interoperating Platform for Spatial Information Based on GIS Web Services", Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on, IEEE, 2007.
- [12] "Infrastructure for Spatial Information in the European Community", <http://www.ec-gis.org/inspire/>
- [13] "National Spatial Data Infrastructure", <http://www.fgdc.gov/framework>
- [14] "Proyecto Sistema de Información Geográfica Nacional de la República Argentina", <http://www.sig.gov.ar/>
- [15] "OpenGis Location Services (OpenLS): Core Services", Open Gis Consortium, 2004.

- [16] “The Open Geospatial Consortium Inc.”, <http://www.opengeospatial.org/>

Herramienta para consultas complejas orientada a usuarios finales

Ing. Pablo A. Vaca – Departamento de Sistemas de Información - Facultad Regional Córdoba - Universidad Tecnológica Nacional. E-mail: vacapablo72@gmail.com

Ing. Maximiliano Abrutsky - Departamento de Sistemas de Información - Facultad Regional Córdoba -Universidad Tecnológica Nacional. E-mail mxtrula@gmail.com

CONTEXTO

El tema que se presenta se encuentra enmarcado por el Proyecto de Investigación PID Promocional de la UTN: Análisis y aplicación de metodologías para la generación de consultas complejas utilizando esquemas OLAP. La presente etapa esta enmarcada dentro de un proyecto que se refiere al estudio de generador y facilitadores de consultas complejas a los usuarios finales. El mismo es el encargado de la extracción de los datos desde las fuentes de para su posterior utilización en el análisis de la información que los mismos contienen.

RESUMEN

La línea de investigación en curso persigue el objetivo de facilitar a los usuarios finales la realización de consultas complejas que respondan a preguntas del negocio. Siendo el principal punto de dificultad la comunicación con la fuente de datos y en este punto en particular la sintaxis a utilizar para extraer la información.

Es así que se plantearon tres alternativas o cursos de avance en la construcción de soluciones, la primera: utilizar el lenguaje nativo de la fuente datos, la segunda: desarrollar un intérprete y compilar que unifique el lenguaje de cara a los usuarios y se encargue de traducir el mismo en el lenguaje específico de la fuente de datos y por último una alternativa mixta, es decir una parte que sea lenguaje propietario y otra parte que utilice el lenguaje de la fuente de datos.

Con estas tres alternativas planteadas se ha desarrollado una serie de pasos tendientes a decidir por cual de nos inclinaremos, colocando en la balanza las ventajas y desventajas de ellas.

Se presenta en este trabajo los avances alcanzados y las líneas de acción propuestas en el marco del proyecto de investigación en curso.

Palabras clave: *Intérpretes – Código Fuente – Lenguajes de Programación – Datamart – OLAP – Extractores de datos*

1. INTRODUCCION

Para la realización del módulo que se encargue de extraer los datos, los cuales se pueden encontrar en una base de datos relacional, una archivo de texto, un archivo dbf o del tipo de x-base, es necesario que todo programa de análisis de información cuente con extractores de información, los cuales no son otra cosa que programas que se comunican a dichas fuentes, y mediante la ejecución de una serie de reglas y sintaxis extraen los datos.

Ahora bien, estas reglas que usan estos programas pueden ser las reglas propias de las fuentes de datos, o sino pueden ser reglas propias que disponga la herramienta en cuestión, y dentro de esta última alternativa disponemos de dos caminos a seguir, uno de los cuales es desarrollar totalmente un lenguaje que se encargue de todos los aspectos de la extracción o en su lugar disponer de un lenguaje que en algunos puntos utilice parte del lenguaje de la fuente de datos.

A continuación se plantean ventajas y desventajas de cada una de estas opciones, las cuales son enumeradas a los fines de justificar luego la elección que tuvo el equipo de investigación sobre cual camino utilizar.

Fuentes consultadas: ver referencia bibliográfica Nro 13.

a. Utilización de lenguajes propietarios

Se entiende por lenguajes propietarios a los que están incorporados como parte de las herramientas que los usan, como por ejemplo el lenguaje PL/SQL que es parte de las bases de datos Oracle y el cual es usado en la elaboración de procedimientos almacenados o disparadores.

A continuación se enumeran ventajas y desventajas observadas si utilizamos este tipo de lenguajes para la extracción de la información.

Fuentes utilizadas: ver referencias bibliográficas Nro 10 – 11 – 12.

Ventajas.

- No se requiere de una traducción antes de ejecutar la extracción en la base de datos, esto redundaría en un mejor rendimiento en la extracción de la información.
- Es posible encontrar expertos en cada uno de los lenguajes, lo cual disminuye los tiempos de aprendizaje de la herramienta.
- Trabajar con los tipos de datos nativos, se disminuyen la probabilidad de introducir errores en la traducción de la información.
- Se podrían ejecutar procedimientos almacenados en la base de datos, los cuales ayudarían a transformar la información.
- No existe límite en cuanto a las fuentes de información a acceder, siempre y cuando tengan un modo de conexión (ej. ODBC u OleDb).

Desventajas.

- Si se quiere acceder a variadas fuentes de datos será necesario conocer todos los lenguajes de cada una de ellas. Esto incluye todas las funciones, por ejemplo se dispondrán de distintas funciones que realicen la misma tarea como ser el formateo de fechas o números.
- Para proyectos que involucren varias fuentes de datos se deberá contar con un súper especialista o con muchas personas que conozcan cada una parte.
- Queda en manos de cada fuente de datos la validación de la sintaxis y tendríamos que manejar innumerables códigos de error para facilitar al usuario la interpretación del mismo.
- Restricción a las fuentes de información para las cuales hayamos desarrollado los traductores correspondientes.
- Si un usuario tuviese que aprender todos los lenguajes antes de usar la herramienta el costo de dicho aprendizaje sería muy alto.

b. Desarrollar un lenguaje propio

Otra alternativa con la que cuentan algunas herramientas es disponer de un lenguaje propio, es decir que cuentan con su propia sintaxis y reglas con las cuales se encargan de extraer los datos desde las fuentes de información, si bien no existen herramientas que apliquen este punto en forma completa es decir que no utilicen para nada lenguaje propio de la fuente de datos, existen lenguajes de programación que para la extracción de datos encapsulan todo lo relacionado con el origen y permiten crear programas que sirvan para varias bases de datos relacionales por ejemplo. Un ejemplo de esto son los Framework de persistencia utilizados en el modelo MVC de desarrollo tales como hibernate.

A continuación se enumeran ventajas y desventajas de desarrollar un módulo con estas características.

Fuentes consultadas: ver referencia bibliográfica Nro 14.

Ventajas.

- Es necesario aprender un solo lenguaje, lo cual disminuye los tiempos de capacitación para los usuarios.
- El conjunto de funciones es único evitando problemas con el uso equivocado y también con las posibles incompatibilidades con los tipos de datos.
- Uniformidad en la escritura del código necesario para extraer la información, facilitando la lectura y comprensión del mismo.

Desventajas.

- Es necesario desarrollar los traductores para la parte del código de extracción que interactúe con las fuentes de datos, dado que será imposible sino lograr comunicarse con las mismas.
- Se agrega una nueva etapa de traducción lo cual atenta contra el rendimiento total de la etapa de extracción de datos.
- La complejidad del nuevo lenguaje puede ser muy elevada si no se delimita adecuadamente el alcance del mismo. Esto podría generar una carga de trabajo excesiva, siendo que solo forma una etapa del proyecto total.
- Restricción a las fuentes de información para las cuales hayamos desarrollado las interfaces necesarias, aunque esto tiene un costo menor que desarrollar un traductor.

c. Desarrollar un lenguaje híbrido

La mayoría de las herramientas utilizan una variante de las dos expuestas anteriormente, es decir cuentan con una parte propietaria y otra parte en la cual utilizan el lenguaje propio de la base de datos o fuente que contiene la información, con lo cual se disminuyen los costos de desarrollo del módulo ya que se reutilizan las funcionalidades del lenguaje ya existentes..

A continuación se muestran ventajas y desventajas de esta alternativa.

Fuentes consultadas: ver referencias Nro 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 15.

Ventajas.

- Es posible unificar las funciones que manipulen

los datos extraídos, como ser funciones de formateo de fechas y números, manejos de cadenas de caracteres, de conversión de datos.

- No es necesario desarrollar traductores que se encarguen por ejemplo de transformar una consulta en el lenguaje específico que utiliza cada fuente de datos.
- Al tener una parte propietaria, no es necesario conocer innumerables funciones para trabajar con los datos lo cual permitiría un rápido aprendizaje de la herramienta.
- Se simplifica el desarrollo de la parte propietaria, lo cual permite incorporar aspectos comunes a un menor costo.
- Al tener una parte en lenguaje de la fuente de datos es posible utilizar procedimientos almacenados.

Desventajas.

- Depende de la fuente de datos. Lo cual implica cierto grado de conocimiento de la misma.
- Es necesario tener mucho cuidado en la etapa de diseño, ya que es necesario definir cuidadosamente las interfaces entre la parte propietaria y el lenguaje de datos que son recuperados de la fuente de información

INFORMACIÓN COMPLEMENTARIA.

Fuentes consultadas: ver referencias Nro 1 a 9.

Desarrollo de un lenguaje propio o un híbrido:

El desarrollo de un lenguaje propio o un lenguaje híbrido merece una breve explicación ya que es una parte del proyecto que permite aportar varios conocimientos en distintas áreas, por otra parte el desarrollo de un lenguaje propio o híbrido persigue aportar para el logro del objetivo principal del proyecto, que es facilitarle al usuario la labor de la realización de consultas complejas.

La definición de un lenguaje propio conlleva la necesidad de la creación de un intérprete del mismo, el cual conceptualmente tiene la finalidad de analizar y ejecutar el código escrito en el lenguaje propietario, instrucción por instrucción a medida que sea necesario. En este sentido deberemos contemplar dos alternativas, usar aplicaciones para la creación del intérprete o desarrollar totalmente un intérprete.

Alternativas.

Utilización de aplicaciones para la creación de analizadores léxicos y sintácticos

- Lex & Yacc (lenguaje c)
- TP Lex & Yacc (lenguaje Pascal)

- Flex & Bison (versión de código abierto, con licencia GPL de Lex & Yacc)
- JLex & CUP (lenguaje Java)

Desarrollo íntegro del intérprete

- Utilizando la API de Reflection (Java, .Net, etc).
- En algún lenguaje que manipule tecnologías XML, utilizando las como base (XSLT, XPath, XQuery, Schema, etc).
- Otros lenguajes / Tecnologías.

2. LINEAS DE INVESTIGACION y DESARROLLO

El proyecto de investigación se encuentra dentro de lo que se llama investigación aplicada, más específicamente investigación en Inteligencia de negocios.

Esta enfocada en investigar y evaluar software existente, comparar sus características y encontrar puntos en los cuales son plausibles de mejoras.

También se orienta a generar conocimientos que puedan ser transferidos a la comunidad académica en general.

3. RESULTADOS OBTENIDOS/ESPERADOS

En base a la investigación realizada se observa que conceptualmente el funcionamiento de las aplicaciones antes mencionadas es el mismo, variando solamente el lenguaje de programación del código generado como salida. Es por tanto que se ha tomado la decisión de optar por Lex & Yacc por ser el de mayor popularidad y ser la más evolucionada.

Los patrones son codificados para utilizados como entrada en Lex, quien con la entrada de código fuente genera un analizador léxico en c. El analizador léxico busca coincidencias en el código fuente de entrada respecto a los patrones definidos, convirtiendo dichas cadenas en tokens quienes son una representación numérica de las cadenas de caracteres simplificando el procesamiento.

La gramática es la entrada para Yacc, quien también genera en código c un analizador sintáctico o parseador. Éste aplica las reglas gramaticales que le permiten utilizando los tokens provenientes del analizador léxico crear el árbol de sintaxis, en este se impone una estructura jerárquica de los tokens.

Finalmente el último paso es la generación de código el cual procesa al árbol de sintaxis. Algunos compiladores generan código máquina.

Las reglas de traducción en Lex representan la escritura de todos los patrones, o sea, a tal expresión regular, tal acción. En cambio para Yacc, cada regla consta de una producción de la gramática y la acción semántica asociada. Todo el conjunto de dichas reglas representan a la gramática libre de contexto.

Como resultado de esta etapa de la investigación estamos en condiciones de adoptar una postura sobre la alternativa a utilizar, a la cual hemos escogido por brindar facilidades a los usuarios finales y al equipo de desarrollo de la herramienta, dicha alternativa es desarrollar un híbrido utilizando una aplicación que ayude en el desarrollo de dicho intérprete.

La alternativa elegida permite al equipo de desarrollo adquirir conocimientos y habilidades en distintas áreas de la informática relacionadas con la creación de intérpretes, las cuales pueden ser utilizadas para generar transmisión del conocimiento dentro y fuera de la universidad.

Los objetivos en curso son los siguientes:

- Desarrollar una herramienta que extraiga y contenga la información para que los usuarios finales puedan realizar consultas complejas sobre dicha información sin requerir conocimientos técnicos específicos sobre consultas.
- Desarrollar modelos de extracción de datos.
- Desarrollar un modelo teórico de un generador de consultas complejas.

4. FORMACION DE RECURSOS HUMANOS

Se formarán recursos humanos en los siguientes aspectos:

* Dirección de proyectos, ya que la participación como co-director del proyecto por parte del Ing Pablo A. Vaca lo formará en los aspectos de gestión y dirección de proyectos de investigación.

* Recursos humanos formados en las áreas de Inteligencia de negocios, al contar con la participación de otros egresados y alumnos, estos adquirirán conocimientos avanzados de DataWarehouse, y de otros aspectos de la especialidad Inteligencia de Negocios, aspectos muy requeridos en las organización actuales y futuras.

* Recursos humanos formados en áreas de desarrollo de algoritmos de extracción y compresión de datos. Estas personas obtendrán como resultado de su participación acabados conocimientos de técnicas de extracción de datos, técnicas de compresión de datos y técnicas de almacenamiento y manejo en memoria de los datos.

* Recursos humanos formados en la metodología de documentación y modelado UML. La cual es

ampliamente usada en todas las organizaciones que desarrollan sistemas informatizados.

5. BIBLIOGRAFIA

1. lex & yacc, Second Edition By John Levine, Tony Mason, Doug Brown - ISBN 10: 1-56592-000-7 | ISBN 13: 9781565920002
2. Compilers: Principles, Techniques, and Tools. by Alfred V. Aho, Ravi Sethi, Jeffrey D. Ullman, Monica S. Lam - ISBN-13: 9780321486813
3. Writing Compilers and Interpreters, by Ronald Mark - ISBN-10: 0471113530 # ISBN-13: 978-0471113539
4. Compiladores - Teoria y Construcción, by F. Sanchis Llorca - ISBN-10: 8428314691 # ISBN-13: 978-8428314695.
5. Web Oficial: <http://dinosaur.compilertools.net/>
6. Publicaciones en ePapers: <http://epaperpress.com>
7. Johnson, Stephen C. [1975]. Yacc: Yet Another Compiler Compiler. Computing Science Technical Report No. 32, Bell Laboratories, Murray Hill, New Jersey.
8. Lesk, M. E. and E. Schmidt [1975]. Lex – A Lexical Analyzer Generator. Computing Science Technical Report No. 39, Bell Laboratories, Murray Hill, New Jersey. A COMPACT GUIDE TO LEX & YACC. - by Tom Niemann
9. Wikipedia: Artículos introductorios de lingüística computacional. Listado: http://en.wikipedia.org/wiki/Category:Computational_linguistics.
10. Introducción a los Sistemas de Bases de Datos - C.J. Date. ISBN: 968-444-419-2 - Editorial Pearson.
11. Lantimes SQL 2 - Autor Groff
12. Oracle9i ISBN: 84-481-3653-5 Editorial: Mc Graw Hil.
13. Data Warehousing. Autor Ing. Alejandro Sueldo. Página de Internet. http://alejandrosueldo.com.ar/joomla/index.php?option=com_content&task=view&id=3&Itemid=2 - Consultado en el mes de Abril de 2007.
14. <http://www.hibernate.org/>. Documentación contenida en el sitio en forma on-line y en formato pdf.
15. <http://www.qliktech.com>. Documentación del producto QlikView. Herramienta de BI basada en conceptos de Lenguaje de consultas por asociación basado en memoria.

Geometría Computacional y Bases de Datos Espacio-Temporales

Maria Gisela Dorzán, Edilma Olinda Gagliardi, María Teresa Taranilla

Departamento de Informática, Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis, Argentina
{mgdorzan, oli, tarani}@unsl.edu.ar

Gregorio Hernández Peñalver

Departamento de Matemática Aplicada, Facultad de Informática
Universidad Politécnica de Madrid, España
gregorio@fi.upm.es

CONTEXTO

El proyecto Tecnologías Avanzadas de Bases de Datos, N° 22/F614, de la Universidad Nacional de San Luis, tiene como objetivo principal el estudio de bases de datos avanzadas, o no convencionales, en donde se involucra el diseño y desarrollo de herramientas para administrar eficientemente sistemas de bases de datos no estructurados. Este proyecto posee tres líneas de investigación orientadas al desarrollo de nuevos modelos para buscar y administrar la información en almacenamientos de este tipo, donde los escenarios de exploración requieren modelos más generales tales como las bases de datos espacio-temporales, bases de datos de texto, espacios métricos, entre otros. En particular, la línea de investigación Geometría Computacional y Bases de Datos Espacio-Temporales, perteneciente a dicho proyecto, viene desarrollando actividades vinculadas al tratamiento de objetos de tipo geométrico, que son de utilidad en diversos campos de aplicación, por ejemplo, robótica, visión artificial, computación gráfica, sistemas de información geográfica, computación móvil, diseño asistido por computadora, entre otras, y que se relacionan en tales bases de datos.

Entre los problemas de estudio, podemos mencionar el diseño de índices apropiados para el almacenamiento y consulta de datos de tipo espacio-temporales, que poseen propiedades geométricas, y que son integrales y adecuados para resolver consultas espacio-temporales. Además, entre los problemas de índole geométrica, han surgido aquellos para los que no existen algoritmos eficientes que los solucionen, porque son del tipo NP-duro, o aún no han sido objeto de estudio, y para los cuales

proponemos encontrar soluciones aproximadas, mediante el uso de técnicas metaheurísticas.

Esta línea viene desarrollándose desde el año 2002, en forma conjunta con investigadores afines de proyectos locales y de la Universidad Politécnica de Madrid, en el marco de los convenios de cooperación institucional.

RESUMEN

Respecto de las temáticas de investigación, hemos vinculado las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas, debido a que en diversas aplicaciones dentro del campo de las Ciencias de la Computación se requiere la construcción y manejo de diferentes objetos geométricos, con propiedades deseables. También, se requiere de repositorios no tradicionales, que conlleven a nuevos modelos de bases de datos para administrar y buscar información en ellos. Así, surge la necesidad de estudiar modelos como las bases de datos espacio-temporales. En particular, algunos de los problemas estudiados necesitan algoritmos eficientes para su resolución, pero dada su naturaleza NP-dura, utilizamos técnicas metaheurísticas para hallar soluciones aproximadas. En este trabajo, presentamos los tópicos más relevantes, actualmente en estudio, con las propuestas más recientes y/o de interés.

Palabras clave: Bases de datos, bases de datos espacio-temporales, geometría computacional, metaheurísticas.

1. INTRODUCCIÓN

Con la evolución de las tecnologías de información y comunicación, han surgido repositorios o almacenamientos no estructurados de información. Se consultan nuevos tipos de datos tales como datos geométricos, imágenes, audio y video, entre otros, siendo en algunos casos, que la información no necesariamente se estructura en claves y registros. Y en particular, algunas aplicaciones requieren guardar y consultar información histórica y actual, acerca de los cambios de forma y/o posición que tuvieron los objetos de estudio en diferentes escenarios a lo largo del tiempo, por lo que se requiere de modelos de bases de datos espacio-temporales para tales requerimientos. En este contexto, es necesario contar con herramientas teóricas, de base, que permitan modelar estos tipos de datos, realizar operaciones sobre ellos, definir lenguajes de consulta, analizar su expresividad, entre otras propiedades.

En el ámbito de la Geometría Computacional, se estudian problemas de tipo geométrico, respecto del diseño de algoritmos y/o estructuras geométricas adecuadas para su resolución. Por lo que, la geometría nos permite tener una visión diferente de los problemas, tal que cuando se aplican técnicas o estructuras geométricas en la resolución de los mismos, las soluciones pueden ser más apropiadas y/o viables de realizar.

Asimismo, en Geometría Computacional, surgen problemas de naturaleza NP-dura, o bien, para los cuales no se han encontrado hasta el momento algoritmos eficientes que los solucionen; incluso, puede ser que se desconozcan soluciones, pero que en cualquier caso puede existir la necesidad de encontrar respuestas a tales problemas, aunque las soluciones sean aproximadas o de naturaleza heurística.

En particular, resulta de interés el estudio de estructuras geométricas y problemas relacionados a las mismas, mediante el análisis de propiedades que constituyen medidas de calidad que permiten estimar cuán “buena” es la misma. La construcción de estas estructuras óptimas es un problema que no necesariamente

admite soluciones exactas con algoritmos eficientes; en algunos casos está demostrado que son problemas NP-duros. Por ello, resulta un campo de interés, en el que se pueden lograr resultados significativos dado que es un área emergente en la comunidad científica, lo que le agrega mayor relevancia para abordar futuras investigaciones. El uso de algoritmos aproximados nos puede dar soluciones cercanas a las óptimas, los cuales pueden ser específicos para el problema tratado o formar parte de una estrategia general que se puede aplicar en la resolución de distintos problemas, como lo son las técnicas metaheurísticas.

Por lo expuesto, en esta línea nos dedicamos al estudio, diseño y desarrollo de índices espacio-temporales, aplicables a diversos escenarios de movimiento (redes, espacios libres de obstáculos, etc.), considerando la geometría como una disciplina marco en la cual se formalizan aspectos propios de los problemas involucrados. Además, en este contexto, nos planteamos el estudio de optimizaciones de estructuras tales como triangulaciones, pseudotriangulaciones y poligonizaciones. Entre las principales medidas de calidad consideramos el peso, la dilación, el número de apuñalamiento, entre otras. También, nos dedicamos al estudio de problemas de minimización de la suma de longitudes, minimización del camino más largo entre sus vértices, optimización del vector de ángulos de la triangulación, entre otros ejemplos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La Geometría Computacional estudia técnicas, algoritmos y estructuras de datos adecuadas para la resolución de problemas geométricos. Estos problemas se encuentran en diversas áreas de las Ciencias de la Computación, tales como Computación Gráfica, Planificación de Movimientos, Visión Computacional, Robótica, Recuperación de la Información, Bases de Datos Espacio-Temporales, entre otras. En particular, los problemas de interés en la línea de investigación son aquellos vinculados al diseño de índices espacio-temporales para resolver integralmente

consultas espacio-temporales y su vinculación con la problemática de objetos móviles; y al tratamiento de problemas geométricos de naturaleza NP-dura. En ambos casos, proponemos la búsqueda de soluciones para ellos, aunque éstas sean aproximadas. Con este fin, utilizamos las técnicas metaheurísticas, que proporcionan estrategias adecuadas para la resolución aproximada de múltiples problemas de optimización, en especial de aquellos para los que no existen algoritmos eficientes de resolución.

Entre los tópicos de estudio de la línea, destacamos los siguientes:

i) Estudio de triangulaciones de puntos que cumplan ciertas medidas de calidad como son el peso, la dilación y el número de apuñalamiento. Optimización de triangulaciones de peso, dilación y número de apuñalamiento mínimos utilizando técnicas metaheurísticas. Estudio de problemas de visibilidad o iluminación en un tipo particular de superficies poliédricas, los poliedros-terreno. Resolución aproximada mediante metaheurísticas a los problemas de iluminación en triangulaciones planas. [MR06] [FLM08] [CS89] [M04].

ii) Optimización de rutas de vigilancia entre obstáculos poligonales mediante la aplicación de técnicas metaheurísticas. Resolución aproximada de diferentes variantes de problemas de vigilancia en polígonos, aplicando metaheurísticas. Metaheurísticas para la optimización de poligonizaciones de puntos en el plano. [Fe00] [LHL93] [CNN93] [Zy06].

iii) Optimización de pseudotriangulaciones, que satisfagan propiedades como bajo peso, dilación pequeña, número de apuñalamiento pequeño, entre otras, mediante la aplicación de técnicas metaheurísticas. [PV96] [RSS06] [GL07].

iv) Indexación espacio-temporal sobre objetos en movimiento para diversos escenarios. Estudio de nuevas estrategias de ruteo en grafos geométricos para su aplicación a objetos móviles. Aplicaciones de la vida real con herramientas de Geometría Computacional

y Bases de Datos Espacio-Temporales. [DGG07] [GGS07].

3. RESULTADOS OBTENIDOS/ESPERADOS

Entre los trabajos de investigación, podemos mencionar los siguientes:

Existen aplicaciones que requieren manejar objetos espacio-temporales, es decir, objetos cuya posición espacial o forma cambia en distintos instantes de tiempo. Por consiguiente, deberían considerarse métodos de acceso espacio-temporales, que sean capaces de soportar estos nuevos tipos de datos y de responder tipos de consultas apropiados, satisfaciendo de esta manera, las demandas de los usuarios. En general, TimeSlice, Intervalo, Eventos y Trayectoria son los tipos de consultas de interés. En la línea de investigación, estudiamos métodos de acceso espacio-temporal, que integralmente, permiten resolver estos cuatro tipos de consultas, sin aumentar la complejidad espacio-temporal, en diversos escenarios. Desarrollamos las estructuras de almacenamiento, los algoritmos de consulta y la evaluación experimental, mostrando el buen desempeño de los distintos índices en aplicaciones de diferentes magnitudes respecto de la población de objetos en movimiento. [DGG06] [DGGG06] [DG07].

Como antecedentes acerca de los problemas geométricos de complejidad NP-duros tratados con técnicas metaheurísticas, en el ámbito de la geometría computacional, que se resolvieron utilizando este tipo de métodos aproximados, podemos mencionar: Minimum Vertex Guard [CH04], Maximum Hidden Vertex Set, [BCHM08], Descomposición de Minkowski [GLHT07].

Como trabajo futuro, específicamente para problemas geométricos, tenemos encaminados los siguientes problemas de optimización de propiedades que pueden satisfacer algunas estructuras geométricas (triangulaciones, pseudotriangulaciones y poligonizaciones), tales como dilación, apuñalamiento, peso, grado, perímetro, entre otras. Por ello, consideraremos los siguientes problemas para

las clases de estructuras geométricas mencionadas: construcciones de peso bajo, de dilación pequeña, con número de apuñalamiento pequeño, de bajo grado de vértices y de perímetro o área mínima.

Por medio de la revisión, estudio y adecuación de técnicas metaheurísticas, proponemos realizar evaluaciones y validación mediante lotes de experimentación que abarquen un amplio espectro de instancias. Primeramente, se prevé la revisión de los algoritmos propuestos en la literatura para resolver de forma aproximada los problemas planteados. Se propone el estudio y análisis de adecuación algunas técnicas de computación evolutiva [BFM97]; y a posteriori, el estudio y análisis de adecuación de técnicas de basadas en el paradigma de Inteligencia Colectiva (Swarm Intelligence) [KE01], Optimización basada en Colonias de Hormigas (Ant Colony Optimization) [DS04], entre otras [MZFD04]. Para determinar la funcionalidad e impacto en la comunidad científica, los análisis incluirán el correspondiente tratamiento estadístico, y en caso de ser necesario, se realizarán comparaciones utilizando diferentes métricas para determinar la eficiencia, robustez e impacto de los resultados que se obtengan.

4. FORMACIÓN DE RECURSOS HUMANOS

Se han llevado a cabo actividades de cooperación mutua e intercambio recíproco de información científica, tecnología, desarrollo de nuevos conocimientos, creación y aplicación de nuevas tecnologías y emprendimientos en todos los campos que desarrollan sus actividades, con investigadores de otras universidades. Entre las actividades desarrolladas, se mencionan las siguientes: i) Actividades de formación académica, tales como cursos de postgrado, jornadas de capacitación, reuniones, seminarios, talleres, encuentros, entre otras. ii) Actividades de divulgación científica tales como charlas, conferencias, seminarios, entre otras, y publicaciones en congresos y revistas, contando alrededor de cincuenta trabajos publicados con referato en el ámbito nacional

e internacional (Chile, Perú, España, Colombia). iii) Formación de Recursos Humanos: trabajos finales de Licenciatura en Ciencias de la Computación, tesis de Maestría en Ciencias de la Computación y tesis doctorales en desarrollo; iv) Direcciones de becas de investigación (CyT-FCFMyN-UNSL y CONICET); v) Pasantías de investigación y docencia.

Las actividades se subvencionaron con fondos provenientes de i) Programa de Cooperación Interuniversitaria de la Agencia Española de Cooperación Iberoamericana (AECI); ii) Fondo para Mejoramiento de la Calidad Universitaria (FOMECE); iii) Proyecto AL2002-1010-2.43 / AL2003-1010-2.55 / AL2004-1010-2.53 / AL2005-PF-004 / AL2006-PF-013 / AL07-PAC-027 Geometría Computacional; iv) Proyecto AL08-PAC-16 Geometría Computacional, Algoritmos Aproximados y Bases de Datos subvencionado por la Universidad Politécnica de Madrid; v) Proyecto Fondo para Mejoramiento de la calidad Institucional (FOMEI) de la Universidad Nacional de San Luis; vi) Proyecto Tecnologías avanzadas de Bases de Datos (22/F614) de la Universidad Nacional de San Luis; vii) Aportes del Departamento de Informática Facultad de Ciencias Físico, Matemáticas y Naturales de la Universidad Nacional de San Luis; viii) Subsidios de la Universidad Nacional de San Luis para traslados y pasantías de docentes, reglamentados por Ordenanzas N° 01/90-CS y N° 18/00-CS.

Conjuntamente, nos proponemos continuar con las actividades integradoras y relacionadas al presente proyecto, proponiendo las siguientes actividades de formación académica, investigación, desarrollo y experimentación, diseño y desarrollo de índices espacio-temporales aplicables a diversos escenarios, y toda otras actividades académico-científicas vinculantes.

5. BIBLIOGRAFÍA

- [BCHM08] Bajuelos A., Canales S., Hernández, G., Martins A. *Estimating the Maximum Hidden Vertex Set in Polygons*. CG&A, Perugia. 2008.
- [BFM97] Bäck T., Fogel D. y Michalewicz Z. *Handbook of Evolutionary Computation*. IOP Publishing Ltd and Oxford University Press. 1997.
- [CH04] Canales S., Hernández Peñalver, G. *Métodos Heurísticos en Problemas Geométricos: Visibilidad, Iluminación y Vigilancia*, Universidad Politécnica de Madrid. 2004.
- [CNN93] Carlsson, S. Nilsson B.J. Ntafos, S. *Optimum guard covers and m-Watchmen Routes for restricted Polygons*, International Journal of computational Geometry and Applications, 3(1) 85-105, 1993.
- [CS89] Cole R. y Sharir M.. *Visibility problems for polyhedral terrains*. Journal of Symbolic Computation, 7, pp. 11–30. 1989.
- [DG07] Dorzán M. G., Gagliardi E. O. *Índice espacio-temporal D*R-Tree: estudio experimental de su desempeño*. XIX Encuentro Chileno de Computación, Jornadas Chilenas de Computación 2007- UNAP, Iquique, Chile. 2007.
- [DGG06] Díaz, A.J.; Gagliardi, E. O.; Gutierrez, G.; *Algoritmo de reunión espacio-temporal usando MVR-Tree*. II Congreso sobre Ingeniería e Investigación Científica, Universidad Tecnológica del Perú (UTP), Lima, Perú. 2006.
- [DGG07] Díaz, A.J.; Gagliardi, E. O.; Gutierrez, G.; *Algoritmo de Reunión Espacio-Temporal usando estructura 3DR-tree podada*. XIII Congreso Argentino de Ciencias de la Computación 2007.
- [DGGG06] Dorzán M. G., Gagliardi E. O., Gómez Barroso J. G.; Gutiérrez Retamal G. A.. *Un nuevo índice eficiente para resolver diversas consultas espacio-temporales*. CLEI 2006, Santiago de Chile, 2006.
- [DS04] Dorigo M., Stützle T. *Ant Colony Optimization*. Massachusetts Institute of Technology. 2004.
- [Fe00] Fekete, S. P. *On simple polygonalizations with optimal area*. Discrete and Computational Geometry, 23, pp. 73-110. 2000.
- [FLM08] Fekete S., Lübbecke M. y Meijer H. *Minimizing the Stabbing Number of Matchings, Trees, and Triangulations*. Proceedings of the 15th ACM-SIAM Symposium on Discrete Algorithms. 2008.
- [GGSH07] Gagliardi, E.O.; Giraudi, D. C.; Segura Guzmán, G. S.; Hernández Peñalver, *Estrategias de ruteo para redes móviles vinculadas a índices espacio-temporales*. XIII Congreso Argentino de Ciencias de la Computación. 2007.
- [GL07] Gudmundsson, J., Levcopoulos, C.; *Minimum weight pseudo-triangulations*. Computational Geometry. Theory and applications. Elsevier Vol. 38-pages 139-153, 2007.
- [GLHT07] Gagliardi E., Leguizamón M., Hernández Peñalver, G. Taranilla, M.T. *Algoritmo genético para la Descomposición de Minkowski de polígonos convexos*, XII Encuentro de Geometría Computacional, Valladolid, España. 2007.
- [LHL93] Liaw, B.C. ,Huang N.F., Lee, R.C.T. *The minimum cooperative guards problem on k-spiral polygons*, in Proc. of CCCG'93, pp. 97–101. 1993.
- [M04] Mulzer V. *Minimum Dilation Triangulation for the Regular n-gon*. 2004
- [MR06] Mulzer W., Rote G. *Minimum weight triangulation is NP-hard*. In Proceedings of the 22nd Annual ACM Symposium on Computational Geometry. 2006.
- [MF04] Michalewicz Z., Fogel D., *How to Solve It: Modern Heuristics*, 2nd Edition, Springer, 2004.
- [PV96] Pocchiola M.; Vegter G.. *Pseudo-triangulations: theory and applications*. Proceedings of the 12th Annual ACM Symposium on Computational Geometry: 291–300. 1996.
- [RSS06] Rote, Günter; Santos, Francisco; Streinu, Ileana. *Pseudo-triangulations - a survey*. 2006.
- [Zy06] Zylinski P., *Watched guards in art galleries*, Journal of Geometry 84.164–185. 2006.

Análisis e Indexación de Datos no Convencionales^{*}

Cristian Bustos, Susana Esquivel, Verónica Ludueña, Nora Reyes, Patricia Roggero
Departamento de Informática, Universidad Nacional de San Luis.
{cjbustos,esquivel,vlud,nreyes,proggero}@unsl.edu.ar

Edgar Chávez
Escuela de Ciencias Físico–Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo
elchavez@umich.mx

Gonzalo Navarro
Departamento de Ciencias de la Computación, Universidad de Chile.
gnavarro@dcc.uchile.cl

1. Introducción y Motivación

La constante aparición de datos en forma digital de diferentes tipos, tamaños y en gran cantidad, concuerda con un crecimiento de las capacidades de almacenamiento a precios más moderados. Por otro lado, dos fenómenos se han manifestado últimamente: (a) mientras la velocidad de procesamiento de la CPU se ha duplicado casi anualmente, la de los almacenamientos masivos ha progresado poco; (b) han aparecido memorias caché con mayor capacidad, más rápidas y más pequeñas, aunque más costosas, que las memorias RAM. Estos fenómenos han cambiado los modelos de costos utilizados para diseñar algoritmos y estructuras de datos eficientes. Por tal motivo los costos que se pagaban al almacenar datos en forma comprimida, en términos de velocidad de procesamiento por la descompresión, hoy en día se tornan despreciables debido a que la diferencia entre los tiempos de CPU y acceso a disco es tan significativa que el esfuerzo de descompresión se paga a cambio de una pequeña disminución en el tiempo de I/O. Además, la transferencia de los datos sobre una red local cuesta casi lo mismo que la transferencia a disco, por lo cual ésta se ve favorecida con la compresión. Este escenario ha originado líneas de investigación que tienen en cuenta estas diferencias de costos de operaciones, así nos dedicamos a las *estructuras de datos*: compactas y/o con I/O eficiente.

Nuestro objetivo es contribuir a estas líneas de investigación, diseñando estructuras de datos más eficientes

para memorias jerárquicas, haciendo uso de la compactidad o la I/O eficiente. Particularmente nos centraremos en las estructuras de datos capaces de manipular los siguientes tipos de datos: secuencias, textos, grafos, y espacios métricos, entre otros, y en estudiar los problemas desde ambos puntos de vista teórico y empírico. Además de diseñar estructuras de datos, planeamos investigar otros aspectos tales como la construcción eficiente (en espacio o en I/O u otras medidas de eficiencia), el dinamismo (es decir actualizaciones eficientes) y operaciones de búsqueda complejas (más allá de las básicas soportadas por las estructuras de datos clásicas).

2. Métodos de Acceso Métricos

Aunque existen numerosas estructuras para búsquedas por similitud en espacios métricos, sólo unas pocas trabajan eficientemente en espacios de alta o mediana dimensión, y la mayoría no admiten dinamismo, ni están diseñadas para trabajar sobre grandes volúmenes de datos; es decir, en memoria secundaria. Por lo tanto, estudiamos distintas maneras de optimizar algunas de las estructuras que han mostrado buen desempeño, con el fin de optimizarlas teniendo en cuenta la jerarquía de memorias.

2.1. SATD

Hemos desarrollado una estructura para búsqueda por similitud en espacios métricos llamado *árbol de Aproximación Espacial Dinámico (SATD)* [14] que permite realizar inserciones y eliminaciones, manteniendo un buen desempeño en las búsquedas. Muy pocos índices para espacios métricos son completa-

^{*}Enmarcado dentro de la línea Bases de Datos no Convencionales del Proyecto “Tecnologías Avanzadas de Bases de Datos” de la Universidad Nacional de San Luis.

mente dinámicos. Esta estructura se basa en el *Árbol de Aproximación Espacial* [13], el cual había mostrado un muy buen desempeño en espacios de mediana a alta dimensión, pero era estático.

El *SAT* está definido recursivamente; la propiedad que cumple la raíz a (y a su vez cada uno de los nodos) es que los hijos están más cerca de la raíz que de cualquier otro punto de S . La construcción del árbol se hace también de manera recursiva. De la definición se observa que se necesitan de antemano todos los elementos para la construcción y que queda completamente determinado al elegirle una raíz, lo cual en el *SAT* original se realizaba al azar.

Actualización de *SATD*

Para el desarrollo del *SATD* [14] se han estudiado distintas maneras de realizar incorporaciones de nuevos elementos sobre el árbol, pero se optó por un método que inserta un nuevo elemento en un punto determinado del árbol, manteniendo la aridad acotada. Gracias a esos trabajos, quedó demostrado que existen otros posibles puntos de inserción válidos, aunque no se analizó cuál de ellos sería el mejor.

Por lo tanto, tiene sentido considerar si los otros puntos posibles de inserción para un elemento consiguen mejorar aún más los costos de búsqueda. Así, como resultado se espera brindar una estructura que mejore el comportamiento durante las búsquedas gracias a elegir adecuadamente los puntos de inserción de los nuevos elementos.

El tema está siendo desarrollado como trabajo final de la Lic. en Ciencias de la Computación.

SATD con Clustering

El *SATD* es una estructura que realiza la partición del espacio considerando la proximidad espacial; pero, si el árbol agrupara los elementos que se encuentran muy cercanos entre sí, lograría mejorar las búsquedas al evitar recorrerlo para alcanzarlos.

Podemos pensar entonces que construimos un *SATD*, en el que cada nodo representa un grupo de elementos muy cercanos (“clusters”) y relacionamos los clusters por su proximidad en el espacio. La idea sería que en cada nodo se mantenga el centro del cluster correspondiente, y se almacenen los k elementos más cercanos a él; cualquier elemento a mayor distancia del centro que los k elementos, pasará a formar parte de otro nodo en el árbol.

Esperamos así obtener una estructura más adecuada a espacios en los que se sabe de antemano que pueden existir “clusters” de elementos, aprovechándonos de la existencia de los mismos para mejorar las búsquedas. Además, esta nueva estruc-

tura podría ser eficiente en memoria secundaria, lo cual la haría más adecuada a aplicaciones que traten con grandes volúmenes de datos.

Otro aspecto a analizar es cuán bueno es el agrupamiento o “clustering” que lograría esta estructura, lo cual podría estudiarse haciendo uso de nuevas estrategias de optimización de funciones a través de heurísticas bioinspiradas, las cuales han mostrado ser útiles en detección de clusters.

Estos temas han dado lugar a dos trabajos finales de la Lic. en Ciencias de la Computación, que se están desarrollando actualmente.

2.2. Join Métricos

El modelo de espacios métricos permite cubrir muchos problemas de búsqueda por similitud o proximidad, aunque en general se deja fuera de consideración al ensamble o “join” por similitud, otra primitiva extremadamente importante [7]. De hecho, a pesar de la atención que esta primitiva ha recibido en las bases de datos tradicionales y aún en las multidimensionales, no han habido grandes avances para espacios métricos generales.

Nos hemos planteado resolver algunas variantes del problema de join por similitud: (1) *join por rango*: dadas dos bases de datos de un espacio métrico y un radio r , encontrar todos los pares de objetos (uno desde cada base de datos) a distancia a lo sumo de r , (2) *k-pares más cercanos*: encontrar los k pares de objetos más cercanos entre sí (uno desde cada base de datos). Para resolver estas operaciones de manera eficiente hemos diseñado un nuevo índice métrico, llamado *Lista de Clusters Gemelos (LTC)* por su sigla en inglés) [15], el cual construye el índice de ambas bases de datos conjuntamente, en lugar de lo que habitualmente se haría de indexar una o ambas bases de datos independientemente. Esta nueva estructura permite además resolver las consultas por similitud clásicas en espacios métricos sobre cada una de las bases de datos independientemente.

A pesar de que esta estructura ha mostrado ser competitiva y obtener buen desempeño en relación a las alternativas más comunes para resolver las operaciones de join, aún queda mucho por mejorar para que se vuelva una estructura práctica y mucho más eficiente para trabajar con grandes bases de datos métricas. De esta manera sería posible pensar en extender apropiadamente el álgebra relacional y diseñar soluciones eficientes para nuevas operaciones, teniendo en cuenta aspectos no sólo de memoria secundaria, sino también de concurrencia, confiabili-

dad, etc. Algunos de estos problemas ya poseen solución en las bases de datos espaciales, pero no en el ámbito de los espacios métricos.

Este tema forma parte de la tesis doctoral de uno de los investigadores de la línea.

3. Dimensión Intrínseca

En los espacios de vectores, la “maldición de la dimensionalidad” describe el fenómeno por el cual el desempeño de todos los algoritmos existentes se deteriora exponencialmente con la dimensión. En espacios métricos generales la complejidad se mide como el número de cálculos de distancias realizados, pero la ausencia de coordenadas no permite analizar la complejidad en términos de la dimensión.

En los espacios vectoriales existe una clara relación entre la dimensión (*intrínseca*) del espacio y la dificultad de buscar. Se habla de “intrínseca”, como opuesta a “representacional”. Los algoritmos más ingeniosos se comportan más de acuerdo a la dimensión intrínseca que a la representacional. Hay varios intentos de medir la dimensión intrínseca en espacios de vectores, como la transformada de *Karhunen-Loève (KL)* y otras medidas tales como *Fastmap* [8]. Otro intento de medir la dimensión de espacios de vectores no uniformemente distribuidos, es la *dimensión fractal* [2].

Existen sólo unas pocas propuestas diferentes sobre cómo estimar la dimensión intrínseca de un espacio métrico tales como el *exponente de la distancia* (basada en una ley de potencias empírica observada en muchos conjuntos de datos) [10], y la medida de dimensión intrínseca como una medida cuantitativa basada en el histograma de distancias [4]. Además, también parece posible adaptar algunos de los estimadores de distancia para espacios de vectores para aplicarlos a espacios métricos generales, como por ejemplo *Fastmap* y *dimensión fractal*.

Muchos autores [1, 3, 6, 4] han propuesto usar histogramas de distancia para caracterizar la dificultad de las búsquedas en espacios métricos arbitrarios. Existe al menos una medida cuantitativa [4], pero ella no refleja fielmente la facilidad o dificultad de buscar en un espacio métrico dado. Así, la idea es buscar una medida posible basada en Half-Space Proximal (HSP) [5].

El test HSP es un nuevo test local para extraer un subgrafo no dirigido de un grafo de disco unitario (GDU)¹. Los vecinos HSP de cada vértice son úni-

¹Un grafo de disco unitario es un grafo bipartito en el cual

cos, dado un GDU fijo. El test HSP es un algoritmo computacionalmente simple y distribuido, que se aplica independientemente a cada vértice de un GDU. Por lo tanto, adaptamos el test HSP para aplicarlo a espacios métricos con el fin de obtener con él una medida de dimensionalidad intrínseca.

En aplicaciones reales de búsqueda en espacios métricos, sería muy importante contar con un buen estimador de la dimensión intrínseca porque nos permitiría decidir el índice adecuado a utilizar en función de la dimensión del espacio. Además, tener una buena estimación de la dimensión nos permitiría, en algunas ocasiones, elegir la función de distancia de manera tal que se obtenga una menor dimensión.

Este tema ha dado lugar a un trabajo final de la Lic. en Ciencias de la Computación, que está en desarrollo y además está siendo analizado desde un punto de vista más amplio por uno de los investigadores del grupo a fin de proponer nuevas medidas.

4. Estructuras de Datos Compactas

Las estructuras de datos *compactas*, son variantes de las estructuras clásicas sólo que funcionan en espacio reducido y se pueden dividir en dos grupos: *sucintas* y *comprimidas*. Una estructura de datos se dice comprimida cuando toma espacio proporcional al de la secuencia comprimida (existe cierta libertad para elegir un método razonable de compresión). En cambio, una estructura de datos es sucinta si necesita espacio asintóticamente despreciable sobre los datos en bruto. Una medida popular de compresibilidad de secuencias es la entropía de orden k -ésimo (H_k), según lo definido por Manzini [12], que mide la complejidad del espacio para cada secuencia individual, sin ninguna suposición sobre la entrada. La medida H_k es un límite inferior para el número de bits de salida cuando se comprime S con cualquier compresor que codifique cada símbolo de la entrada que dependa solamente de los k símbolos precedentes. Esto abarca los estándares populares tales como *PPM*, *la familia de Lempel-Ziv* y compresores basados en *Burrows-Wheeler*.

Sea s el alfabeto de una secuencia de símbolos S , $n(i)$ el número de ocurrencias del i -ésimo símbolo, y n la longitud de S . Entonces $H_0(S) = \sum(n(i) \log(n/n(i)))$. Sea w una secuencia de longitud k y $w(S)$ la subsecuencia de símbolos de S que

los vértices son objetos del universo y dos vértices estarán conectados por un arco si la distancia entre ellos es menor que una unidad dada.

siguen a w , entonces $H_k(S) = 1/n \sum (|w| H_0(w))$ sobre todos los posibles w . Esta medida es popular también en las estructuras de datos comprimidas, como aquellos índices comprimidos que codifican un texto T de n símbolos sobre un alfabeto s usando espacio de $H_k(T) + o(n \log s)$ bits y además pueden buscar eficientemente patrones en el texto. Observar que la codificación llana del texto toma $n \log s$ bits. La idea de las estructuras de datos compactas se diferencia de la compresión pura en su capacidad de manipular los datos en forma comprimida, sin tener que descomprimirlos primero. En la actualidad las estructuras de datos compactas pueden manipular secuencias de bits o de símbolos generales, árboles en general, grafos, colecciones de texto, permutaciones y mapping, sumas parciales, búsqueda por rango en una y más dimensiones, etc.

4.1. Búsqueda en Texto con Estructuras Comprimidas

El texto puede convertirse en un medio en el cual se necesite realizar búsquedas por similitud. El problema de la *búsqueda aproximada de un patrón en el texto* puede verse como: dado un *texto* $T = T[1, n]$ y un *patrón* $P = P[1, m]$ en el alfabeto Σ (con $m \ll n$) y un entero k , se desea encontrar y devolver todos los *substrings* en el texto que sean una ocurrencia aproximada de P , con a lo más k diferencias. La diferencia entre dos strings α y β se determina con la *distancia de edición* d ; $d(\alpha, \beta)$ es el mínimo número de inserciones, eliminaciones y/o sustituciones de caracteres que se deben realizar para convertir β en α . Este tipo de búsqueda tiene aplicaciones tales como la recuperación de errores (en reconocimiento óptico de caracteres, spelling), biología computacional, comunicaciones de datos, data mining, bases de datos textuales, entre otras.

La solución básica al problema de búsqueda aproximada de patrones se basa en la *programación dinámica* [11, 16]. Se puede mejorar introduciendo algún esquema de *filtración* para extraer las áreas del texto con potenciales coincidencias para luego inspeccionarlas por medio de programación dinámica, en una segunda fase del algoritmo.

Luego, el propósito es encontrar en el texto q -gramas que coincidan con los del patrón; esto se puede realizar de dos maneras: *dinámica* que escanea el texto en tiempo lineal, y *estática* que utiliza un *índice de q -gramas*. La propuesta estática es superior cuando el índice puede usarse varias veces.

Implementación de índices de q -gramas. Un índice

de q -gramas es una estructura de datos que permite encontrar rápidamente en el texto todas las ocurrencias de un q -grama dado. Existen distintas implementaciones de índices para los métodos estáticos. La básica es un arreglo de punteros de tamaño $|\Sigma|^q$ (una entrada para cada posible subcadena). Cada posición del arreglo referencia a la lista de ocurrencias en el texto del q -grama correspondiente. El índice puede resultar inútilmente grande [9]. Una mejora a este esquema de indexación es reducir el tamaño del arreglo de punteros por medio de un hashing eficiente, sin una demora significativa en las búsquedas. Otro enfoque usa una estructura de trie construido con los distintos q -gramas que aparecen en el texto. Cada hoja del trie contiene un puntero a la lista de ocurrencias del correspondiente q -grama en el texto. El trie puede construirse con un algoritmo similar al de la creación del árbol de sufijos [17].

Las implementaciones anteriores encuentran todas las ocurrencias de un q -grama dado en tiempo óptimo en función del número de ocurrencias encontradas. Pero todas sufren el mismo inconveniente: el tamaño del índice se vuelve impráctico al crecer la longitud del texto.

Un índice que usa compresión. En este índice *Lempel-Ziv* (LZ) para q -gramas las listas de ocurrencias se reemplazan con una estructura de datos más compacta. La estructura (hashing o trie) para los distintos q -gramas, ahora llamada *índice primario*, es aún necesaria para proveer un punto de comienzo para las búsquedas.

La representación compacta de las listas de ocurrencias toma ventaja de las *repeticiones* en el texto. La primera ocurrencia del string será llamada *definición* y las siguientes se llamarán *frases*. Luego, cada ocurrencia de un q -grama es o el primero de su clase o parte de alguna frase. La primera ocurrencia se almacenará en el índice primario y las demás se encontrarán usando un parsing de Lempel-Ziv [18] que usa la información sobre repeticiones. El índice LZ encontrará todas las ocurrencias de un q -grama en igual tiempo que los índices tradicionales.

Esta temática se investiga en el marco de la tesis de maestría de un investigador de la línea.

5. Conclusiones y Trabajos Futuros

Nuestras investigaciones además se encuadran en el marco de dos proyectos dentro del Programa de Promoción de la Universidad Argentina para el Fortalecimiento de Redes Interuniversitarias III en los que participa nuestra universidad junto con las uni-

versidades de: Chile, da Coruña (España), Michoacana (México) y Zaragoza (España).

Como trabajo futuro de esta línea de investigación se consideran varios aspectos relacionados al diseño de estructuras de datos que, concientes de que existe una jerarquía de memorias y de las características particulares de los datos a ser indexados, saquen el mejor partido haciendólas eficientes tanto en espacio como en tiempo.

Se trabajará en particular con estructuras de datos compactas para textos, implementando un índice LZ para q -gramas y estudiando su comportamiento en comparación con los índices tradicionales.

En el caso de espacios métricos, se intentará que las estructuras se adapten mejor al espacio métrico particular considerado, gracias a la determinación de su dimensión intrínseca, y también al nivel de la jerarquía de memorias en que se deba almacenar. Es importante destacar que estos estudios sobre espacios métricos y sobre algunas estructuras de datos particulares (como el *SATD*) permitirán no sólo mejorar el desempeño de las mismas sino también aplicar, eventualmente, muchos de los resultados que se obtengan a otras estructuras para espacios métricos.

Referencias

- [1] S. Brin. Near neighbor search in large metric spaces. In *In Proc. 21st Conference on Very Large Databases*, page 574.
- [2] Francesco Camastra. Data dimensionality estimation methods: a survey. *Pattern Recognition*, 36(12):2945–2954, 2003.
- [3] E. Chávez and J. Marroquín. Proximity queries in metric spaces. In *In R. Baeza-Yates, editor. American Workshop on String Processing*, pages 21–36. Carleton University Press, 1997.
- [4] E. Chávez and G. Navarro. Towards measuring the searching complexity of metric spaces. In *Proc. International Mexican Conference in Computer Science*, volume II, pages 969–978, 2001. Sociedad Mexicana de Ciencias de la Computación.
- [5] E. Chávez, S. Dobrev, E. Kranakis, J. Opatrny, L. Stacho, H. Tejada, and J. Urrutia. Half-space proximal: A new local test for extracting a bounded dilation spanner of a unit disk graph. In *OPODIS*, pages 235–245, 2005.
- [6] P. Ciaccia, M. Patella, and P. Zezula. A cost model for similarity queries in metric spaces. In *Proc. 17th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 1998.
- [7] V. Dohnal, C. Gennaro, P. Savino, and P. Zezula. Similarity join in metric spaces. In *Proc. 25th European Conf. on IR Research*, LNCS 2633, pages 452–467, 2003.
- [8] C. Faloutsos and K. Lin. Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In M. J. Carey and D. A. Schneider, editors, *Proc. of the 1995 ACM SIGMOD International Conference on Management of Data*, pages 163–174. ACM Press, 1995.
- [9] Petteri Jokinen and Esko Ukkonen. Two algorithms for approximate string matching in static texts. In *MFCS*, pages 240–248, 1991.
- [10] C. Traina Jr., A. M. Traina, and C. Faloutsos. Distance exponent: A new concept for selectivity estimation in metric trees. In *ICDE*, page 195, 2000.
- [11] G. Landau and U. Vishkin. Fast string matching with k differences. *J. Comput. Syst. Sci.*, 37(1):63–78, 1988.
- [12] G. Manzini. An analysis of the burrows-wheeler transform. *J. ACM*, 48(3):407–430, 2001.
- [13] G. Navarro. Searching in metric spaces by spatial approximation. *The Very Large Databases Journal*, 11(1):28–46, 2002.
- [14] G. Navarro and N. Reyes. Fully dynamic spatial approximation trees. In *Proceedings of the 9th International Symposium on String Processing and Information Retrieval*, LNCS 2476, pages 254–270. Springer, 2002.
- [15] R. Paredes and N. Reyes. List of twin clusters: a data structure for similarity joins in metric spaces. In *Proc. of the 1st International Workshop on Similarity Search and Applications*, pages 131–138. IEEE, 2008.
- [16] E. Ukkonen. Finding approximate patterns in strings. *J. Algorithms*, 6(1):132–137, 1985.
- [17] E. Ukkonen. On-line construction of suffix trees. *Algorithmica*, 14(3):249–260, 1995.
- [18] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, 1977.

Índices para Bases de Datos de Texto

Gonzalo Navarro

Departamento de Ciencias de la Computación
Universidad de Chile, Chile
gnavarro@dcc.uchile.cl

Nieves Rodríguez Brisaboa

Facultad de Informática
Universidad de A Coruña, España
brisaboa@udc.es

Norma Herrera, Carina Ruano, Darío Ruano, Ana Villegas

Departamento de Informática
Universidad Nacional de San Luis, Argentina
{nherrera, cmruano, dmruano, anaville}@unsl.edu.ar

Contexto

El presente trabajo se desarrolla en el ámbito de la línea Técnicas de Indexación para Datos no Estructurados del Proyecto Tecnologías Avanzadas de Bases de Datos, cuyo objetivo principal es realizar investigación básica en problemas relacionados al manejo y recuperación eficiente de información no tradicional, diseñando nuevos algoritmos de indexación que permitan realizar búsquedas eficientes sobre datos no estructurados.

Resumen

Mientras que en bases de datos tradicionales los índices ocupan menos espacio que el conjunto de datos indexados, en bases de datos de texto el índice generalmente ocupa más espacio que el texto pudiendo necesitar de 4 a 20 veces el tamaño del mismo. Una alternativa para reducir el espacio ocupado por el índice es buscar una representación compacta del mismo, manteniendo las facilidades de navegación sobre la estructura. Pero en grandes colecciones de texto, el índice aún comprimido suele ser demasiado grande como para residir en memoria principal. En estos casos, la cantidad de accesos a memoria secundaria realizados durante el proceso de búsqueda es un factor crítico en la performance del índice. En este trabajo estamos interesados en el diseño de índices comprimidos y en memoria secundaria para búsquedas en texto, un tema de creciente interés en la comunidad de bases de datos.

Palabras claves: *Bases de Datos de Texto, Índices, Compresión, Memoria Secundaria.*

1. INTRODUCCIÓN

Una base de datos de texto es un sistema que mantiene una colección grande de texto y que provee acceso rápido y seguro al mismo. Las tecnologías tradicionales de bases de datos no son adecuadas para manejar este tipo de bases de datos dado que no es posible organizar una colección de texto en registros y campos. Además, las búsquedas exactas no son de interés en este contexto. Sin pérdida de generalidad, asumiremos que la base de datos de texto es un único texto posiblemente almacenado en varios archivos.

Las búsquedas en una base de texto pueden ser búsquedas sintácticas, en las que el usuario especifica la secuencia de caracteres a buscar en el texto, o pueden ser búsquedas semánticas en la que el usuario especifica la información que desea recuperar y el sistema retorna todos los documentos que son relevantes. En este trabajo estamos interesados en búsquedas sintácticas.

Una de las búsquedas sintácticas más sencilla en bases de datos de texto es la *búsqueda de un patrón*: el usuario ingresa un string P (*patrón de búsqueda*), y el sistema retorna todas las posiciones del texto donde P ocurre. Para resolver

<p>T: a b c c a b c a \$</p> <p>\$=00 a=01 b=10 c=11</p>	<table border="0"> <thead> <tr> <th>Pos.</th> <th>Sufijo</th> </tr> </thead> <tbody> <tr><td>9</td><td>\$</td></tr> <tr><td>8</td><td>a\$</td></tr> <tr><td>5</td><td>a b c a \$</td></tr> <tr><td>1</td><td>a b c c a b c a \$</td></tr> <tr><td>6</td><td>b c a \$</td></tr> <tr><td>2</td><td>b c c a b c a \$</td></tr> <tr><td>7</td><td>c a \$</td></tr> <tr><td>4</td><td>c a b c a \$</td></tr> <tr><td>3</td><td>c c a b c a \$</td></tr> </tbody> </table>	Pos.	Sufijo	9	\$	8	a\$	5	a b c a \$	1	a b c c a b c a \$	6	b c a \$	2	b c c a b c a \$	7	c a \$	4	c a b c a \$	3	c c a b c a \$
Pos.	Sufijo																				
9	\$																				
8	a\$																				
5	a b c a \$																				
1	a b c c a b c a \$																				
6	b c a \$																				
2	b c c a b c a \$																				
7	c a \$																				
4	c a b c a \$																				
3	c c a b c a \$																				

Figura 1: Un ejemplo de un texto y sus correspondientes sufijos ordenados lexicográficamente.

este tipo de búsqueda podemos o trabajar directamente sobre el texto sin preprocesarlo o preprocesar el texto para construir un índice que será usado posteriormente para acelerar el proceso de búsqueda. En el primer enfoque encontramos algoritmos como Knuth-Morris-Pratt [12] y Boyer-Moore [2], que básicamente consisten en construir un autómata en base al patrón P que guiará el procesamiento secuencial del texto; estas técnicas son adecuadas cuando el texto ocupa varios megabytes. Si el texto es demasiado grande se hará necesario la construcción de un índice.

Construir un índice tiene sentido cuando el texto es grande, cuando las búsquedas son más frecuentes que las modificaciones (de manera tal que los costos de construcción se vean amortizados) y cuando hay suficiente espacio como para contener el índice. Un índice debe dar soporte a dos operaciones básicas:

Count : consiste en contar el número de ocurrencias de un patrón P en un texto T .

Locate : consiste en ubicar todas las posiciones del texto T donde el patrón de búsqueda P ocurre.

Dado un texto $T = t_1, \dots, t_n$ sobre un alfabeto Σ de tamaño σ , donde $t_n = \$ \notin \Sigma$ es un símbolo menor en orden lexicográfico que cualquier otro símbolo de Σ , denotaremos con $T_{i,j}$ a la secuencia t_i, \dots, t_j , con $1 \leq i \leq j \leq n$. Un sufijo de T es cualquier string de la forma $T_{i,n} = t_i, \dots, t_n$ y un prefijo de T es cualquier string de la forma $T_{1,i} = t_1, \dots, t_i$ con $i = 1..n$.

SA =	9	8	5	1	6	2	7	4	3
	1	2	3	4	5	6	7	8	9

Figura 2: Arreglo de sufijos para el ejemplo de la figura 1.

Un patrón de búsqueda $P = p_1 \dots p_m$ es cualquier string sobre el alfabeto Σ . La figura 1 muestra un ejemplo de un texto y sus correspondientes sufijos ordenados lexicográficamente, suponiendo que la codificación de los símbolos del alfabeto es $\$ = 00$, $a = 01$, $b = 10$, $c = 11$.

Entre los índices más populares para resolver búsqueda de patrones encontramos el *arreglo de sufijos* [15] y el *árbol de sufijos* [24]. Estos índices se construyen basándose en la observación de que un patrón P ocurre en el texto si es prefijo de algún sufijo del texto.

Arreglo de sufijos: un arreglo de sufijos $A[1, n]$ es una permutación de los números $1, 2, \dots, n$ tal que $T_{A[i],n} \prec T_{A[i+1],n}$, donde \prec es la relación de orden lexicográfico. Buscar un patrón P en T equivale a buscar todos los sufijos de los cuales P es prefijo, los cuales estarán en posiciones consecutivas de A . El proceso de búsqueda consiste entonces en dos búsquedas binarias que identifiquen el segmento del arreglo A que contiene todas las posiciones de T donde P ocurre. La figura 2 muestra el arreglo de sufijos para el texto de la figura 1.

Árbol de sufijos: un árbol de sufijos es un Pat-Tree [8] construido sobre el conjunto de todos los sufijos de T codificados sobre alfabeto binario. Cada nodo interno mantiene el número de bit del patrón que corresponde usar en ese punto para direccionar la búsqueda y las hojas contienen una posición del texto que representa al sufijo que se inicia en dicha posición. La figura 3 muestra el árbol de sufijos para el texto de la figura 1.

Mientras que en bases de datos tradicionales los índices ocupan menos espacio que el conjunto de datos indexados, en bases de datos de texto el índice generalmente ocupa más espacio que el texto pudiendo necesitar de 4 a 20 veces el tamaño del mismo [8, 15]. Una alternativa para reducir el espacio ocupado por el índice es buscar una re-

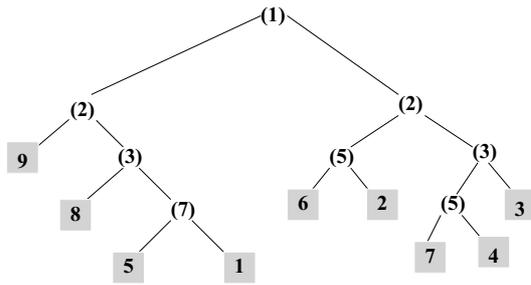


Figura 3: Árbol de sufijos para el ejemplo de la figura 1.

presentación compacta del mismo, manteniendo las facilidades de navegación sobre la estructura [6, 7, 9, 10, 13, 19, 20, 22]. Una línea de investigación reciente se enfoca en construir índices que no necesitan almacenar de manera explícita el texto que indexan. Estos índices, denominados **autoíndices**, mantienen información que permite reconstruir el texto indexado [3, 11, 21, 22].

Pero en grandes colecciones de texto, el índice aún comprimido suele ser demasiado grande como para residir en memoria principal. Como un ejemplo de este caso podemos nombrar las bases de datos conteniendo secuencias de ADN y secuencias de proteínas, que requieren la construcción de un índice de texto completo y cuyo tamaño implicará que el índice resida en memoria secundaria. En estos casos, la cantidad de accesos a memoria secundaria realizados durante el proceso de búsqueda es un factor crítico en la performance del índice [23].

Entre los índices para texto en memoria secundaria más relevantes encontramos:

String B-Tree [5]: consiste básicamente en un B-Tree en el que cada nodo es representado como un Pat-Tree [8]. Este índice requiere tanto para *count* como para *locate* $O(\frac{m+occ}{b} + \log_b n)$ accesos a memoria secundaria en el peor caso, donde *occ* es la cantidad de ocurrencias de *P* en *T* y *b* es el tamaño de páginas de disco medido en enteros. No es un índice comprimido y su versión estática requiere en espacio de 5 a 6 veces el tamaño del texto más el texto.

Compact Pat Tree [4]: representa un árbol de sufijos en memoria secundaria y en forma compacta. Si bien no existen desarrollos teóricos que

garanticen el espacio ocupado por el índice y el tiempo insumido en resolver la búsqueda, en la práctica el índice tiene un muy buen desempeño requiriendo de 2 a 3 accesos a memoria secundaria tanto para *count* como para *locate*, y ocupando entre 4 y 5 veces el tamaño del texto más el texto.

Disk-based Compressed Suffix Array [14]: adapta el autoíndice comprimido para memoria principal presentado en [22] a memoria secundaria. Requiere $n(H_0 + O(\log \log \sigma))$ bits de espacio (donde $H_k \leq \log \sigma$ es la entropía de orden *k* de *T* [16]). Para la operación *count* realiza $O(m \log_b n)$ accesos. Para la operación *locate* realiza $O(\log n)$ accesos lo cual es demasiado costoso.

Disk-based LZ-Index [1]: adapta a memoria secundaria el autoíndice comprimido para memoria principal presentado en [19]. Utiliza $8n H_k(T) + o(n \log \sigma)$ bits; los autores no proveen límites teóricos para la complejidad temporal, pero en la práctica es muy competitivo.

El estudio de índices comprimidos y en memoria secundaria para búsquedas en texto es un tema de creciente interés en la comunidad de bases de datos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo principal de esta línea de investigación es el diseño de índices comprimidos en memoria secundaria. Para lograr este objetivo hemos tomado como base dos índices: el *Compact Pat Tree* (CPT), un índice comprimido dinámico para memoria secundaria, y el *String B-Tree* (SBT) un índice dinámico para memoria secundaria.

Sobre el SBT el objetivo principal es lograr una reducción en el espacio utilizado por el mismo manteniendo los costos de búsquedas de la versión original. Para ello, se han diseñado dos variantes que consisten en modificar la representación de cada nodo del árbol B subyacente. Una de las variantes consiste en usar un Pat-Tree como originalmente proponen los autores para los nodos pero usando representación de paréntesis para el mismo [17]. La otra variante consiste en

representar cada nodo con la representación de arreglos propuesta en [18] que ofrece las mismas funcionalidades que un Pat-Tree pero que tienen las características necesarias como para permitir una posterior compresión de los mismos.

Sobre el CPT hemos diseñado e implementado los algoritmos de creación y búsqueda en un CPT en el cual se reemplaza la codificación de la forma del árbol originalmente propuesta por los autores, por la representación de paréntesis propuesta en [17]. Esta representación utiliza $2n$ bits para codificar un árbol de n nodos en lugar de los $B(n)$ bits (con $2 < B(n) < 3n$) del CPT original. Si bien esta representación es menos eficiente en búsquedas, como sólo se buscará en memoria principal sobre subárboles pequeños (limitados por el tamaño de página de disco) el desempeño del índice no se verá afectado. Los tiempos inclusive podrían mejorar dado que, reducir el espacio usado para codificar la forma del árbol, permite que cada página mantenga subárboles más grandes y la altura total del CPT (en cantidad de páginas) disminuya.

Se ha diseñado y se está implementado una segunda modificación al CPT que consiste en almacenar la hojas del árbol que son índices de sufijos en un archivo separado. El objetivo de esta modificación es lograr que las partes contengan subárboles más grandes y, en consecuencia, la altura total en cantidad de páginas sea menor.

Se está analizando además como extender el algoritmo de paginación propuesto por los autores del CPT para árboles r -arios con el fin de poder paginar otras estructuras para búsquedas en texto.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se espera que los cambios en el diseño del CPT y del SBT permitan obtener índices con las mismas funcionalidades que los originales pero reduciendo el espacio necesario para su representación. El desempeño de los índices obtenidos será medido tanto analíticamente como en forma empírica. Para esto último se cuenta con un conjunto de textos de prueba ampliamente usados y aceptados por la comunidad científica del área de estudio; los

mismos se encuentran disponibles en el sitio <http://pizzachili.dcc.uchile.cl>.

4. FORMACIÓN DE RECURSOS HUMANOS

El trabajo en curso forma parte del desarrollo de un Trabajo Final de la Licenciatura en Ciencias de la Computación, dos Tesis de Maestría en Ciencias de la Computación y una Tesis de Doctorado en Ciencias de la Computación, todos realizados en el ámbito de la Universidad Nacional de San Luis, con el asesoramiento del Dr. Gonzalo Navarro de la Universidad de Chile y de la Dra. Nieves Brisaboa de la Universidad da Coruña, España.

REFERENCIAS

- [1] D. Arroyuelo and G. Navarro. A Iempel-ziv text index on secondary storage. In *Proc. 18th Annual Symposium on Combinatorial Pattern Matching (CPM)*, LNCS 4580, pages 83–94, 2007.
- [2] R. S. Boyer and J. S. Moore. A fast string searching algorithm. *Communications of the ACM*, 20(10):762–772, 1977.
- [3] N. Brisaboa, A. Fariña, G. Navarro, A. Places, and E. Rodríguez. Self-indexing natural language. In *Proc. 15th International Symposium on String Processing and Information Retrieval (SPIRE)*, LNCS. Springer, 2008.
- [4] D. Clark and I. Munro. Efficient suffix tree on secondary storage. In *Proc. 7th ACM-SIAM Symposium on Discrete Algorithms*, pages 383–391, 1996.
- [5] P. Ferragina and R. Grossi. The string B-tree: a new data structure for string search in external memory and its applications. *Journal of the ACM*, 46(2):236–280, 1999.
- [6] P. Ferragina and G. Manzini. Indexing compressed text. *J. ACM*, 52(4):552–581, 2005.

- [7] P. Ferragina, G. Manzini, V. Mäkinen, and G. Navarro. Compressed representations of sequences and full-text indexes. *ACM Trans. Algorithms*, 3(2):20, 2007.
- [8] G. H. Gonnet, R. Baeza-Yates, and T. Snider. *New indices for text: PAT trees and PAT arrays*, pages 66–82. Prentice Hall, New Jersey, 1992.
- [9] R. González and G. Navarro. A compressed text index on secondary memory. In *Proc. 18th International Workshop on Combinatorial Algorithms (IWOCA)*, pages 80–91. College Publications, UK, 2007.
- [10] R. González and G. Navarro. Compressed text indexes with fast locate. In *Proc. 18th Annual Symposium on Combinatorial Pattern Matching (CPM)*, LNCS 4580, pages 216–227, 2007.
- [11] R. Grossi, A. Gupta, and J. Vitter. High-order entropy-compressed text indexes. In *Proc. 14th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA'03)*, pages 841–850, 2003.
- [12] D. E. Knuth, J. H. Morris, and V. R. Pratt. Fast pattern matching in strings. *SIAM Journal of Computing*, 6(2):323–350, 1977.
- [13] V. Mäkinen and G. Navarro. *Compressed Text Indexing*, pages 176–178. Springer, 2008.
- [14] V. Mäkinen, G. Navarro, and K. Sadakane. Advantages of backward searching - efficient secondary memory and distributed implementation of compressed suffix arrays. In *Proc. 15th Annual International Symposium on Algorithms and Computation (ISAAC)*, LNCS 3341, pages 681–692. Springer, 2004.
- [15] U. Manber and G. Myers. Suffix arrays: A new method for on-line string searches. *SIAM Journal of Computing*, 22(5):935–948, 1993.
- [16] G. Manzini. An analysis of the burrows—wheeler transform. *J. ACM*, 48(3):407–430, 2001.
- [17] J. Ian Munro and Venkatesh Raman. Succinct representation of balanced parentheses and static trees. *SIAM J. Comput.*, 31(3):762–776, 2001.
- [18] Joong Chae Na and Kunsoo Park. Simple implementation of string b-trees. In Alberto Apostolico and Massimo Melucci, editors, *SPIRE*, volume 3246 of *Lecture Notes in Computer Science*, pages 214–215. Springer, 2004.
- [19] G. Navarro. Indexing text using the ziv-lempel trie. *Journal of Discrete Algorithms (JDA)*, 2(1):87–114, 2004.
- [20] G. Navarro and V. Mäkinen. Compressed full-text indexes. *ACM Computing Surveys*, 39(1):2, 2007.
- [21] K. Sadakane. Compressed text databases with efficient query algorithms based on the compressed suffix array. In *ISAAC '00: Proceedings of the 11th International Conference on Algorithms and Computation*, pages 410–421, London, UK, 2000. Springer-Verlag.
- [22] K. Sadakane. New text indexing functionalities of the compressed suffix arrays. *J. Algorithms*, 48(2):294–313, 2003.
- [23] J. Vitter. External memory algorithms and data structures: Dealing with massive data. *ACM Computing Surveys*, 33(2):209–271, 2001.
- [24] P. Weiner. Linear pattern matching algorithm. In *Proc. 14th IEEE Symposium Switching Theory and Automata Theory*, pages 1–11, 1973.

Selección de Centros para Índices en Espacios Métricos

Ariel Lucero, Norma Edith Herrera, Carina Mabel Ruano

Departamento de Informática

Universidad Nacional de San Luis

San Luis, Argentina

{3033402,nherrer,a,cmruano}@unsl.edu.ar

Contexto

El presente trabajo se desarrolla en el ámbito de la línea Técnicas de Indexación para Datos no Estructurados del Proyecto Tecnologías Avanzadas de Bases de Datos, cuyo objetivo principal es realizar investigación básica en problemas relacionados al manejo y recuperación eficiente de información no tradicional, diseñando nuevos algoritmos de indexación que permitan realizar búsquedas eficientes sobre datos no estructurados.

Resumen

El concepto de búsquedas por similitud, es decir buscar elementos en una base de datos que sean similares o cercanos a uno dado, tiene aplicación en diversas áreas de computación. Las bases de datos que soportan este tipo de consultas pueden ser modelizadas mediante el concepto de espacio métrico. Un espacio métrico es un par (\mathcal{X}, d) , donde \mathcal{X} es un conjunto de objetos y d es una función de distancia definida entre ellos que mide cuán diferentes son. El procesamiento de consultas en espacios métricos es un tema de investigación emergente tanto desde el punto de vista de los algoritmos que las implementan como de los índices que las soportan. En este trabajo abordamos el estudio de algoritmos de indexación basados en particiones compactas buscando mejorar la eficiencia de los mismos.

Palabras claves: Espacios Métricos, Búsquedas por Similitud, Índices, Particiones Compactas.

1. INTRODUCCIÓN

El concepto de *búsquedas por similitud o por proximidad*, es decir buscar elementos de una base de datos que sean similares o cercanos a uno dado, aparece en diversas áreas de computación, tales como reconocimiento de voz, reconocimiento de imágenes, compresión de texto, biología computacional, inteligencia artificial, minería de datos, entre otras.

En [9] se muestra que este problema se puede expresar como sigue: dado un conjunto de objetos \mathcal{X} y una función de distancia d definida entre ellos que mide cuán diferentes son, el objetivo es recuperar todos aquellos elementos que sean similares a uno dado. Esta función d cumple con las propiedades características de una función de distancia: *positividad* ($d(x, y) \geq 0$), *simetría* ($d(x, y) = d(y, x)$) y *desigualdad triangular* ($d(x, y) \leq d(x, z) + d(z, y)$).

El par (\mathcal{X}, d) se denomina *espacio métrico*. La base de datos será un subconjunto finito $\mathcal{U} \subseteq \mathcal{X}$. En este nuevo modelo de bases de datos, una de las consultas típicas que implica recuperar objetos similares es la *búsqueda por rango*, que denotaremos con $(q, r)_d$. Dado un elemento $q \in \mathcal{X}$, al que llamaremos *query* y un radio de tolerancia r , una búsqueda por rango consiste en recuperar los objetos de la base de datos cuya distancia a q no sea mayor que r , es decir, $(q, r)_d = \{u \in \mathcal{U} : d(q, u) \leq r\}$.

El tiempo total de resolución de una búsqueda

da contiene tres términos, a saber: $T = \#evaluaciones\ de\ d \times complejidad(d) + tiempo\ extra\ de\ CPU + tiempo\ de\ I/O$. En muchas aplicaciones la evaluación de la función d es tan costosa que las demás componentes de la fórmula anterior pueden ser despreciadas. Éste es el modelo usado en este trabajo; por consiguiente, nuestra medida de complejidad será la cantidad de evaluaciones de la función de distancia d .

Una forma trivial de resolver una búsqueda por rango es examinando exhaustivamente la base de datos. Para evitar esta situación, se preprocesa la base de datos por medio de un *algoritmo de indexación* con el objetivo de construir una *estructura de datos o índice*, diseñada para ahorrar cálculos en el momento de resolver una búsqueda.

En [9] se presenta un desarrollo unificador de las soluciones existentes en la temática. En dicho trabajo, se muestra que todos los enfoques para la construcción de índices en espacios métricos consisten en particionar el espacio en clases de equivalencia e indexar las clases de equivalencia. Luego, durante la búsqueda, por medio del índice descartar algunas clases, y buscar exhaustivamente en las restantes. La diferencia entre los distintos algoritmos radica en cómo construyen esta relación de equivalencia. Básicamente se pueden distinguir dos grupos: *algoritmos basados en pivotes* y *algoritmos basados en particiones compactas*.

Algoritmos basados en pivotes: en los algoritmos basados en pivotes [1, 3, 5, 7, 8, 9], la relación de equivalencia se define tomando en cuenta la distancia de los elementos de la base a un conjunto preseleccionado de elementos denominados *pivotes*; en este sentido, dos elementos son considerados equivalentes si están exactamente a la misma de distancia de todos los pivotes. El proceso de indexación consiste en seleccionar k pivotes $\{p_1, p_2, \dots, p_k\}$, y asignar a cada elemento a el vector o firma $\delta(a) = (d(a, p_1), d(a, p_2), \dots, d(a, p_k))$. Ante una búsqueda $(q, r)_d$, se usa la desigualdad triangular junto con los pivotes para filtrar elementos de la base de datos sin medir su distancia a la query q . Para ello se computa la distancia de q a cada uno de los pivotes p_i , y luego se descartan todos aquellos elementos a , tales que para algún pivote p_i se cumple

que $|d(q, p_i) - d(a, p_i)| > r$. Los elementos no descartados pasan a formar parte de un conjunto de elementos que se comparan directamente con q para determinar si forman o no parte de la respuesta.

Algoritmos basados en particiones compactas:

En el caso de los algoritmos basados en particiones compactas [4, 13, 11, 12], la relación de equivalencia se define teniendo en cuenta la cercanía de los elementos a un conjunto preseleccionado de elementos denominados *centros*; en este caso dos elementos son equivalentes si tienen al mismo centro c como su centro más cercano. El objetivo final es dividir el espacio en zonas tan compactas como sea posible. Para ello seleccionan un conjunto de *centros* $\{c_1, c_2, \dots, c_k\}$ y dividen el espacio asociando a cada centro. La partición asociada a un centro c_i está formada por el conjunto de puntos que tienen a c_i como su centro más cercano. Existen muchos criterios posibles para descartar zonas o particiones durante una búsqueda. Los dos más populares son:

- a. **Criterio del hiperplano:** es el más básico y el que mejor expresa la idea de partición compacta. Básicamente, si c es el centro de la clase $[q]$ (es decir, el centro más cercano a q) entonces la bola con centro q no interseca $[c_i]$ si $d(q, c) + r < d(q, c_i) - r$. Es decir, si la bola asociada a q no interseca el hiperplano que divide su centro más cercano c y el centro c_i , entonces cae fuera de la clase de c_i .
- b. **Criterio del radio de cobertura:** en este caso se trata de limitar la clase $[c_i]$ considerando la bola centrada en c_i que contiene todos los elementos de \mathcal{U} que caen en la clase. Definimos el radio de cobertura de c en el espacio \mathcal{U} como $cr(c) = \max_{u \in [c] \cap \mathcal{U}} d(c, u)$. Luego, podemos descartar $[c_i]$ si $d(q, c_i) - r > cr(c_i)$.

Uno de los principales obstáculos en el diseño de buenas técnicas de indexación es lo que se conoce con el nombre de *maldición de la dimensionalidad*. El concepto de dimensionalidad

está relacionado a la dificultad o facilidad de buscar en un determinado espacio métrico. La dimensión intrínseca de un espacio métrico se define en [9] como $\rho = \frac{\mu^2}{2\sigma^2}$, siendo μ y σ^2 la media y la varianza respectivamente de su histograma de distancias. Es decir que, a medida que la dimensionalidad intrínseca crece, la media crece y su varianza se reduce. Esto significa que el histograma de distancia se concentra más alrededor de su media, lo que influye negativamente en los algoritmos de indexación.

La figura 1 da una idea intuitiva de por qué el problema de búsqueda se torna más difícil cuando el histograma es más concentrado. Los histogramas de la figura representan posibles distribuciones de distancias respecto de algún elemento c (histogramas locales respecto de c). Considerando una búsqueda $(q, r)_d$, las áreas sombreadas de la figura muestran los puntos que no podrán descartarse si se utiliza c como centro. Puede observarse que a medida que el histograma se concentra más alrededor de su media, disminuye la cantidad de puntos que pueden descartarse usando como dato $d(c, q)$. Este fenómeno es independiente de la naturaleza del espacio métrico, y nos brinda una forma de cuantificar cuán dura es una búsqueda sobre el mismo.

El procesamiento de consultas en espacios métricos es un tema de investigación emergente tanto desde el punto de vista de los algoritmos que las implementan como de los índices que las soportan. Por esta razón, en este trabajo abordamos el estudio de algoritmos de indexación basados en particiones compactas buscando mejorar la eficiencia de los mismos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Se sabe que la forma en que se seleccionan los centros afecta en gran medida el desempeño del índice creado. La selección trivial es la random, pero la experiencia marca que aquellas tareas realizadas aleatoriamente pueden mejorarse incorporando alguna política específica. El grupo de centros seleccionados durante la construcción del índice no afecta en absoluto la efectividad del mismo pero es crucial para su eficiencia.

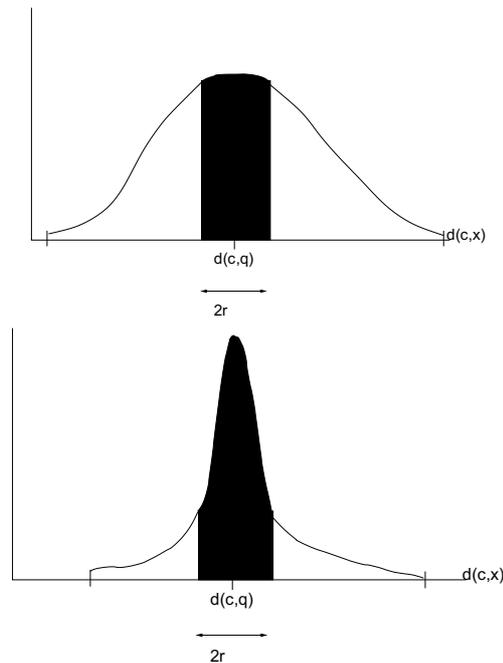


Figura 1: Histogramas de distancias de baja dimensionalidad (arriba), y de alta dimensionalidad (abajo)

Una buena selección de centros debería elegir un conjunto de elementos que permitan agilizar las búsquedas mediante el uso de la información obtenida al calcular las distancias entre los centros y la query q . Es decir, se espera que un buen conjunto de centros forme particiones en el espacio que minimicen la cantidad de evaluaciones de la función de distancia necesarias para responder una búsqueda por similitud.

Una característica de un buen conjunto de centros es que sus elementos no estén muy cercanos unos de otros o concentrados en una pequeña zona del espacio ya que, si esto ocurre, es muy probable que la bola de la query intersekte varias zonas del espacio las que no podrán ser descartadas.

En [4] se propone una técnica que intenta evitar elegir centros que estén muy cercanos unos de otros. Para ello se toma una muestra del espacio y luego se seleccionan los elementos de la muestra que están más alejados unos de otros. Para lograr esto, en cada paso, se elige como centro c_i aquel elemento que esté más alejado simultáneamente de c_1, c_2, \dots, c_{i-1} . Este proceso puede resolverse con una complejidad de $O(nm)$ donde n es el número de elementos en la muestra y m es la cantidad de centros deseados.

El proceso de descarte de las búsquedas por si-

militud en espacios métricos es sensible a la distribución de los elementos en el espacio. Es decir, en las zonas del espacio con mayor concentración de elementos, el proceso de descarte de la búsqueda se hace más dificultoso que en otras zonas de menor concentración de elementos. Como ya vimos, una forma de visualizar la distribución de los elementos de un espacio métrico es utilizando histogramas de distancia. En [2] los autores hacen uso de histogramas de distancias para definir el concepto de *núcleo duro* y *núcleo blando* de un espacio métrico. El núcleo duro está formado por aquellos elementos que se ubican en la zona de mayor concentración, que se corresponde con la zona que se encuentra alrededor de la media del histograma de distancias, si el mismo tiene forma de campana de Gauss; el núcleo blando está formado por el resto de los elementos en el espacio métrico.

En [6] se proponen dos técnicas de selección de centros basadas en los conceptos de núcleo duro y núcleo blando. Una de ellas, denominada *closer element* consiste en seleccionar los centros desde el conjunto de elementos pertenecientes al núcleo blando y la otra, denominada *high density zone*, consiste en elegirlos desde el conjunto de elementos pertenecientes al núcleo duro. Estas técnicas no utilizan los núcleos duro y blando del espacio métrico (como se sugiere en [2]) sino que en cada paso se crea el histograma local del centro elegido en el paso anterior y la selección del próximo centro se basa sólo en lo que el centro anterior vería como núcleo duro o núcleo blando (que puede no corresponderse con los núcleos reales del espacio). Los autores muestran que *high density zone* es la más competitiva logrando importantes reducciones en la cantidad de evaluaciones de distancias cuando se la compara con una selección aleatoria de centros.

En este trabajo proponemos el estudio de nuevas técnicas de selección de centros para índices métricos basados en particiones compactas. Hemos diseñado hasta el momento dos nuevas políticas de selección, las que explicamos a continuación:

- Se sabe que el histograma local puede ser muy diferente del histograma global del espacio; pero si los histogramas locales de di-

ferentes puntos de referencia son similares, entonces podemos predecir a través de ellos la distribución de los elementos del espacio métrico. Esta es la observación que usan los autores en [2] para proponer como método de detección del núcleo duro, la intersección de varios histogramas locales. Basándonos en esto, la técnica *high density zone* debería mejorar su aproximación al núcleo duro, y en consecuencia mejorar su desempeño, si en lugar de considerar sólo el histograma del último centro elegido c_i considera la intersección de los histogramas de todos los centros elegidos hasta ese momento c_1, c_2, \dots, c_i .

- La técnica *high density zone* se basa en histogramas con forma de campana de Gauss y como se muestra en [10] no todos los histogramas tienen esta forma; en algunos casos el histograma pueden tener varios máximos locales que por lo general no se corresponden con la media del mismo. Una adaptación de *high density zone* a otros tipos de histogramas es realizar la intersección de aquellas zonas que se encuentren aledañas a todos los máximos locales.

Actualmente nos encontramos implementando estas técnicas para su posterior evaluación experimental sobre índices basados en particiones compactas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se espera que las técnicas diseñadas resulten más competitivas que *high density zone* dado que estarán aproximando de manera más real el histograma global del espacio métrico y además tendrán en cuenta las distintas formas que puede tener un histograma. Los algoritmos implementados serán evaluados empíricamente utilizando los espacios de prueba ampliamente usados y aceptados por la comunidad científica del área de estudio, los que encuentran disponibles en el sitio de Similarity Search and Applications (SISAP) <http://www.sisap.org>.

4. FORMACIÓN DE RECURSOS HUMANOS

El presente trabajo se desarrolla en el ámbito de la línea *Técnicas de Indexación para Datos no Estructurados* del proyecto *Tecnologías Avanzadas de Bases de Datos* de la Universidad Nacional de San Luis. El desarrollo de este trabajo es parte de un Trabajo Final de la Licenciatura en Ciencias de la Computación de dicha Universidad.

REFERENCIAS

- [1] R. Baeza-Yates. Searching: an algorithmic tour. In A. Kent and J. Williams, editors, *Encyclopedia of Computer Science and Technology*, volume 37, pages 331–359. Marcel Dekker Inc., 1997.
- [2] R. Baeza-Yates, B. Bustos, E. Chávez, N. Herrera, and G. Navarro. *Clustering in Metric Spaces and Its Application to Information Retrieval*. Kluwer Academic Publishers, 2003. ISBN 1-4020-7682-7.
- [3] R. Baeza-Yates, W. Cunto, U. Manber, and S. Wu. Proximity matching using fixed-queries trees. In *Proc. 5th Combinatorial Pattern Matching (CPM'94)*, LNCS 807, pages 198–212, 1994.
- [4] S. Brin. Near neighbor search in large metric spaces. In *Proc. 21st Conference on Very Large Databases (VLDB'95)*, pages 574–584, 1995.
- [5] W. Burkhard and R. Keller. Some approaches to best-match file searching. *Comm. of the ACM*, 16(4):230–236, 1973.
- [6] N. Herrera C. Mendoza Alric. Center selection techniques for metric indexes. *Journal of Computer Science & Technology*, 7(1):98–104, 2007.
- [7] E. Chávez and K. Figueroa. Faster proximity searching in metric data. In *Proceedings of MICAI 2004*. LNCS 2972, Springer, Cd. de México, México, 2004.
- [8] E. Chávez, J. Marroquín, and G. Navarro. Fixed queries array: A fast and economical data structure for proximity searching. *Multimedia Tools and Applications (MTAP)*, 14(2):113–135, 2001.
- [9] E. Chávez, G. Navarro, R. Baeza-Yates, and J.L. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.
- [10] E. Chávez, N. Herrera, C. Ruano, and A. Villagas. Funciones de discretización basadas en histogramas de distancia. In *Actas de la Conferencia Latinoamericana de Informática (CLEI'06)*, Santiago, Chile, 2006.
- [11] I. Kalantari and G. McDonald. A data structure and an algorithm for the nearest point problem. *IEEE Transactions on Software Engineering*, 9(5):631–634, 1983.
- [12] G. Navarro. Searching in metric spaces by spatial approximation. In *Proc. String Processing and Information Retrieval (SPIRE'99)*, pages 141–148. IEEE CS Press, 1999.
- [13] J. Uhlmann. Satisfying general proximity/similarity queries with metric trees. *Information Processing Letters*, 40:175–179, 1991.

IDENTIFICACIÓN DE REQUERIMIENTOS DE SOFTWARE EN LAS PEQUEÑA Y MEDIANA EMPRESA (PYMES) DE LA PROVINCIA DE CÓRDOBA

Alejandra Jewsbury – ajewsbury@gmail.com

Susana Rey – srey@sistemas.frc.utn.edu.ar

Silvia Lanza Castelli – slcastell@sistemas.frc.utn.edu.ar

Germán Vélez – gvelez@sistemas.frc.utn.edu.ar

Universidad Tecnológica Nacional – Facultad Regional Córdoba – Departamento de Ingeniería en Sistemas de Información

CONTEXTO

Desde los trayectos académicos muchas veces nos encontramos con la presencia de investigaciones que aportan herramientas para el desarrollo de sistemas de información en las organizaciones pero no siempre con producciones que demuestren efectivamente el grado de inserción de estas aplicaciones en el contexto productivo real de los sectores a los cuales se dirigen y menos aún del impacto que éstas tienen en el crecimiento económico de la región toda.

Por otro lado, en lo relacionado a la formación de personal en tecnologías de información, no se sabe si está bien orientada y especializada, debido a que no se cuenta con información confiable de estudios en las organizaciones de nuestra región, como hacen uso de la tecnología de la información, cuales son las estrategias para la recolección de información en sus procesos de negocio, la interpretación de la información que necesitan los procesos como requerimientos esenciales, el estado actual de la documentación que respalda dicho software y el mantenimiento posterior de esa documentación.

Por otro lado, se ha detectado que muchas empresas hacen el uso del software sin licencias y demoran en regularizar esta situación legal a corto plazo.

La asociación civil Software Legal, que nuclea a 25 fabricantes de soluciones informáticas nacionales e internacionales y que persigue la protección de la propiedad intelectual, publicó un informe donde se consigna que cuatro de cada diez pymes utiliza programas informáticos ilegales.

En consecuencia, las pymes enfrentarían distintos riesgos que van desde pérdida de

información hasta exclusiones empresarias por operar con software pirata-

Debido a lo anterior, la hipótesis que se plantea en esta investigación es que la mayoría de las PyMES en esta región carecen de un Sistema de Información apropiado o no cubren todas las expectativas de los procesos de la organización, además de no aprovechar eficientemente la información que éste les proporciona.

Este es el caso especial de las PYMES del sector metalmecánico de Argentina. En general no se encuentran investigaciones concretas sobre el grado de desarrollo informático y por ende no existe información que permita conocer el uso, impacto, y aplicación sobre estas industrias.

Estas inquietudes son las que dan lugar al proyecto de investigación PID Promocional del Departamento de Sistemas de UTN – FRC y que se titula “Identificación de Requerimientos de Software en las Pequeña y Mediana Empresa (PYMES) metal-mecánica de la Provincia de Córdoba”.

Este proyecto expresa las motivaciones de conocer, de comprender cuales son los procesos que se siguen en estas empresas para expresar sus requerimientos iniciales de información, validar los mismos en productos finales y completar esta documentación a modo de actualizaciones continuas.

En base a las investigaciones emergentes de este proyecto se presentan los avances del proyecto en los que se conjugan los aspectos teóricos y metodológicos de la ingeniería de requerimientos con aquellos que provienen del conocimiento específico sobre los que se enmarcan los procesos de negocios del sector metalmecánica en la región central del país.

RESUMEN

El trabajo investiga los métodos que se utilizan en la actualidad en las PYMES de la Provincia de Córdoba (sector metalmeccánico) para la identificación, indagación y especificación de los Requerimientos de información en referencia a futuros desarrollos de software.

A través de la aplicación del instrumento de recolección de información, en entrevistas a niveles de usuarios no expertos, se pudo observar a modo de hipótesis de trabajo que en muchos casos la introducción de tecnología informática en las organizaciones no se ha asociado a incrementos en los niveles de productividad que luego se traducen en valor agregado a los clientes

Se plantean los aspectos propios de la industria y de sus procesos asociados y el tratamiento de información asociados a éstos.

El estudio se basa en investigaciones provenientes de la ingeniería de requerimientos y se propone aplicación concreta de éstos lineamientos, propios de la disciplina informática, a la indagación y luego elaboración de modelos conceptuales que permitan la generación de documentos de Especificación de Requerimientos de Software.

Palabras clave: Requerimientos – Software-PYMES

1. INTRODUCCION

La captura de requerimientos es una de las actividades de mayor relevancia dentro del proceso de desarrollo de software, ya que una correcta especificación de requerimientos garantiza el desarrollo de productos asociados a las expectativas del cliente, utilización adecuada del tiempo del equipo de trabajo y menores costos de reproceso y por ende mejores resultados en el proceso de producción de software. Este proceso de especificación de requerimientos implica una conceptualización y expresión de los conceptos y necesidades planteadas por los futuros usuarios en forma concreta, mediante la generación de especificaciones

(documentos, modelos) correctas que describan con claridad, sin ambigüedades, en forma consistente y completa el comportamiento deseado para el software que soporte los procesos del negocio implicados en el emprendimiento bajo análisis (Sommerville, 2002).

La industria metalmeccánica de la provincia de Córdoba se caracteriza por ser un sector de amplia cobertura geográfica, con un importante número de empresas de pequeña dimensión y con un grado de utilización de tecnología dispar. En referencia a las actividades asociadas a la producción de software nos encontramos sin áreas informáticas o de escala pequeña, y con una fuerte tercerización de sus servicios informáticos y del desarrollo de software. Es por tanto que el estudio de la forma en la que estos usuarios no expertos desarrollan las comunicaciones que dan lugar a futuros desarrollos reviste un interés especial.

Todas estas situaciones han provocado que las PyMEs en general, y en particular en el ámbito metalmeccánico, no consideren necesario o de importancia para el negocio el desarrollo, implementación y/o mantenimiento de los sistemas informáticos. Y en directa relación con este tema, se encuentra la importancia de la especificación de requerimientos de software como parte de las tareas vinculadas al proceso de desarrollo: sin una especificación de requerimientos no será posible establecer con claridad las metas a alcanzar, realizar las pruebas e inspecciones al sistema desarrollado, controlar la producción, proveer una forma eficaz para el mantenimiento del software, guiar el desarrollo hacia el sistema correcto; en definitiva no se podrá alcanzar un nivel óptimo en la construcción del sistema informático que satisfaga a los clientes.

Las pequeñas y medianas empresas (PyMEs), para salir adelante, conviviendo con las grandes corporaciones, requieren de mecanismos que permitan canalizar sus necesidades, priorizarlas, implementarlas y rápidamente retroalimentarse para seguir con su camino.

Día a día buscan participar más en los mercados y desean ser más competitivas y productivas para poder asegurar su sostenibilidad y éxito económico.

Una gran apuesta para estas organizaciones es implementar un sistema de gestión ya sea de calidad o integral ajustado a sus necesidades y enmarcado no solamente a conseguir una certificación que mejore su imagen y posibilidades comerciales, sino que asegure que la empresa puede garantizar la satisfacción de las necesidades de sus clientes y desarrollar una cultura de mejoramiento continuo que garantice su sostenibilidad futura.

Se tiene conocimiento que el liderazgo de la mayoría de las Pymes está en manos de sus fundadores y familiares directos o indirectos, y es en éste marco que se puede observar, en base a estudios realizados para el sector económico, que la delegación de funciones es precaria y que la toma de decisiones no se realiza en base a información sistémicamente organizada.

Los ciclos de vida de estos proyectos guardan estrecha relación con la naturaleza de sus productos y/o servicios que atienden, condicionando para ello una infraestructura que permita canales de atención de requerimientos claros, y un proceso de planificación de proyectos muy exigente y flexible a la vez. La cobertura del negocio obliga a agrupar los proyectos en Carteras de Requerimientos, siendo sus objetivos seguidos y controlados según sus salidas al mercado.

La globalización de los mercados y la creciente demanda de clientes de calidad, precio y eficiencia en sus productos o servicios, ha puesto a las empresas ante exigencias inapelables de optimizar sus procesos de negocio y aumentar sus recursos tecnológicos, como así también la eficiencia de sus recursos humanos para poder ser versátil a la situación actual e impacte de manera notable en su estructura de costos y pueda implementarse en plazos razonables.

La Industria Metalmeccánica constituye un eslabón fundamental en el entramado productivo de una nación. No sólo por su

contenido tecnológico y valor agregado, sino también por su articulación con distintos sectores industriales. Prácticamente todos los países con un desarrollo industrial avanzado cuentan con sectores metalmeccánicos consolidados.

Parecería que es una constante de este tipo de organizaciones se encuentren en crisis permanente. A los efectos de delimitar los aspectos que determinan la crisis constante en este sector se observan:

Determinantes Externos:

- Los vinculados con la adquisición de tecnología de alto costo y alta complejidad de manipulación, que producen serias limitaciones de producción en escala y, por lo tanto, de riesgosa amortización, así como el requerimiento de capitales no disponibles.
- La existencia o formación de mano de obra altamente calificada, difícil de preparar y retener por parte de las pequeñas y medianas empresas.
- Los vinculados con la permeabilidad de mercados, nuevos productos, nuevos sistemas de distribución, menores costos por mejor adquisición de materia primas o productos intermedios, mejor incorporación de recursos humanos.

Determinantes Internos:

- La personalidad de los dirigentes.
- La presencia, en mayor parte, de empresas de familia.

En la actualidad, existen una infinidad de herramientas de software que son utilizadas por las organizaciones con el fin de optimizar sus procesos operacionales, mejorar su inserción en el mercado, aumentar sus ganancias y disminuir sus costos.

No obstante ello el acceso a estas herramientas se ve recortada no por el costo o precio sino por el tiempo demandado en su implementación.

La ingeniería de requerimientos disminuye los costos y retrasos del proyecto; muchos

estudios han demostrado que reparar errores por un mal desarrollo no descubierto a tiempo, es sumamente caro.

La ingeniería de requerimientos es un enfoque sistémico para recolectar, organizar y documentar los requerimientos del sistema; es también el proceso que establece y mantiene acuerdos sobre los cambios de requerimientos, entre los clientes y el equipo del proyecto.

La especificación de requerimientos debe atender a mejorar la gestión del cambio en la organización, integrar visiones dentro de la misma y vincular los Sistemas de Información con la estrategia organizacional.

En el marco de la ingeniería de requerimientos, la especificación incluye la producción de documentos y modelos que capturen diversos aspectos relacionados con los requerimientos definidos por los futuros usuarios. La gestión de estos requisitos pretende ser un enfoque sistemático para encontrar, documentar y seguir la pista de los requisitos cambiantes de un sistema. Al analizar este concepto, se visualiza el hecho de que para documentar y especificar los requerimientos, el paso previo indispensable es encontrar los mismos, utilizando diversas técnicas de recolección de información. Y luego de obtener dichos requerimientos, los mismos deberán ser documentados, en el marco de algún documento de especificación.

Y en este sentido es importante destacar la situación actual de las PyMEs metalmeccánicas de la provincia de Córdoba: la mayoría no cuenta con especificaciones de requerimientos de los sistemas con los que trabajan. Y cuando nos referimos a requerimientos, no sólo se contemplan las capacidades con las cuales debe ser conforme el sistema, también se deben considerar requisitos de calidad o no funcionales, tales como fiabilidad, rendimiento, soporte y otros. ¿Cuál es la consecuencia de esta escasez o carencia de documentación? Los actuales sistemas informáticos son difícilmente mantenibles; las complejas tareas tales como modificar características actuales del software, solucionar problemas o ampliar las funcionalidades existentes resultan costosas.

Ampliando el horizonte del análisis, también se deben atender las consecuencias relativas al desarrollo de nuevos sistemas ¿De qué manera se establece la especificación de requerimientos de software en los procesos en la construcción de nuevos sistemas? En resumen no permite se predecir el tamaño, el esfuerzo y planificar adecuadamente las tareas.

Y la siguiente pregunta que se podría plantear es ¿Cuál es la importancia dentro de la empresa que se otorga a estas actividades y a los sistemas informáticos en general? ¿Se presenta la especificación de requerimientos como una necesidad con carácter prioritario en la organización? En la mayoría de los casos se observa que los sistemas fueron desarrollados con una pobre captura de las necesidades del usuario o mala interpretación de las características de lo que se necesitaba en realidad. Inmediatamente la pregunta podría ser ¿Por qué estas actividades no son consideradas importantes en la organización? ¿Acaso se desconoce la importancia y los beneficios de realizar tareas como la especificación de requerimientos? ¿O el contexto en el cual se encuentran inmersas las organizaciones no permite destinar recursos a estos procesos? Seguramente una combinación de todos estos elementos llevan a las PyMEs metalmeccánicas a no realizar estas actividades.

Otro interrogante a analizar es el siguiente: las actividades vinculadas a la ingeniería de requerimientos ¿no son realizadas en la organización o los procedimientos que se utilizan no son eficientes para el dominio en el cual se trabaja? Seguramente estos interrogantes se podrán responder en el transcurso de la investigación que se realiza.

Un estudio sobre la Promoción de la Pequeña y Mediana Empresa de la República Argentina”, organizado por la Secretaría de Industria, Comercio y PyMES, el Instituto Nacional de Tecnología Industrial y la Agencia de Cooperación Internacional de Japón (JICA); dicho proyecto se focalizó en la mejora de productividad de una selección de empresas pertenecientes a los sectores autopartista, agropartista y componentes para

maquinarias de procesamiento de alimentos a partir de implementación de “tecnologías blandas” y en la sistematización de datos e intercambios comerciales entre PYMES, aumentando la productividad y la competitividad en el mercado internacional. Según el estudio, dicho segmento de las Pymes tienen pocas oportunidades de recibir formación sobre nuevos conocimientos en este tipo de tecnologías.

2. LINEAS DE INVESTIGACION y DESARROLLO

Por su finalidad, la investigación se encuadra en una investigación básica para que permita ser fundamento para otras investigaciones y se pretende conocer, explicar y comprender los fenómenos actuantes en las empresas.

En palabras de Mario Bunge:

...” A la par que el sociólogo o el economista básico estudia sociosistemas (sistemas sociales) con el fin de comprender cómo funcionan (bien o mal), el científico social aplica los estudia con el fin de averiguar qué favorece u obstaculiza su mantenimiento o su desarrollo en algún sentido. Y lo hace con la esperanza (o el temor) de que los resultados de su estudio sean utilizados por quienes ejercen poder, para modificar dichos sistemas. El técnico, en cambio, puede investigar o utilizar los resultados de investigaciones (propias o ajenas), pero en ningún caso se queda con el conocimiento: aspira a poner el saber en acción. El centro de la actividad cognoscitiva técnica es el diseño de dispositivos o planes de acción que permitan crear o controlar cosas concretas.”¹

Si nos situamos en el momento epistémico el objeto de investigación está definido por la identificación del problema de la cual atraviesan las empresas en éste contexto actual socio-político, en la identificación e interpretación de sus necesidades actuales y una escasa visión de futuras necesidades.

Delineando como marco teórico las empresas que se encuadran como pequeñas y medianas tomamos como concepto que cada una de ellas es particular pero el hecho en si es uno sólo. Donde el contexto observable es un

continuo transcurrir de lo que existe. Identificamos ese transcurrir como lo real.

Los hechos emergen a partir del cruce entre lo que existe - efectivamente del mundo real - y los conceptos, las teorías, los métodos y los instrumentos.

-1.-Bunge M.op.cit.,p.353

3. RESULTADOS OBTENIDOS

Los objetivos de la presente línea de investigación se encuentran enmarcados en el proyecto de investigación antes citado y son la determinación de las métodos y prácticas que se utilizan en el sector de la industria para enunciar y describir requerimientos de información.

Es objetivo de esta investigación también definir los lineamientos metodológicos generales para la identificación, especificación y uso de instrumentos de documentación de requerimientos de información para la concreción de productos de software.

Las preguntas que guían ésta investigación se refieren a: ¿Las técnicas actuales de captura de requerimientos permiten detectar necesidades de información no contempladas hasta el momento en los procesos de negocio? ¿Cómo es el estado actual en la producción de documentación de requerimiento de información del software? ¿Cuáles es la relación dentro de la organización de los encargados de generar, mantener y actualizar los registros de requerimientos? ¿Será la tecnología el instigador real de la eficiencia productiva en el sector?, ó ¿Cuál es el apoyo que estos sectores tienen en el tratamiento informático de la información? ¿Cuáles son los procesos prioritarios en el ingreso a la administración informática?

En el momento actual nos encontramos con los siguientes avances:

- Se han determinado los sectores productivos y se ha diseñado el instrumento de recolección de información a aplicar.
- Se ha iniciado la indagación encontrando perfiles comunes que

permiten describir inicialmente los procesos productivos prioritarios.

- Se ha relevado la estructura informática del sector.
- Se han realizado investigaciones bibliográficas que permiten conocer el grado de avance de la ingeniería de software en el aspecto especial de la elicitación de requerimientos.

Es importante destacar que desde las actividades académicas y profesionales de los participantes se han abordado distintas instancias de estudio y aplicación de las técnicas de Ingeniería de Requerimientos, por lo cual este estudio de aplicación reviste especial interés.

4. FORMACION DE RECURSOS HUMANOS

En el plano económico de las empresas del sector metalmeccánica tendrán con los resultados de los estudios presentes un insumo para mejorar el conocimiento hacia el interior de sus operaciones, en especial, al respecto de un recurso estratégico como es la tecnología de la información. Los resultados se difundirán entre las empresas participantes de la muestra de investigación. Es de notar que los informes finales de la investigación realizada se colocará a disposición de la Cámara Industrial de Metalúrgicos y componentes de la Provincia de Córdoba y de la Unión Industrial de la Provincia de Córdoba.

Los participantes del proyecto son docentes del área de sistemas de la carrera en Ing. en Sistemas de Información por lo cuál el conocimiento y prácticas adquiridas permiten fortalecer la transferencia de aprendizajes en el aula. Este proyecto permite que los docentes participantes se involucren en actividades de investigación iniciales conformando la primera actividad en la disciplina, y por tanto permitiéndoles ingresar a la carrera formal de docente investigador.

La aplicabilidad será directa, ya que la carrera de Ingeniería en Sistemas dentro de su currícula brinda conocimientos de las diferentes gestiones de las organizaciones y

los tipos de entrevistas a llevar a cabo para la identificación de Modelos de Negocio y de los Requerimientos de información para un Desarrollo de Software.

5. BIBLIOGRAFIA

- León, Orfelio. Diseño de Investigación. Líneas de Investigación. México McGraw Hill 2000.

- Taylor, S. Líneas de investigación . Introducción a los métodos cualitativos de investigación España Paidós 2002.

-Líneas de investigación Sierra Bravo, Restituto Técnicas de investigación social: teoría y ejercicios España .Paraninfo.

- Senn, James. Analysis and Design of Information Systems, McGraw Hill, 1989.

- Sommerville Ian. Ingeniería de Software, Addison Wesley, 2002. Sexta Edición.

- Jacobson, Booch, Rumbaugh. El Proceso de Desarrollo Unificado, Addison Wesley 1999.

- Braude, Eric J. Ingeniería de Software, una perspectiva Orientada a Objetos, AlfaOmega-2003.

<http://www.inti.gov.ar/sabercomo/sc39/inti5.php>.

MODELOS Y MÉTRICAS PARA EVALUAR CALIDAD DE SOFTWARE

Estayno, Marcelo⁽¹⁾; Dapozo, Gladys⁽²⁾; Cuenca Pletch Liliana⁽³⁾, Greiner, Cristina⁽²⁾

(1) Departamento de Informática. Facultad de Ingeniería. Universidad Nacional de Lomas de Zamora
mestayno@fibertel.com.ar

(2) Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura
Universidad Nacional del Nordeste
{gndapozo, cgreiner}@exa.unne.edu.ar

(3) Departamento de Ingeniería en Sistemas de Información. Facultad Regional Resistencia
Universidad Tecnológica Nacional
cplr@frre.utm.edu.ar

CONTEXTO

Las líneas de I/D presentadas en este trabajo forman parte de las actividades definidas en el marco del proyecto “Modelos y métricas para la evaluación de la calidad de software orientados a Pymes”, presentado para su acreditación ante la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE). En este proyecto participan tres unidades académicas del país, los integrantes pertenecen a la Universidad Tecnológica Nacional Regional Resistencia (UTN-FRRe) y a la Facultad de Ciencias Exactas de la UNNE, bajo la dirección de un docente investigador de la Universidad Nacional de Lomas de Zamora (UNLZ).

El objetivo fundamental del proyecto es contribuir a la mejora en la calidad de los productos de software mediante modelos y métricas aplicados al producto y al proceso de creación, diseño, desarrollo y mantenimiento de software, como medio para aumentar la competitividad de las PYMES de la región NEA en el contexto de la industria del software.

RESUMEN

En este trabajo se describen las líneas de investigación y desarrollo que se realizarán en el marco de un proyecto mayor referido a calidad de software orientado a Pymes de la región NEA del país, en el contexto de las iniciativas gubernamentales de promoción de la industria del software. La calidad del software, de por sí compleja y subjetiva, se basa en modelos y métricas que intentan realizar la medición de los distintos aspectos que afectan el proceso de desarrollo y el producto software. Los resultados de la medición y evaluación de la calidad aportarán al mejoramiento del software que se elabora y, consecuentemente, al incremento de la competitividad de quienes desarrollan, como así también, y principalmente, mejorar las aplicaciones y sistemas, y sus ámbitos de aplicación. En particular, en este proyecto se plantea además, estimular la vinculación de las universidades y las empresas Pymes de la región NEA, como así también, el mejoramiento de sus recursos humanos, como forma de contribuir al desarrollo local.

Palabras clave: Ingeniería de software, calidad del software, modelos y métricas de evaluación, calidad de aplicaciones web, programación orientada a objetos, programación orientada a aspectos, técnicas metaheurísticas.

1. INTRODUCCION

En la actualidad, gracias a los avances de la Informática, el software se encuentra en casi todos los campos de la actividad humana: la industria, el comercio, las finanzas, el gobierno, la salud, la educación, las artes, etc. La dependencia de las organizaciones respecto de este recurso es crítica. La automatización de las actividades y la generación y disponibilidad de información para la toma de decisiones, logradas a través del software, son claves para el logro de los objetivos y supervivencia de las organizaciones.

Por las razones expuestas, existe una creciente preocupación por lograr que los productos software cumplan con ciertos criterios de calidad. Para ello, se avanza en la definición e implementación de estándares que fijan los atributos deseables del software de calidad, a la vez que surgen modelos y metodologías para la evaluación de la calidad.

Por otra parte, en los últimos años el sector TIC ha tomado una relevancia particular en la Argentina, y se está asentando como un nuevo pilar del desarrollo nacional. La concentración geográfica de empresas, universidades y todos los actores involucrados en dicho sector, incrementa la utilización de la capacidad instalada de las regiones, potencia su industria y mejora la calidad de sus productos e insumos [1]. En consonancia con estos propósitos, a partir del año 2005, se constituyen en la región, primero el Polo IT Chaco y luego el Polo IT Corrientes, buscando mediante la asociatividad el crecimiento individual y conjunto de las empresas que lo componen.

Calidad del software

La calidad del software es una compleja combinación de factores, que variarán entre diferentes aplicaciones. Diversos autores como Pressman [2], McCall [3] y estándares, como ISO 9126 han tratado de determinar y categorizar los factores que afectan a la calidad del software.

Una definición amplia de calidad, planteada en la norma UNE-EN ISO 8402 [4], expresa que “la calidad es el conjunto de propiedades y características de un producto o servicio que le confieren su aptitud para satisfacer unas necesidades explícitas o implícitas”. Llevada esta definición al campo de la ingeniería de software, la IEEE Std 610 [5], señala que “la calidad del software es el grado con el que un sistema, componente o proceso cumple los requerimientos especificados y las necesidades o expectativas del cliente o usuario”.

Por su parte Pressman [2], se refiere a la calidad del software como “la concordancia con los requisitos funcionales y de rendimientos explícitamente establecidos, estándares de desarrollo explícitamente documentados y características implícitas que se espera de todo software desarrollado profesionalmente”.

Sommerville [6] sostiene que la calidad del software es un concepto complejo que no es directamente comparable con la calidad de la manufactura de productos. En la manufacturación, la noción de calidad viene dada por la similitud entre el producto desarrollado y su especificación. En un mundo ideal, esta definición debería aplicarse a todos sus productos, pero, para sistemas de software, existen cuestiones específicas que impiden aplicar este mecanismo.

Calidad del Producto y del Proceso

Al intentar definir el concepto de calidad del software se debe diferenciar entre la calidad del Producto de software y la calidad del Proceso de desarrollo del mismo. No obstante, las metas que se establezcan para la calidad del producto van a determinar las metas a establecer para la calidad del proceso de desarrollo, ya que la calidad del producto va a estar en función de la calidad del proceso de desarrollo. Sin un buen proceso de desarrollo es casi imposible obtener un buen producto [7].

Hay un vínculo claro entre la calidad del proceso y del producto en producción debido a que el proceso es relativamente fácil de estandarizar y monitorizar. Cada sistema de producción se calibra, y debe producir una y otra vez productos de alta calidad. Sin embargo, el software no se manufactura, sino que se diseña. El desarrollo de software es un proceso más creativo que mecánico, donde las experiencias y habilidades individuales son importantes. La calidad del producto, sea cual fuere el proceso utilizado, también se ve afectada por factores externos, como la novedad de una aplicación o la presión comercial para sacar un producto rápidamente.

Modelos de calidad de software

A lo largo del tiempo se han desarrollado diferentes modelos para evaluar la calidad del software, que intentan descomponer la calidad en una categoría de características más sencillas. Entre ellos puede

mencionarse el de McCall, Evans y Marciniak, Deutch y Willis, FURPS, entre otros [8].

Un hito en la definición de estándares de calidad de producto software, lo constituye la publicación del ISO9126 en el año 1991. Luego, en el año 2001, este estándar fue reemplazado por dos estándares relacionados: el ISO/IEC 9126, que especifica características y métricas de la calidad del software; y el estándar ISO/IEC 14598, que especifica la evaluación de productos de software [9].

El estándar ISO/IEC 9126 se compone de cuatro partes: modelo de calidad [10], métricas externas [11], métricas internas [12] y métricas para la calidad en uso [13]. Propone un modelo de calidad categorizando la calidad de los atributos software en seis características (funcionalidad, fiabilidad, usabilidad, eficiencia, mantenibilidad y portabilidad), las cuales son subdivididas en subcaracterísticas. La calidad de uso es definida como “la capacidad del software que posibilita la obtención de objetivos específicos con efectividad, productividad, satisfacción y seguridad” [14].

El modelo más actual está representado por las normas ISO 25000:2005, conocidas con el nombre de SQuaRE (*Software Quality Requirements and Evaluation*), basada en ISO 9126 y en ISO 14598, se desagrega en 5 tópicos: 1-Gestión de la Calidad (2500n), 2- Modelo de Calidad (2501n), 3- Medidas de Calidad (2502n), 4- Requerimientos de Calidad (2503n) y 5-Evaluación de la Calidad (2504n) [8].

La especificación de requisitos de calidad y la evaluación de productos software son dos procesos que por su inherente complejidad pueden beneficiarse del proceso que regule su realización. Sin embargo, y como señala el estándar SQuaRE, es importante que sus objetivos estén alineados. Por ello, la creación de una norma que regule su realización pueda ser muy beneficiosa, en cuanto a la consistencia de los resultados obtenidos [14].

Otro aspecto destacable de SQuaRE es la incorporación de una normalización de la terminología, considera la Metrología como la ciencia de la medida y la necesidad de amoldar los conceptos usados en Ingeniería del Software a los utilizados en otras disciplinas que hacen uso de la medición [15].

Métricas y Medición.

La medición es un elemento clave en cualquier proceso de ingeniería. Las medidas se emplean para comprender mejor los atributos de los modelos que se crean y evaluar la calidad de los productos de la ingeniería. Por las características inherentes al software, sus medidas y métricas son indirectas y, por lo tanto, expuestas al debate [2].

Una métrica contiene la definición de un método de medición o un método de cálculo y la escala asociada.

El método de medición es la secuencia lógica particular de operaciones y posibles heurísticas, especificada para permitir la realización de la

descripción de una métrica por una actividad de medición. Por otro lado, la escala se define como un conjunto de valores con propiedades definidas.

La propiedad más importante de una escala es su tipo, considerando que puede ser Categórica o Numérica. A su vez, dependiendo de la naturaleza de la relación entre los componentes de la escala, pueden clasificarse en: nominal, ordinal, intervalo, proporción o absoluta. El tipo de escala de los valores medidos define las transformaciones admisibles y afecta las operaciones matemáticas y estadísticas que pueden ser aplicadas.

Las métricas pueden ser directas, sobre las que puede aplicarse un método de medición (objetivo o subjetivo); o indirectas, que son aquellas definidas en función de otras métricas y se calculan en base al método de cálculo asociado, es decir en base a una fórmula.

Calidad en aplicaciones Web

Los avances en Internet han conducido a un desarrollo impactante de sistemas y aplicaciones basadas en la Web, suceso que se presenta como el más significativo en la historia de la Computación

Muchas de las nuevas tecnologías y estándares de la Web han surgido en los últimos años para mejorar el apoyo a nuevas aplicaciones Web: XML, servicios Web, Web semántica, técnicas de personalización de la Web, minería Web, inteligencia, *contextaware* y móviles y servicios Web.

Las aplicaciones web son diferentes de otras categorías de software; son eminentemente de red, las gobiernan los datos y se encuentran en evolución continua. La inmediatez dirige su desarrollo, la seguridad es un requisito prioritario y la demanda de estética, así como la entrega de contenido funcional, son factores diferenciales adicionales [2].

El estudio de la calidad de productos y procesos de desarrollo para la Web es muy reciente y todavía no se dispone de métodos de evaluación ampliamente difundidos para este tipo de entorno, por lo tanto, existe la necesidad de metodologías efectivas para la obtención de aplicaciones Web de calidad.

La Ingeniería Web surge debido a la necesidad de lograr enfoques disciplinados y nuevos métodos y herramientas para desarrollar, desplegar y evaluar los sistemas y aplicaciones basados en la Web. Estos enfoques y técnicas deben considerar las particularidades del nuevo medio, el contexto y los escenarios operativos y, principalmente, la diversidad de perfiles de usuarios que constituyen desafíos adicionales al desarrollo de aplicaciones Web [2].

Cómo se mide la calidad del software para la web? En general, con los mismos modelos que para el software tradicional. Sin embargo, hay características que son más relevantes en este contexto, como por ejemplo, la facilidad de uso, funcionalidad, confiabilidad, eficiencia y facilidad de mantenimiento. Olsina [16] define un “árbol de requisitos de calidad” para aplicaciones Web y

Offut [17] agrega otros atributos como Seguridad, Disponibilidad, Escalabilidad, Tiempo en el Mercado.

Calidad en Programación Orientada a objeto

El desarrollo de programas orientados a objetos (POO) es cada vez mayor, sin embargo, no ha evolucionado al mismo ritmo el uso de métricas para este paradigma. Las métricas orientadas a objetos, al igual que las del software convencional, buscan poder entender mejor la calidad del producto, evaluar la efectividad del proceso y mejorar la calidad del trabajo llevado a cabo al nivel del proyecto.

Sin embargo, la POO difiere en importante medida del desarrollado utilizando enfoques tradicionales. Por esta razón las métricas deben ajustarse a las características que lo distinguen, como ser encapsulamiento, ocultamiento de información, herencia y técnicas de abstracción de objetos que hagan única a esa clase.

Entre las métricas encontradas en la literatura que han tenido relevancia en la orientación a objetos sobresalen las definidas por Abreu, Chidamber y Kemerer y Lorenz y Kidd, que abordan todos los posibles niveles de granularidad y características en sistemas OO, como ser: Métricas a nivel de sistema, Métricas a nivel de acoplamiento, Métricas a nivel de herencia, Métricas a nivel de clases, Métricas a nivel de métodos [18].

Calidad en Programación Orientada a Aspectos

En el modelo de Programación Orientado a Objetos (POO) toda tarea específica debe ser responsabilidad de una clase o de un pequeño número de clases agrupadas de alguna forma lógica. Sin embargo, existen ocasiones en las que determinados servicios se utilizan en diversas clases y no tienen suficiente entidad para incluirlos en una clase específica, lo que provoca repetición de código a lo largo de toda la aplicación. Por ejemplo, los bloques de código dedicados a la sincronización de los accesos a los recursos, a la persistencia de los datos, al registro de auditorías (logs), etc. Estos bloques de código son características o temas de interés (*concerns*) dentro del sistema software. La diseminación de estos concerns a través de varias clases son conocidos como *crosscutting concerns*.

Normalmente los aspectos están mezclados en los principales módulos de los componentes del sistema causando el problema de tener código desordenado. La Programación Orientada a Aspectos (POA) trata de encapsular estas características en módulos en vez de tenerlos dispersos en los componentes del sistema.

En sus inicios, la POA se centró principalmente en el nivel de implementación y codificación, pero en los últimos tiempos esta separación se está llevando a cabo a nivel de diseño. Esto impacta positivamente en la calidad integral del software.

Dado el continuo desarrollo de aplicaciones que contemplan este nuevo concepto, diferentes modelos de métricas han aparecido como una forma de evaluación de la calidad de las mismas [19] [20] [21]. Si bien estos nuevos modelos tienen características similares a los modelos de métricas para sistemas orientados a objetos, tienen también características particulares a los aspectos.

Las métricas sobre las cuales se trabaja consideran: Cantidad de Aspectos, Cantidad de relaciones existentes entre aspectos y una clase, Cantidad de clases relacionadas con un mismo aspecto, Cantidad de puntos de enlace en una clase, Cantidad de Clases Tejidas.

La POA es un nuevo paradigma que aún adolece de madurez y formalidad, por lo que las líneas de investigación apuntan a definir métricas que reúnan las mejores características de las existentes, entre otras.

Técnicas metaheurísticas orientadas a la calidad del proceso de desarrollo

Con el aumento de las prestaciones de las computadoras y recursos asociados el desarrollo de Software se hizo más complejo, pasando de ser una tarea realizada por una sola persona en pocas horas a convertirse en un conjunto de actividades interrelacionadas que deben realizarse en grandes equipos de trabajo durante meses.

Los problemas de optimización se plantean todas las ingenierías y la Ingeniería del Software, a pesar de ser una disciplina joven, no es una excepción. Según Chicano [22] en la actualidad existe un creciente interés por aplicar técnicas de optimización a problemas de Ingeniería del Software, ya sea mediante el uso de algoritmos exactos, heurísticos ad hoc o metaheurísticos. Si bien los algoritmos exactos garantizan encontrar el óptimo global de cualquier problema, tienen el grave inconveniente de que en problemas reales su tiempo de ejecución crece de forma exponencial con el tamaño del problema. Los algoritmos heurísticos *ad hoc*, en cambio, suelen ser bastante rápidos pero las soluciones no suelen ser óptimas. Las metaheurísticas ofrecen un equilibrio adecuado entre ambos extremos: son métodos genéricos que ofrecen soluciones de buena calidad en un tiempo moderado. En esta línea de investigación se analizan las alternativas para aplicar, a los problemas de optimización en Ingeniería de Software, algoritmos metaheurísticos que ofrezcan soluciones de cierta calidad en un breve periodo de tiempo: un compromiso entre calidad de la solución y rapidez. Existen numerosos trabajos en este sentido, que abordan distintas etapas del desarrollo de Software: Des Creer y Ruhe [23] abordan el problema de la selección de requisitos para cada iteración del proceso de desarrollo, Simons y Parmee [24] plantean el diseño conceptual como un problema de optimización. El problema de la generación automática de código paralelo óptimo a partir de

código secuencial ha sido abordado por Nisbet [25] y Williams [26]. Soluciones a los problemas de optimización relacionados con la planificación de proyectos y la generación automática de casos de prueba han sido propuestas por Enrique Alba et al en [27] y [28]. En [22] Chicano presenta un análisis de la evolución de este campo de investigación, demostrando un aumento casi lineal de los trabajos a partir de 1999. El mismo informe demuestra que el mayor interés de la comunidad científica se ha centrado en la fase de pruebas, seguida por la fase de gestión de proyectos. Esto podría deberse, según el autor, a que aproximadamente la mitad del tiempo de un proyecto software y más de la mitad de su costo se dedica a la fase de pruebas. Se trata entonces de un campo de investigación que es necesario explorar para definir las áreas de vacancia donde sea posible realizar aportes sustantivos con relevancia regional.

2. LINEAS DE INVESTIGACION y DESARROLLO

- Análisis, estudio y discusión de modelos de evaluación de calidad, estándares y metodologías.
- Relevamiento de metodologías y herramientas utilizadas por las empresas Pymes de la región para la evaluación de la calidad de sus productos software y la calidad del proceso de desarrollo de software.
- Análisis y estudio de modelos y métricas de evaluación de calidad de uso de aplicaciones Web. Propuesta metodológica para la aplicación de un modelo adaptado a las características de las Pymes locales.
- Relevamiento de modelos y métricas de evaluación de calidad de aplicaciones orientadas a objetos. Aplicación y comparación de, por lo menos, dos metodologías relevantes.
- Análisis y estudio de modelos y métricas de evaluación de calidad de aplicaciones orientadas a objetos. Propuesta de un modelo para evaluar aplicaciones orientadas a aspectos.
- Identificación de problemas de optimización en Ingeniería del Software.
- Descripción, formalización, aplicación y evaluación de técnicas metaheurísticas para resolver problemas de optimización de la Ingeniería del Software.
- Selección, descripción y formalización de los problemas que se detecten como relevantes para el campo científico y para el desarrollo regional.

3. RESULTADOS ESPERADOS

Dentro de las líneas de trabajo mencionadas, se espera obtener los siguientes resultados:

- Conocimiento actualizado de los distintos modelos, métricas, estándares y herramientas de evaluación de calidad de software para su

aplicación al ámbito de las PYMES en el contexto regional NEA.

- Diagnóstico de la importancia que las Pymes de software confieren a los modelos y métodos de aseguramiento de la calidad, así como a las herramientas utilizadas para tal fin.
- Metodología que permita evaluar la calidad en uso de productos software sobre plataformas Web y que brinde información que aporte a la mejora del proceso de creación de software.
- Nuevas herramientas o adaptación de las existentes para automatizar la generación de los indicadores que conforman las métricas de evaluación de la calidad en uso de aplicaciones web.
- Vinculaciones con otras instituciones, universitarias, gubernamentales o empresariales (especialmente las PYMES regionales), para el estudio y desarrollo de metodologías y/o desarrollo de herramientas relacionadas con la evaluación de la calidad en el software.-
- Especialización de recursos humanos en los temas propios de la evaluación de calidad de software a través de la elaboración de tesis de carreras de posgrado y trabajos de investigación realizados en el marco de becas de investigación para alumnos y/o graduados.
- Realización de transferencias tecnológicas o de capacitación en temas relacionados con la mejora de la calidad de productos software desarrollados por Pymes regionales/nacionales.

La implementación de este proyecto permitirá el desarrollo de actividades de investigación en un área actualmente considerada de vacancia en la región de influencia de la UNNE y la UTN - Facultad Regional Resistencia, posibilitando la inserción y el crecimiento en el camino de la investigación de alumnos y recientes graduados de las carreras de Informática ofrecidas por ambas universidades.

A su vez, contribuirá a mejorar la formación de los profesionales de la Informática dado que la generación y la actualización de los conocimientos, se trasladará a los alumnos a través del proceso de enseñanza y aprendizaje, en cada una de las asignaturas dictadas por los integrantes del proyecto. Finalmente, permitirá que las universidades realicen transferencia de conocimientos al medio regional, contribuyendo de esta forma a mejorar la competitividad de las empresas Pymes, generar valor para la zona mediante la creación de puestos de trabajo, facilitar la inserción de los profesionales formados en la universidad, evitando el desarraigo de los jóvenes graduados y la migración a los grandes centros urbanos, cambios importantes que conllevan un deterioro de la calidad de vida de las personas.

4. FORMACION DE RECURSOS HUMANOS

En el marco del proyecto se encuentran en este momento en desarrollo un plan de trabajo de beca de investigación de pregrado de la SECYT-UNNE, se proyecta el desarrollo de cinco tesis de la Maestría en Ingeniería de Software de la UNLP dictada en FaCENA-UNNE y UTN-FRRe, mediante acuerdo institucional, como así también, la elaboración de dos tesis doctorales, correspondientes al Doctorado en Ingeniería de Sistemas y Computación de la Universidad de Málaga- España.

5. BIBLIOGRAFIA

- [1] Primer Foro Nacional de experiencias en el Sector TIC. "El Encuentro de los Polos de Argentina". 2007. Disponible en: http://www.ses.me.gov.ar/spu/Noticias/Noticias_Universitarias_2007/noticias_marzo_2007_asociacion_de_tec.htm
- [2] Pressman, R. S. "Ingeniería de Software. Un enfoque práctico". Editorial MCGRAW-HILL - 2005
- [3] McCall, J.A. ; Richards, P.K. ; Walters, G.F. – "Factors in Software Quality". Vols I, II, III. NTIS AD-AO49-014, 015, 055, Nov. 1977
- [4] AENOR. UNE-EN-ISO 8402 "Gestión de la calidad y aseguramiento de calidad". Vocabulario (ISO 8402:1994).
- [5] IEEE Standard 610. Institute of Electrical and Electronics Engineers Computer dictionary. Compilation of IEEE Standard Computer Glossaries. 1990
- [6] Ian Sommerville. "Ingeniería del Software". 7ª Ed. Cap. 27. Pearson Educacion S.A., Madrid 2005.
- [7] Scalone F. "Estudio comparativo de los modelos y estándares de calidad del software". Tesis de Maestría en Ingeniería en Calidad. Universidad de Buenos Aires. 2006. Disponible en <http://www.fi.uba.ar/laboratorios/lisi/scalone-tesis-maestria-ingenieria-en-calidad.pdf>
- [8] Piattini, M., García F., Caballero, I. "Calidad de los Sistemas Informáticos". Editorial Alfaomega. 2007.
- [9] Marín B., Condori-Fernández N., Pastor O., "Calidad en Modelos Conceptuales: Un Análisis Multidimensional de Modelos Cuantitativos basados en la ISO 9126". Revista de Procesos y Métricas.
- [10] ISO, "ISO/IEC 9126-1 – Software engineering – Product quality – Part 1: Quality Model", 2001.
- [11] ISO, "ISO/IEC 9126-2 – Software engineering – Product quality – Part 2: External Metrics", 2003.
- [12] ISO, "ISO/IEC 9126-3 – Software engineering – Product quality – Part 3: Internal Metrics", 2003.
- [13] ISO, "ISO/IEC 9126-4 – Software engineering – Product quality – Part 4: Quality in Use Metrics", 2004.
- [14] Ruiz Morilla, J., "ISO 9126 vs. SQuARE". Material del curso de posgrado Calidad y Medición de Sistemas de Información. Escuela Superior de Informática. Universidad de Castilla-La Mancha.

2008. Disponible en: <http://alarcos.inf-cr.uclm.es/doc/cmsi/trabajos/Joaquin%20Ruiz%20Expo.pdf>

[15] Gómez García, O. “SQuaRE: Una unificación de normas para la especificación de requisitos y la evaluación de la calidad”. Material del curso de posgrado Calidad y Medición de Sistemas de Información. Escuela Superior de Informática. Universidad de Castilla-La Mancha. 2009. Disponible en: <http://alarcos.inf-cr.uclm.es/doc/cmsi/trabajos/Oscar%20Gomez.pdf>

[16] Olsina L. et al “Specifying Quality Characteristics and attributes for web sites”. ACM. 1999.

[17] Offutt, J. “Quality attributes of web software applications”, in *IEEE Software*. 2002, pp.25-32

[18] Rodríguez, D., Harrison, R., "Medición en la Orientación a Objeto", in *Medición para la Gestión en la Ingeniería del Software*, Dolado, J. and Fernández, L., Eds., RA-MA, 2000, ISBN 84-7897-403-2.

[19] Marchetto, A. “A concerns-based metrics suite for web applications”. In *INFOCOMP Journal of Computer Science*, 4(3):11–22, 2005.

[20] Grigoreta, M. S.; Serban, G. - “Quality measures for evaluating the results of clustering based aspect mining techniques”. In *Proc. of TEAM '06 (Towards Evaluation of Aspect Mining) Workshop in ECOOP '06*, pages 13–16. TU Delft, June 2006.

[21] Zhao, J. - “Towards a Metric Suite for Aspect-Oriented Software”. - Technical Report SE-136-25, Information Processing Society of Japan (IPSJ), Marzo 2002.

[22] Chicano García, Francisco. “Metaheurísticas e Ingeniería del Software”. Phd Thesis. Departamento de Lenguajes y Ciencias de la Computación. Universidad de Málaga. España. 2007.

[23] Des Greer and Günther Ruhe. “Software release planning: an evolutionary and iterative approach.” En *Information and Software Technology*, 46(4), 243:253, 2004.

[24] C. L. Simons and I. C. Parmee. “Single and multi-objective genetic operators in object-oriented conceptual software design”. In *GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pages 1957:1958. ACM Press, 2006.

[25] Andy Nisbet. “GAPS: A compiler framework for genetic algorithm (GA) optimised parallelisation”. In *High-Performance Computing and Networking, International Conference and Exhibition, HPCN Europe 1998, Amsterdam, The Netherlands, April 21-23, 1998, roceedings*, volume LNCS 1401, pages 987:989. Springer, 1998

[26] Kenneth Peter Williams. “Evolutionary Algorithms for Automatic Parallelization”. PhD thesis, University of Reading, UK, 1998.

[27] Enrique Alba y Francisco Chicano. “Software project management with Gas”. *Information Sciences*, 177(11), 2380:2401, Junio 2007.

[28] Enrique Alba y J. Francisco Chicano. “Software testing with evolutionary strategies”. En *Proceedings of the 2nd International Workshop on Rapid Integration of Software Engineering Techniques, LNCS 3943*, páginas 50:65, Heraklion, Grecia, Septiembre 2005.

CALIDAD TOTAL EN UN MODELO INTEGRADO DE PROCESO SOFTWARE Y CICLO DE VIDA

Alicia Mon¹ y Javier Garzás²

¹Grupo GIS-UNLaM

Universidad Nacional de La Matanza, Argentina

aliciamon@fibertel.com.ar

²Kybele Research, Dpto. de Lenguajes II;

Universidad Rey Juan Carlos; Madrid, España.

Javier.Garzas@urjc.es

CONTEXTO

La presente línea de investigación se desarrolla en el contexto de I/D del Grupo de Ingeniería de Software GIS de la Universidad Nacional de La Matanza que trabaja en proyectos sobre proceso software y modelos de calidad para la industria del software. Asimismo, la investigación forma parte de una tesis doctoral que se está desarrollando en el Depto. de Lenguajes II de la Universidad Rey Juan Carlos de Madrid, participante de numerosos proyectos de I/D sobre Fábricas de Software y mejora de procesos.

RESUMEN

La modelización del Proceso Software constituye un marco de referencia para la organización de las actividades que involucran todas las etapas del desarrollo. La representación del ciclo de vida define los estados por los que pasa un producto software y la representación del proceso software define el conjunto de actividades esenciales no ordenadas en el tiempo que requiere el desarrollo de software.

En el presente artículo se exponen los avances de una investigación que propone una modelización integrada del Proceso Software y del Ciclo de Vida del producto en la que se diferencian actividades y productos con incorporación de mejoras inspiradas en los Modelos de Proceso industriales, tales como la Calidad Total.

Palabras Clave: Proceso Software, Modelos de Proceso, Ciclos de Vida Software

1. INTRODUCCION

La modelización de los procesos para la industria en general, requiere de la Ingeniería de Procesos como un factor clave para brindar niveles de calidad predecibles y escalables, basándose en la clara definición de las actividades de los procesos. La forma de aumentar la eficiencia en este sentido, consiste en centrar la producción en procesos y mejorar la capacidad de éstos.

El diseño, la medición y la mejora de los procesos en ciclos se han revelado como la clave para mejorar de forma continua la eficiencia y la calidad productiva. En esta irrupción de la relevancia del proceso, algunos modelos de proceso industrial configuran una ecuación en la que, las personas ayudadas por la tecnología actúan como recursos para ejecutarlos.

En los aspectos organizativos centrales del proceso en la industria del software, la materia prima que posibilita la construcción del producto software es el conocimiento, constituido en el único elemento capaz de generar valor, a través del capital intelectual y la capacidad de transformarlo en producto.

La modelización del proceso software debe considerar los productos del software que va construyendo, tanto como la relación entre los

procesos específicos y los entornos socio-organizacionales en los cuales se desarrolla que se determinan por esa capacidad intelectual generada por el conocimiento.

Los diferentes estándares, Modelos de Proceso [IEEE,1997]; [ISO/IEC,2002] o Modelos de Madurez [ISO,1998]; [CMMI,2006][CMMI,2008] y Modelos de Ciclos de Vida [Boehm, 1988]; [Boehm,1994]; [Alexander,1991]; [Jacobson, 1999] comparten la concepción de que una buena definición del proceso software facilita el aseguramiento de que cada elemento de trabajo se asigna y se desarrolla adecuadamente, lo que indica que la calidad del proceso determina la calidad del producto.

Un análisis global de estos marcos de referencia determina la falta de consenso en la terminología empleada para caracterizar los enfoques de modelización del proceso software y de los modelos de ciclos de vida. En este campo se trabaja esta perspectiva diferencial entre el producto y el proceso, no obstante, en la definición y descripción de la mayoría de los modelos de ciclos de vida propuestos en la actualidad, hacen referencia al ciclo de vida del producto software a través del conjunto de actividades del proceso software. Por lo tanto, esta separación entre producto y proceso no queda claramente diferenciada en el desarrollo de software.

El desarrollo conceptual del proceso software no es ajeno a la evolución de las formas de organización en la producción industrial en general. El surgimiento de Metodologías Ágiles [Beck, 2001]; [Fowler,2001]; [Larman,2004] en el desarrollo de software da cuenta de ello, al introducir formas de organización provenientes de la mejora de los modelos de proceso industriales en la actualidad [Coriat,1992a]; [Coriat,1992b], tales como el capital intelectual, la importancia de los conocimientos el desarrollo modular, la flexibilidad en los procesos, o la programación por pares [Beck, 1999], todas ellas tendentes a quebrar las fronteras en los compartimentos estancos de los equipos de desarrollo.

En la producción industrial, los modelos de proceso se han desarrollado en la medida que la ciencia y la tecnología han ido evolucionando. Los cambios sucedidos en los modelos de proceso industriales no responden exclusivamente a los avances tecnológicos, sino a las formas de organización del proceso de trabajo en el ámbito de las fábricas y las empresas.

Si bien el producto software no puede ser comparado con los productos ingenieriles o industriales clásicos, es posible establecer similitudes entre el proceso de desarrollo o de construcción del software, en términos de organización, y los elementos de los modelos de proceso industriales.

Las innovaciones provenientes de los procesos industriales, podrían incorporarse al desarrollo de software con el fin de obtener mejoras semejantes a las ocurridas con su introducción en la producción industrial.

2. LINEAS DE INVESTIGACION Y DESARROLLO

En la presente investigación, se propone un Modelo de Proceso Software, que representa un enfoque dirigido por el desarrollo del producto en cuanto al ciclo de vida del software y por el proceso software en cuanto al conjunto de actividades sistematizadas.

El modelo propuesto diferencia conceptualmente el proceso software con una definición clara de actividades y el ciclo de vida de producto, incorporando ambos conceptos en un único modelo. Esta diferenciación, permite separar las actividades del proceso, de las características técnicas de los productos que construye, y comparar el proceso software con los procesos industriales para incorporar al desarrollo de software las mejoras generadas en otros campos de la producción.

El modelo propuesto está centrado en el conocimiento de los recursos humanos que requiere el desarrollo de software, como capital intelectual intrínseco de la producción. Representa

la construcción del producto software en una serie de versiones, conformando cada una de ellas un ciclo de desarrollo indeterminado en el tiempo, en una secuencia no ordenada ni predefinida de actividades y en una secuencia indeterminada de productos construidos, que incorpora los conceptos de calidad total al producto[Coriat,1992b], garantizado por lo particular de la organización de las actividades del proceso basadas en el conocimiento y en la detención del proceso ante la presencia de fallos.

3. RESULTADOS OBTENIDOS

3.1. Características del Modelo propuesto

El Modelo de Proceso Software desarrollado, representa un enfoque dirigido por el desarrollo del ciclo de vida de la construcción del producto software y por el proceso software con el conjunto de actividades sistematizadas, repetibles, no ordenadas en el tiempo como algo diferente del ciclo de vida, es decir que las actividades serán repetibles e independientes del ciclo de vida que adopte un producto software específico.

Dirigido por la gestión para el análisis y estructuración del proceso software, se propone un modelo de desarrollo evolutivo que acompaña la naturaleza iterativa y concurrente de las actividades de construcción del producto en su ciclo de vida software, con los aspectos controlados y sistemáticos del proceso software.

El Modelo propone la construcción del producto software en una serie de versiones, conformando cada una de ellas un ciclo de desarrollo indeterminado en el tiempo, en la secuencia de actividades y en la secuencia de productos construidos. Las actividades del proceso son acciones que realiza un equipo de desarrollo, que están definidas por los estándares de proceso y el ciclo de vida del producto constituye los estados de transformación por los que va pasando el producto software a lo largo de las actividades del proceso software.

Esta diferenciación, permite separar las actividades del proceso, de las características de los productos que construye, y de esta manera poder analizar genéricamente el proceso software comparativamente con los procesos industriales, por una parte, y los productos software, por otra parte.

3.2. Conocimientos del equipo

El conocimiento es considerado como una materia prima inherente al producto que debe desarrollarse. La planificación de la producción debe ser consensuada y coordinada. Debe realizarse en cada puesto o módulo funcional y para construir la planificación global coordinada.

La pluriespecialización de los integrantes de un equipo de trabajo, implica que el amplio conocimiento y la alta cualificación generan una relación participativa en la toma de decisiones, y en el conocimiento de todas las instancias de la producción o desarrollo.

El modelo propone incorporar prácticas que implican un mayor acercamiento en el conjunto de tareas, así como un vínculo más estrecho entre la concepción y la ejecución de tareas, lo que se logra a partir del conocimiento como capital e insumo de la producción.

3.2. Calidad

El Modelo propuesto incorpora los conceptos de calidad total basados en la detención del proceso de producción ante la presencia de fallos, en la que cada instancia del producto, desarrollada en una fase del Modelo, puede pasar a otra fase del desarrollo, cualquiera que sea, solo si tiene una aceptación de que todo lo necesario ha sido completado, y si el producto que la requiere para iniciar su ejecución cuenta con los conocimientos necesarios para comenzar dicha instancia.

Si se detecta un fallo, el producto no pasa de fase y se detiene el proceso hasta que el problema está resuelto. Es condición, la existencia de cero defectos para que una parte del producto pase a la siguiente fase de su desarrollo.

De este modo, cada fase continúa trabajando en forma concurrente en cualquier instancia del producto si es que tiene lo estrictamente necesario para avanzar, provisto por los procesos anteriores, sin ninguna organización temporal y/o secuencial determinada a priori.

La Figura 1 expone una representación gráfica del modelo, en el cual las líneas paralelas representan las actividades del proceso software y las líneas tipo meridianos, representan los estados del producto software o sub productos por los que pasa a través del ciclo de vida, que es finalizado al llegar a cada uno de los polos del círculo.

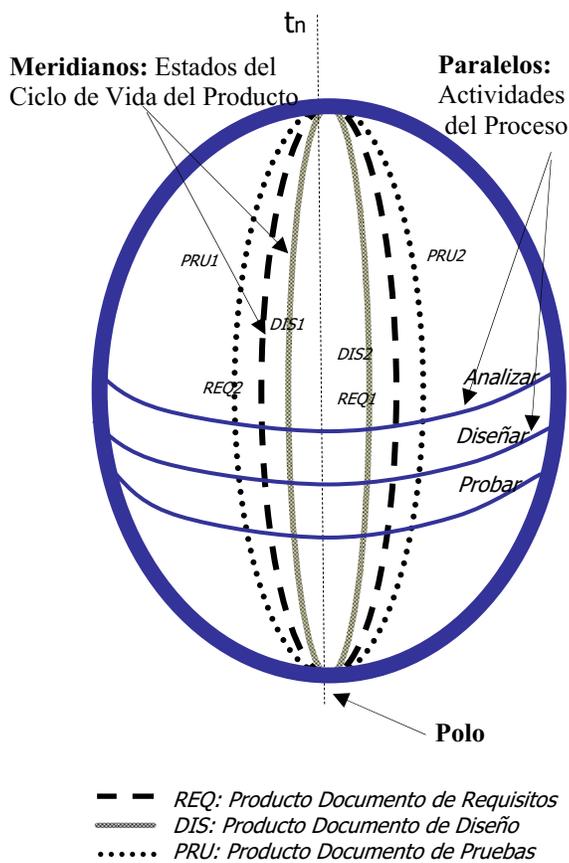


Figura 1. Representación del Modelo

3.4. Iteratividad versus Secuencialidad

En el Modelo desarrollado, la construcción de los productos se inicia a partir de la entrega de un producto en un polo, requiere, como condición esencial el conocimiento aportado por otro producto anterior. Esta entrega de producto

conlleva el conocimiento necesario para que se dispare el inicio de otro producto.

La construcción de los productos es realizada por las actividades del proceso, en tanto que, la transmisión de la condición de inicio entre los productos determina la falta de secuencialidad en las actividades del proceso.

La iteratividad y sucesión no secuencial de actividades es generada por las características de cada producto y por la indeterminación del tiempo de desarrollo de cada una de las actividades del proceso.

Las condiciones para que cada producto sea entregado en un polo, la brindan las diferentes técnicas aplicadas en el proceso para cada una de las actividades que garantizan la calidad en el producto, como por ejemplo, las técnicas de requisitos, las técnicas de validación de diseño, las técnicas de verificación del código, etc.

Las técnicas utilizadas para cada instancia del producto accionan para probar el conocimiento acerca del producto, que puede ser transmitido a los demás productos. En este sentido, cada producto entregable en un polo está conformado por el conocimiento sobre sí mismo.

3.5. Estados del ciclo de vida

Cada producto, en el estado en que se encuentre en el ciclo de vida, se comporta como un dispositivo receptor y transmisor de conocimiento sobre cómo deben seguir los otros productos. De este modo, cada instancia del producto, desarrollado en una fase del ciclo de vida puede pasar a otra fase, cualquiera que sea, solo si tiene una aceptación de que todo lo necesario para esa parte ha sido completado y si el producto que lo requiere para iniciar su ejecución cuenta con los conocimientos necesarios para comenzar dicha instancia.

El desarrollo concurrente y paralelo de diversos productos en una misma fase del ciclo de vida, no se encuentra determinado en el orden de sucesión de instancias, sino que se va determinando de acuerdo a las condiciones para que cada producto,

se construya, se detenga o se entregue en un polo. Cada actividad comienza a desarrollar un producto cuando tiene las condiciones para hacerlo.

Los productos que se construyen no tienen una secuencia definida en el ciclo de vida, sino que la consecución del producto está definida por la indeterminación de lo que cada producto va construyendo y la indeterminación del tiempo para realizar las actividades de su construcción.

4. FORMACION DE RECURSOS HUMANOS

El presente artículo expone las líneas generales de una investigación que se está llevando a cabo en el Departamento de Lenguajes II de la Universidad Rey Juan Carlos de Madrid, España, en el marco de la tesis doctoral “Incorporación de la calidad total de la producción industrial en un modelo integrado de proceso software y ciclo de vida” de una investigadora del Grupo GIS de la Universidad Nacional de La Matanza.

En la Tesis se desarrolla una propuesta de modelo de proceso que integra la ingeniería de proceso con la ingeniería de producto para la industria del software, incorporando de manera definida y sistematizada las mejoras provenientes de los modelos de procesos industriales, tales como la detención del proceso ante la presencia de fallos y la transferencia de conocimiento, como elementos organizativos de la Calidad Total del producto.

5. BIBLIOGRAFIA

[Alexander, 1991] Alexander, L. and Davis, A. “Criteria for selecting software process models”. Proceedings of COMPSAC’91. 521-528. 1991.

[Beck, 1999] Beck, K. Extreme Programming Explained: Embrace Change. (Addison Wesley, 1999).

[Beck, 2001] Beck, K; Beedle, M; Cockburn, A; Cunningham, W; Fowler, M; Agile Manifesto. web site. (2001). <http://agilemanifesto.org>.

[Boehm, 1988] Boehm, B. A Spiral Model of Software Development and Enhancement Computer, pp. 61-78, May 1988.

[Boehm, 1994] Boehm, B. & R. Ross A Collaborative Spiral Software Process Model Based on Theory W. Proceedings, ICSP3. IEEE, Reston, VA, Octubre 1994.

[CMMI, 2006] Capability Maturity Model Integration for Development (CMMI-DEV), Version 1.2. Software Engineering Institute, Carnegie Mellon University. 2006.

[CMMI, 2008] Chrissis, M.B., Konrad, M., Shrum, S. CMMI: Guidelines for Process Integration and Product Improvement (Sei Series in Software Engineering). Amazon, 2008.

[Coriat, 1992a] B. Coriat. El Taller y el Robot. Siglo XXI, 1992.

[Coriat, 1992b] B. Coriat. Pensar al Revés: Trabajo y Organización en la Empresa Japonesa. Siglo XXI, 1992.

[Fowler, 2001] M. Fowler. “Is design dead?” Proceedings XP2000. Web site (2001). <http://www.martinfowler.com/articles/designDead.html>. <http://www.refactoring.com/>

[IEEE, 1997] IEEE. IEEE Standard for Developing Software Life Cycle Processes, IEEE Standard 1074-1997. IEEE, 1997.

[ISO/IEC, 2002] ISO/IEC. International Standard: Information Technology. Software Life Cycle Processes, ISO/IEC Standard 12207-1995/Amd. 1-2002.

[ISO/IEC, 1998] ISO/IEC. ISO/IEC TR 15504. Information Technology – Software process assessment. International Organization for Standardization, International Electrotechnical Commission, 1998. <http://www.sel.iit.nrc.ca/spice>

[Jacobson, 1999] Jacobson, I; Booch, G; Rumbaugh, J. El Proceso Unificado de Desarrollo de Software, Addison Wesley, 1999.

[Larman, 2004] C. Larman. Agile & Iterative Development. A Manager’s Guide. (Addison Wesley, 2004).

MODELADO DE APLICACIONES SENSIBLES AL CONTEXTO

Quincoces, V.E.; Gálvez, M.; Méndez, S. A.; Zapana, J.V.; Ramos, H. O; Cáceres, N. R.; Vega, A. A
Grupo GRISECO
Facultad de Ingeniería
Universidad Nacional de Jujuy
vquincoces@fi.unju.edu.ar

CONTEXTO

Esta línea de investigación está desarrollada por el grupo GRISECO (Grupo de Ingeniería de Aplicaciones Sensibles al Contexto), integrado por docentes de distintas asignaturas y alumnos de la Facultad de Ingeniería de la Universidad Nacional de Jujuy, que se constituyó en el mes de agosto del año 2007. El proyecto que se expone tuvo su iniciación en el año 2008. Algunos de los integrantes del grupo han participado en otras líneas de investigación con anterioridad al presente.

RESUMEN

Los Sistemas Informáticos fueron evolucionando desde aplicaciones científicas, comerciales y de escritorio, hasta el momento actual, en que pueden brindar servicios de acuerdo a la ubicación, tiempo y perfil del usuario. Las aplicaciones que brindan este tipo de funcionalidad se denominan Aplicaciones Sensibles al Contexto (CA – Context Aware), donde tienen predominancia los dispositivos móviles y el perfil del usuario. También va cambiando la forma de uso de la computadora, evolucionando desde brindar servicios a pedido del usuario hasta brindarlos en forma automática y extender la percepción y la interacción del usuario con el mundo real (realidad aumentada).

La Ingeniería de Aplicaciones CA se ocupa de proporcionar un enfoque sistematizado para desarrollar este tipo de aplicaciones. El propósito de este proyecto es abarcar los aspectos de Ingeniería del software para el desarrollo de aplicaciones CA y en particular aplicar patrones en el modelado.

Palabras clave: Ingeniería de Software - Aplicaciones sensibles al contexto - Orientación a Objetos – Patrones – Modelado

1. INTRODUCCION

Korkea-aho (2000) expresa que cuando las personas interactúan entre sí o con el ambiente circundante se hace uso de información situacional implícita. Intuitivamente una persona interpreta el contexto de una situación y reacciona apropiadamente. Las computadoras, en cambio, no son tan buenas como los humanos deduciendo información situacional del ambiente y usándola en sus interacciones. Esto constituye un desafío para las aplicaciones con respecto a la captura de la información y servicios ofrecidos de acuerdo a la ubicación, la actividad y las preferencias del usuario. Hay muchas formas diferentes de usar la información de contexto para hacer los sistemas de computación y las aplicaciones más amigables, flexibles y adaptables. Su uso produce valor agregado y es especialmente importante en un ambiente móvil, donde la interacción, ejecución y uso necesita ser cambiado rápidamente.

Contexto significa información situacional. Según Dey & Abowd (1999) "*Contexto es cualquier información que puede ser usada para caracterizar la situación de una entidad. Una entidad es una persona, lugar u objeto que es considerado relevante para la interacción entre un usuario y una aplicación, incluyendo el usuario y la aplicación en sí mismas*". Así también, cualquier información disponible en el tiempo de una interacción puede ser considerada como información de contexto. Algunos ejemplos de ésta son: identidad, información espacial del personal y maquinaria (ubicación, orientación, velocidad, aceleración), información temporal (hora del día, fecha, estación del año), información del ambiente (temperatura, calidad de aire, nivel de luz o

ruido), situación social (con quién está una persona y gente que está a su alrededor), recursos que están próximos (dispositivos accesibles y hosts), disponibilidad de recursos (batería, pantalla, red, ancho de banda), medidas fisiológicas (presión, ritmo cardíaco, ritmo respiratorio, actividad de los músculos y tono de voz), actividad (hablando, leyendo, caminando, corriendo), cronogramas y agendas.

Un sistema es sensible al contexto (context aware – CA) si puede extraer, interpretar y usar información de contexto y adaptar su funcionalidad al contexto actual de uso. El desafío para tales sistemas radica en la complejidad de capturar, representar y procesar los datos contextuales. Para capturar información de contexto generalmente se necesitan sensores o programas adicionales y para transferirla a aplicaciones, para que éstas la puedan usar, debe existir un formato de representación común.

Según Dey & Abowd (1999), Schilit *et al.* (1994) y Pascoe *et al.* (1999), las características que distinguen las aplicaciones sensibles al contexto son:

- Información y servicios que pueden ser presentados al usuario de acuerdo al contexto actual.
- La ejecución automática de un servicio cuando el usuario se halla en un cierto contexto.
- Guardar información de contexto para una recuperación posterior.

Según Abowd (1999), las aplicaciones sensibles al contexto implican mucho trabajo para su construcción y son aún más difíciles de mantener debido a su naturaleza “orgánica”. Por esta razón, mejorar la modularidad y reducir el acoplamiento entre objetos es extremadamente necesario cuando se diseña este tipo de software.

Entre las aplicaciones CA se puede destacar, entre otras, las de Ambient Learning, Realidad Aumentada y Adaptive Hipermedia.

Del resultado de diversos estudios en el campo de las aplicaciones CA han surgido diversas propuestas de frameworks en los

cuales se implementan modelos de arquitecturas para aplicaciones sensibles al contexto. En Dey (2001) se propone el framework Context Toolkit basado en widgets, en Winograd (2001) se presenta la arquitectura denominada Blackboard centrada en los datos más que en los procesos, en Finkelstein (2002) se presenta un framework para desarrollo de aplicaciones Web basado en reglas, en Gordillo *et al.* (2006) se propone una aproximación de diseño para ingeniería de servicios sensibles al contexto basado en capas y en los mecanismos de dependencia en orientación a objetos. Fortier *et al.* (2005) proponen una aproximación para diseñar servicios sensibles a la ubicación y describen cómo extender las aplicaciones con nuevo comportamiento sensible al contexto.

Todos coinciden en que es necesario separar la aplicación de todo lo concerniente a la captación, interpretación, control y supervisión de la información contextual; como así también de los servicios que se puedan generar a partir del procesamiento de ésta.

Un patrón es una solución a un problema recurrente. Los patrones surgen de la experiencia y aceleran los tiempos de desarrollo de software porque captan la esencia y la solución de un problema y mejoran la comunicación entre desarrolladores. Se enuncian en la forma de problema-solución-contexto donde se aplican. Los patrones de diseño son utilizados a nivel de microarquitectura y los patrones arquitectónicos aparecen a un alto nivel de abstracción (interacción de componentes).

El trabajo de Chung *et al.* (2004) presenta 45 pre-patrones de diseño para computación ubicua y en el trabajo de Tesis de Yorío (2006), se especifica una clasificación de patrones arquitectónicos en aplicaciones móviles.

En Distant et al. (2007), se propone un modelo lógico intermedio denominado UML-MVC para diseño e implementación de aplicaciones Web Ubicuas, que está

basado en diagramas de UML y el patrón de arquitectura MVC.

Se considera que la nueva generación de aplicaciones CA serán tan extensas que una sola compañía o grupo de desarrollo no las va a poder proporcionar por sí sola. Para solucionar este problema es necesario implementar una plataforma integrada, para que grupos independientes de construcción de software creen módulos que interactúen (Fortier, 2006).

2. LINEAS DE INVESTIGACION y DESARROLLO

- Se investiga sobre el modelado de aplicaciones para procesos industriales y en particular para la extensión de sistemas existentes de soporte a la toma de decisiones.
- Se realizan modelos de prototipos para procesos industriales que interactúen con el contexto en diferentes escenarios, es decir, experimentar con diferentes tipos de sensores, actuadores y diferentes contextos.
- Se realizan modelos de arquitectura CA aplicando diferentes tipos de patrones utilizando distintos dispositivos móviles.

3. RESULTADOS OBTENIDOS/ESPERADOS

Objetivos generales en curso:

- Investigar sobre el modelado de Aplicaciones Sensibles al Contexto en particular el Modelado de patrones.
- Consolidar en la Facultad de Ingeniería de la UNJu un grupo de investigación que se dedique al modelado de aplicaciones sensibles al contexto.

Objetivos particulares:

- Realizar el modelado de los Sistemas de Soporte a la toma de decisiones de procesos industriales.
- Realizar el modelado de los servicios y del perfil de usuario de los sistemas

que contemplen la utilización de dispositivos móviles.

- Aplicar patrones existentes, modificados o desarrollados por **GRISECO** que sean pertinentes a los dominios involucrados en la investigación.

Actividades propuestas en el año 2008

- Realizar el estudio sobre los dominios y obtención de especificaciones.
- Analizar perfiles de usuarios.
- Realizar el modelado de aplicaciones CA.
- Revisión bibliográfica encarada hacia patrones.

Resultados

Trabajos realizados y presentados en las IV Jornadas Regionales de Ciencia y Tecnología de las Facultades de Ingeniería del NOA, Universidad Nacional de Santiago del Estero, 4 y 5 de diciembre de 2008:

- Modelo orientado a objetos sensible al contexto basado en capas para aplicaciones de automatización de procesos industriales. Viviana E. Quincoces, María del P. Gálvez, Sandra A. Méndez, Héctor O. Ramos. José V. Zapana, Ariel A. Vega, Nélica R. Cáceres. Investigaciones en Facultades de Ingeniería del NOA, ISBN 978-987-1341-37-5, Vol I, pág 34-39,
- J2ME en Aplicaciones Sensibles al Contexto: Diseño de una aplicación móvil. José V. Zapana, Viviana E. Quincoces, María del P. Gálvez.. Investigaciones en Facultades de Ingeniería del NOA, ISBN 978-987-1341-37-5, Vol I, pág 46-51.
- Uso de patrones en el modelado de aplicaciones sensibles al contexto que utilizan dispositivos GPS. Ariel A. Vega, Sandra A. Méndez, María del P. Gálvez, Viviana E. Quincoces. Investigaciones en Facultades de Ingeniería del NOA, ISBN 978-987-1341-37-5, Vol I, pág 64-70.

Trabajo realizado y presentado en IX Jornadas Regionales de Investigación en Humanidades y Ciencias Sociales. Facultad de Humanidades. UNJu:

- Aula Inteligente: Nuevas Formas de Interacción con la Computadora. Pilar Gálvez, Viviana E. Quincoces. Sandra Méndez. Sergio Rodríguez. Resumen en Cuadernos 34, Facultad de Humanidades y Ciencias Sociales, Suplemento, pág 125. 14-16/05/2008.

4. FORMACION DE RECURSOS HUMANOS

Entre los integrantes del grupo cuatro de ellos están realizando estudios en maestrías y uno en especialización.

Los integrantes de esta línea de investigación participan en el dictado de asignaturas de postgrado de la Universidad Nacional de Jujuy y la Universidad Tecnológica Nacional (Regional Tucumán). Dirección de becarios y tesinas en temas relacionados a la línea de investigación de GRISECO.

5. BIBLIOGRAFIA

Abowd, G.D., Software engineering issues for ubiquitous computing, *Proceedings of the 21st International Conference on Software Engineering*, 75-84, Los Alamitos, CA, USA, 9 IEEE, Computer Society Press, 1999.

<ftp://ftp.cc.gatech.edu/pub/gvu/tr/1999/99-22.pdf>

Chung, E.S; Hong, J.I.; Lin, J; Prabaker, M.K.; Landay, J.A and A.L. Liu. *Development and Evaluation of Emerging Design Patterns for Ubiquitous Computing*. Symposium on Designing Interactive Systems .Proceedings of the 2004 Conference on Designing interactive systems: processes, practices, methods, and techniques, Cambridge, MA, USA, pp: 233 – 242. ISBN:1-58113-787-7, 2004.

Dey, A.K & G.D. Abowd, Towards a better understanding of context and context-awareness, *GVU Technical Report GIT-*

GVU-99-22, College of Computing, Georgia Institute of Technology, 1999.

Dey, A.K., *Providing architectural support for building context aware applications*, PHD Thesis. Georgia Institute Technology, USA, 2001.

<ftp://ftp.cc.gatech.edu/pub/gvu/tr/1999/99-22.pdf>

Distante, D.; Pedone, P.; Rossi, G. and G. Cantora. *Model-Driven Development of Web Applications with UWA, MVC y JavaServer Faces*. In Proceedings of ICWE 2007, Springer Verlag LNCS, 2007.

Fortier, A.; Rossi, G. and S. Gordillo. *Decoupling Design Concerns in Location Aware Services*. In International Conference on Mobile Information Systems, *MOBIS 2005, Dec. 2005*.

Fortier, A.; Cañibano, N.; Grigera, J.; Rossi, G. & S. Gordillo. *An Object-Oriented Approach for Context-Aware Application, (DRAFT) International Smalltalk Conference*, Prague, 2006.

Gordillo, S., G. Rossi & A. Fortier, *Engineering Pervasive Services for Legacy Software, Proceedings of the 1st International Workshop on Software Engineering for Pervasive Services*, Lyon, France, 2006.

Korkea-aho, M., *Context-Aware Applications Survey*. Department of Computer Science. Helsinki University of Technology, 2000.

Lee, C.J.; Bonanni, L.; Espinosa, J:H; Lieberman, H. And Ted Selker. *Augmenting kitchen appliances with a shared context using knowledge about daily events*. Proceedings of the 11th International Conference on Intelligent User Interfaces. Sydney. Australia. Session: Short papers. *ACM Press. New York, NY; USA. Pages: 348 – 350. ISBN:1-59593-287-9. 2006*.

Pascoe, J., N.S Ryan & D.R. Morse, *Issues in Developing Context-Aware Computing, Proceedings of the International Symposium on Handheld and Ubiquitous Computing*, 208-221, Karlsruhe, Germany, Springer-Verlag, 1999.

Schilit, B.N., N.I. Adams & R. Want, Context-Aware Computing Applications, *Proceedings of the Workshop on Mobile Computing Systems and Applications*, 85-90, Santa Cruz, CA, IEEE Computer Society, 1994.

Want, R., A. Hopper, V. Falcao & J. Gibbons, The Active Badge Location System. *ACM Transactions on Information Systems*, 10 (1), 91-102, 1992.

Winograd, T, *Architectures for Context*, Computer Science Department, Stanford University, 2001

<http://hci.stanford.edu/~winograd/papers/context/context.pdf>

Yorio, D. Identificación y Clasificación de Patrones en el Diseño de Aplicaciones Móviles. Tesis de Magíster en Ingeniería de Software. *Facultad de Informática, UNLP*, 2006.

Bases de Datos Métrico-Temporales

Anabella De Battista , Andrés Pascal

Departamento de Sistemas de Información

Universidad Tecnológica Nacional

Fac. Reg. Concepción del Uruguay

Entre Ríos, Argentina

{debattistaa, pascalj}@frcu.utn.edu.ar

Norma Edith Herrera

Departamento de Informática

Universidad Nacional de San Luis

San Luis, Argentina

nherrera@unsl.edu.ar

Gilberto Gutierrez

Facultad de Ciencias Empresariales

Universidad del Bio-Bio

Chillán, Chile

ggutierr@ubiobio.cl

Contexto

El presente trabajo se desarrolla en el ámbito del Grupo de Investigación en Bases de Datos (Proy. Nro 25-D040) perteneciente al Departamento de Sistemas de la Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay, cuyo objetivo principal es el estudio de métodos de acceso, procesamiento de consultas y aplicaciones de bases de datos no tradicionales.

Resumen

Las bases de datos métrico-temporales constituyen un nuevo modelo de bases de datos orientado al procesamiento de consultas por similitud en un intervalo o instante de tiempo. Este modelo está basado en la combinación de espacios métricos con bases de datos temporales. Para resolver eficientemente consultas métrico-temporales, se han propuesto varios índices cuyas evaluaciones empíricas demuestran que son competitivos. En este trabajo estamos interesados en el diseño de índices eficientes para el procesamiento de consultas métricas temporales.

Palabras claves: *Espacios Métricos, Bases de Datos Temporales, Bases de Datos Métrico-Temporales, Índices*

1. INTRODUCCIÓN

Las operaciones de búsquedas en una base de datos requieren de algún soporte y organización especial a nivel físico. En el caso de las bases de datos clásicas, la organización de la información se basa en el concepto de búsqueda exacta sobre datos estructurados. Esto significa que la información se organiza en registros con campos completamente comparables. Una búsqueda en la base retorna todos aquellos registros cuyos campos coinciden con los aportados en la consulta (búsqueda exacta). Otra característica importante de las bases de datos clásicas es que capturan sólo un estado de la realidad modelizada, usualmente el más reciente. Por medio de las transacciones, la base de datos evoluciona de un estado al siguiente descartando el estado previo.

Actualmente las bases de datos han incluido la capacidad de almacenar otros tipos de datos tales como imágenes, sonido, texto, video, datos geométricos, etc. La problemática de almacenamiento y búsqueda en estos tipos de base de datos difiere notablemente de las bases de datos clásicas en tres aspectos: primero los datos no son estructurados, esto significa que es imposible organizarlos en registros y campos, segundo la búsqueda

exacta carece de interés y tercero resulta de interés mantener todos los estados de la base de datos y no sólo el más reciente a fin de poder consultar el instante o intervalo de tiempo de vigencia de dichos objetos. Como solución a esta problemática surgen modelos que permiten procesar esta clase de datos. Entre estos nuevos modelos encontramos los siguientes:

Espacios métricos [1, 2, 6, 8, 9, 10, 5, 17, 12, 13], que permiten almacenar objetos no estructurados y realizar búsquedas por similitud sobre los mismos. Un espacio métrico es un par (U, d) donde U es un universo de objetos y $d : U \times U \rightarrow R^+$ es una función de distancia definida entre los elementos de U que mide la similitud entre ellos. Una de las consultas típicas en este nuevo modelo de bases de datos es la búsqueda por rango, denotado por $(q, r)_d$, que consiste en recuperar los objetos de la base de datos que se encuentren como máximo a distancia r de un elemento q dado.

Bases de datos temporales [16, 11], que incorporan al tiempo como una dimensión, por lo que permiten asociar tiempos a los datos almacenados. Existen tres clases de bases de datos temporales en función de la forma en que manejan el tiempo: *de tiempo transaccional* (*transaction time*), donde el tiempo se registra de acuerdo al orden en que se procesan las transacciones; *de tiempo vigente*, que almacenan el momento en que el hecho ocurrió en la realidad (puede no coincidir con el momento de su registro) y *bitemporales*, que integran la dimensión transaccional y la dimensión vigente a través del versionado de los estados, es decir, cada estado se modifica para actualizar el conocimiento de la realidad pasada, presente o futura, pero esas modificaciones se realizan generando nuevas versiones de los mismos estados.

Bases de datos métrico-temporales [3, 4, 15], que permiten almacenar objetos no estructurados con tiempos de vigencia asociados y realizar consultas por similitud y por tiempo en forma simultánea. Formalmente un *Espacio Métrico-Temporal* es un par (U, d) , donde

$U = O \times N \times N$, y la función d es de la forma $d : O \times O \rightarrow R^+$. Cada elemento $u \in U$ es una triupla (obj, t_i, t_f) , donde obj es un objeto (por ejemplo, una imagen, sonido, cadena, etc) y $[t_i, t_f]$ es el intervalo de vigencia de obj . La función de distancia d , que mide la similitud entre dos objetos, cumple con las propiedades de una métrica (positividad, simetría y desigualdad triangular). Como un ejemplo de aplicación podemos mencionar una base de datos de rostros de delincuentes y cada foto tiene un intervalo de vigencia asociado, que representa el intervalo de tiempo en que el delincuente tenía el aspecto representado en esa foto; en este caso sería de interés, dada una foto y un intervalo de tiempo, poder recuperar de la base todos aquellos rostros parecidos al dado en el intervalo de tiempo especificado. Formalmente una *consulta métrico-temporal* se define como una 4-upla $(q, r, t_{iq}, t_{fq})_d$, tal que $(q, r, t_{iq}, t_{fq})_d = \{o / (o, t_{io}, t_{fo}) \in X \wedge d(q, o) \leq r \wedge (t_{io} \leq t_{fq}) \wedge (t_{iq} \leq t_{fo})\}$

Una forma trivial de resolver una consulta métrico-temporal, sin realizar un barrido secuencial sobre todos los elementos de la bases de datos, es construir un índice métrico agregándole a cada objeto el intervalo de tiempo de vigencia del mismo. Luego, ante una consulta $(q, r, t_{iq}, t_{fq})_d$ primero se utiliza el índice métrico para descartar aquellos objetos obj que están a distancia mayor que r de q ; posteriormente se realiza un barrido secuencial sobre el conjunto de elementos no descartados por el paso anterior a fin de determinar cuáles objetos son realmente respuesta a la consulta, es decir, cuáles tienen un intervalo de vigencia que se superpone con $[t_{iq}, t_{fq}]$.

La desventaja que tiene esta solución trivial es que no se usa la componente temporal para mejorar el filtrado en el índice; en este proceso sólo se aprovecha la componente métrica. Una mejor estrategia es que durante el proceso de búsqueda se utilice tanto la componente métrica como la componente temporal para descartar elementos.

Varios índices métrico-temporales se han propuesto en este ámbito; todos ellos han tomado como base el Fixed Height Queries Tree, un índice para espacios métricos. El Fixed-Height FQT (FHQT) [2] construye un árbol a partir de un ele-

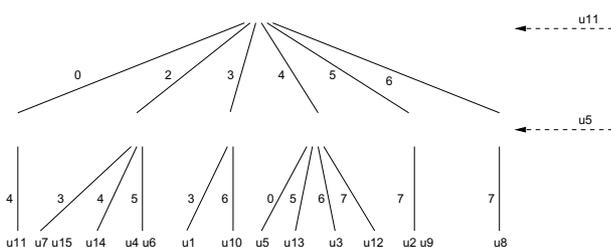


Figura 1: Un ejemplo de un FHQT sobre un conjunto de 15 elementos

mento p (pivote) que puede ser elegido arbitrariamente, o mediante algún procedimiento de selección de pivotes [7], del universo U . Para cada distancia i se crea el conjunto C_i formado por todos aquellos elementos de la base de datos que están a distancia i de p . Luego, para cada C_i no vacío se crea un hijo del nodo correspondiente a p , con rótulo i , y se construye recursivamente un FHQT teniendo en cuenta que todos los subárboles del mismo nivel usarán el mismo pivote como raíz. Este proceso recursivo se continúa hasta lograr que todas las hojas estén en un mismo nivel y tengan menos de b elementos, siendo b un valor fijado previamente. La figura 1 muestra un ejemplo de un FHQT conjunto de 15 elementos en los que se ha elegido u_{11} como pivote en el primer nivel y u_5 como pivote del segundo nivel. Ante una consulta $(q, r)_d$, se comienza por la raíz y se descartan todas aquellas ramas con rótulo i tal que $i \notin [d(p, q) - r, d(p, q) + r]$ siendo p el pivote utilizado en la raíz. La búsqueda continúa recursivamente en todos aquellos subárboles no descartados, utilizando el mismo criterio.

Damos a continuación una breve reseña de los índices métricos-temporales que se basan en el FHQT:

FHQT-Temporal [15]. Este índice es una adaptación del Fixed Height Queries Tree (FHQT) en la que se agrega un intervalo de tiempo en cada nodo del árbol. Este intervalo representa el período máximo de vigencia para todos los objetos del subárbol cuya raíz es dicho nodo. En cada nodo hoja, este intervalo es el período total de vigencia de los objetos que contiene. Para cada nodo interior, el intervalo se calcula tomando el tiempo inicial mínimo, y el tiempo final máximo de sus hijos. Cuando se realiza una consulta

métrico-temporal se procede de la siguiente manera: en cada nivel del árbol se filtran los subárboles hijos por el intervalo de tiempo de la consulta y luego de acuerdo a la distancia entre la consulta y el pivote. Al llegar al último nivel, se realiza una búsqueda secuencial sobre las hojas que no fueron descartadas seleccionando los objetos que cumplen con las condiciones temporales y de similitud.

Historical-FHQT [4]. Consiste en una lista de instantes válidos donde cada uno contiene un FHQT correspondiente a todos los objetos vigentes en dicho instante. Esta estructura es eficiente en bases de datos métrico-temporales en las que los objetos tienen vigencia en un solo instante de tiempo. Los FHQT tienen distintas profundidades en función de la cantidad de elementos que deban indexar. La cantidad de pivotes utilizada en un árbol se calcula como $\lceil \log_2(|o_i|) \rceil$ donde $|o_i|$ es la cantidad de objetos vigentes en el instante i . De esta manera se evita que haya árboles con mayor profundidad de la necesaria, con el fin de que la estructura no tenga un costo excesivo en almacenamiento. Las consultas métrico-temporales se efectúan de la siguiente manera: en primer lugar se seleccionan los instantes incluidos en el intervalo de consulta. Luego se realizan consultas por similitud usando cada uno de los FHQT correspondientes, y finalmente se unen los conjuntos resultantes.

Event-FHQT [14]. Consiste en una lista de intervalos de tiempo válido consecutivos de tamaño fijo. Cada intervalo contiene un FHQT que indexa los objetos vigentes en el primer instante de dicho intervalo. Las hojas del FHQT contienen listas de eventos que indican los cambios que se produjeron entre dos intervalos. Presenta una ventaja respecto del Historical-FHQT, y es que no necesita duplicar los objetos vigentes en más de un instante de tiempo. Ante una consulta métrico-temporal primero se filtran los intervalos de la lista que se intersectan con el intervalo de consulta, luego por cada intervalo se realiza la consulta por similitud sobre el FHQT, se recorren las listas de eventos para determinar que objetos cumplen la restricción temporal de la consulta, por último se unen

los conjuntos resultantes y se compara cada elemento de ese conjunto con la consulta.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Nuestra principal línea de estudio e investigación es el desarrollo de índices métrico-temporales eficientes. El trabajo en curso se puede resumir en los siguientes puntos:

- Se sabe que la dimensionalidad de un espacio métrico afecta el desempeño de los índices [10]. En bases de datos métrico-temporales podría suceder que la dimensión de un conjunto de elementos en el instante i sea distinta a la dimensión del conjunto de elementos en otro instante j y en ese caso las decisiones tomadas con respecto a la construcción del índice deberían variar de un instante a otro. Por esta razón, un aspecto interesante a estudiar es el concepto de dimensionalidad aplicado a bases de datos métrico-temporales con el fin de encontrar una definición que se adecue a este nuevo modelo de bases de datos y que permita comprender mejor el desempeño de los índices.
- En base al punto anterior, se puede diseñar un índice híbrido que permita tener distintos índices métricos en distintos instantes de tiempo, según sea la dimensionalidad del conjunto de elementos almacenados en cada instante.
- Los índices desarrollados hasta el momento se basan en el supuesto de que la memoria principal tiene capacidad suficiente como para mantener tanto el índice como la base de datos. Si esto no es así, la cantidad de accesos a memoria secundaria realizados durante el proceso de búsqueda es un factor crítico en la performance del índice [18]. Nos proponemos explorar técnicas de paginado que sean aplicables a los índices métrico-temporales a fin de lograr que los mismos resulten eficientes también en memoria secundaria.
- Otro aspecto interesante a estudiar es el referido al espacio necesario para mantener el índice, dado que esto decide si el índice se mantendrá en

memoria principal o en memoria secundaria. Una forma de reducir el espacio utilizado es tratar de reutilizar subárboles: si un subárbol del instante i está también en el instante j (con $j > i$), entonces el instante j debería reutilizar el subárbol del instante i en lugar de crearlo de nuevo. Esto implica diseñar un algoritmo que permita detectar subárboles isomorfos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se espera contar con índice eficiente métrico-temporal en memoria secundaria que sea eficiente tanto en los tiempos de respuesta como en el espacio ocupado por el mismo.

4. FORMACIÓN DE RECURSOS HUMANOS

El trabajo desarrollado hasta el momento forma parte del desarrollo de dos Tesis de Maestría en Ciencias de la Computación, una de ellas fue defendida y aprobada en marzo del corriente año. Se cuenta con el asesoramiento del Dr. Gilberto Gutiérrez, de la Universidad del Bio Bio, Chile. El grupo cuenta además con dos alumnos becarios que se están iniciando en las temáticas desarrolladas por el grupo.

REFERENCIAS

- [1] R. Baeza-Yates. Searching: an algorithmic tour. In A. Kent and J. Williams, editors, *Encyclopedia of Computer Science and Technology*, volume 37, pages 331–359. Marcel Dekker Inc., 1997.
- [2] R. Baeza-Yates, W. Cunto, U. Manber, and S. Wu. Proximity matching using fixed-queries trees. In *Proc. 5th Combinatorial Pattern Matching (CPM'94)*, LNCS 807, pages 198–212, 1994.
- [3] De Battista, A. Pascal, G. Gutierrez, and N. Herrera. Búsqueda en bases de datos métricas-temporales. In *Actas del VIII Workshop de Investigadores en Ciencias de*

- la Computación*, Buenos Aires, Argentina, 2006.
- [4] De Battista, A. Pascal, G. Gutierrez, and N. Herrera. Un nuevo índice métrico-temporal: el historical fhqt. In *Actas del XIII Congreso Argentino de Ciencias de la Computación*, Corrientes, Argentina, 2007.
- [5] S. Brin. Near neighbor search in large metric spaces. In *Proc. 21st Conference on Very Large Databases (VLDB'95)*, pages 574–584, 1995.
- [6] W. Burkhard and R. Keller. Some approaches to best-match file searching. *Comm. of the ACM*, 16(4):230–236, 1973.
- [7] B. Bustos, G. Navarro, and E. Chávez. Pivot selection techniques for proximity searching in metric spaces. In *Proc. of the XXI Conference of the Chilean Computer Science Society (SCCC'01)*, pages 33–40. IEEE CS Press, 2001.
- [8] E. Chávez and K. Figueroa. Faster proximity searching in metric data. In *Proceedings of MICAI 2004. LNCS 2972*, Springer, Cd. de México, México, 2004.
- [9] E. Chávez, J. Marroquín, and G. Navarro. Fixed queries array: A fast and economical data structure for proximity searching. *Multimedia Tools and Applications (MTAP)*, 14(2):113–135, 2001.
- [10] E. Chávez, G. Navarro, R. Baeza-Yates, and J.L. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.
- [11] C. S. Jensen. A consensus glossary of temporal database concepts. *ACM SIGMOD Record*, 23(1):52–54, 1994.
- [12] I. Kalantari and G. McDonald. A data structure and an algorithm for the nearest point problem. *IEEE Transactions on Software Engineering*, 9(5):631–634, 1983.
- [13] G. Navarro. Searching in metric spaces by spatial approximation. In *Proc. String Processing and Information Retrieval (SPIRE'99)*, pages 141–148. IEEE CS Press, 1999.
- [14] A. Pascal, A. De Battista, G. Gutierrez, and N. Herrera. Índice métrico-temporal event-fhqt. In *Actas del XIII Congreso Argentino de Ciencias de la Computación*, La Rioja, Argentina, 2008.
- [15] A. Pascal, De Battista, G. Gutierrez, and N. Herrera. Procesamiento de consultas métrico-temporales. In *XXIII Conferencia Latinoamericana de Informática*, pages 133–144, San José de Costa Rica, 2007.
- [16] B. Salzberg and V. J. Tsotras. A comparison of access methods for temporal data. *ACM Computing Surveys*, 31(2), 1999.
- [17] J. Uhlmann. Satisfying general proximity/similarity queries with metric trees. *Information Processing Letters*, 40:175–179, 1991.
- [18] J. Vitter. External memory algorithms and data structures: Dealing with massive data. *ACM Computing Surveys*, 33(2):209–271, 2001.

Participación de los usuarios en el proceso de desarrollo de software

Marcelo G. Estayno/Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales/Universidad de Morón

Marisa D. Panizzi/Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales/Universidad de Morón

Resumen: El Diseño Participativo en los procesos de desarrollo de software es un método que permite a probar utilizar los conocimientos de los usuarios. Consideramos que los usuarios tienen mucho que decir en el desarrollo de un proceso software. La participación de los usuarios contribuye a la satisfacción de ellos mismos por considerarse constructores del producto que utilizarán de ahora en más el cual debe satisfacer sus necesidades. Esta forma de trabajo requiere que para una participación activa de los usuarios, los mismos deben estar no solo muy bien informados sino que hay que hacer dos procesos los que en la jerga se llaman evangelización y alfabetización. El presente método de trabajo incluye la participación de los usuarios no solo en las etapas de ingeniería de requerimientos y de implementación sino que estos no solo se comprometan sino que también se involucren a lo largo de todo el desarrollo de software con lo cual las discrepancias que aparecen en tiempo de implementación no solo se eliminan sino que también esta etapa disminuye en su tiempo de ejecución.

Palabras clave: diseño participativo/participación de usuarios/satisfacción usuario/etapa de implementación/ principio socio-técnico/desarrollo de software.

Introducción:

Comprensión de las necesidades de información hacia una clara percepción del usuario.

La clave para la comprensión del comportamiento del usuario y sus percepciones radica en conocer las causas de su conducta por medio de la observación de su personalidad. Un aspecto que se debe considerar con respecto al usuario es que su actividad general responde siempre a varias necesidades que, por consiguiente, son estimuladas por varios motivos. Sin embargo, en una actividad concreta se puede distinguir una causa, que imprime a la conducta un sentido determinado (González Suárez, 2000

citado por Núñez Paula, 2004). Las necesidades surgen entonces, en el ámbito de varios contextos que interactúan de tal forma que resultan ser determinantes en cada uno de los motivos rectores que regulan la conducta del usuario, es decir, contextos sociales, culturales, económicos, políticos, etcétera, que conducen al usuario a encontrar nuevos motivos que satisfagan sus principales intereses.

Esta idea constituye, como se verá más adelante, gran parte de la esencia cualitativa y razón de ser del Diseño Participativo de Sistemas de Información. En este sentido, Enrique González Suárez (1995 citado por Núñez Paula, 2004) expresa textualmente:

El objetivo de las reuniones informativas consiste en crear el motivo de la actividad informativa de los participantes, principalmente con el auxilio de la dinámica de grupo; y mostrarles la posibilidad de encontrar el objetivo de la actividad informativa [necesidad de información y satisfacción de la misma].

El funcionamiento de este diseño se materializa en la interrelación y la comunicación usuario-sistema informativo. El motivo informativo se crea al relacionar la información que se disemina en las reuniones informativas con las necesidades de Información del grupo según el tema en que está trabajando el grupo.

De esta forma, la información adquiere una función impulsora y rectora de la actividad científico-informativa del usuario.

La satisfacción es un resultado subjetivo que el sistema desea alcanzar, y busca que dependa tanto del servicio prestado, como de los valores y percepciones del propio usuario (Rey Martín, 2000, 141).

El diseño participativo de servicios de información:

“el usuario está en el sistema”

La participación de los usuarios en el proceso de desarrollo de software tiene el objetivo de

acercar a los profesionales de la información con las demandas de sus usuarios; esta propuesta comprende un enfoque organizacional que implica necesariamente un adoctrinamiento socio-técnico, entiéndase la parte social; entiéndase la parte técnica; entiéndanse aquellas herramientas que fueron creadas para el uso y beneficio de las personas. La parte social es la de los participantes; la parte técnica es la de los sistemas simples y/o complejos. Enid Mumford adoptó los principios del enfoque socio-técnico y desarrolló en 1983 ETHICS, un método participativo en el que los usuarios participan en la creación y desarrollo de un sistema de información conjuntamente con los principales actores de la unidad, directores o analistas de la información (González Solano, 2008). Ningún sistema de información puede ser creado a expensas del usuario: *el usuario está en el sistema*.

Esta investigación junto con otros proyectos, conducirían a proponer principios útiles y relevantes utilizando ideas que se han desarrollado y publicado sobre:

- El concepto de sistemas socio-técnicos
- Definición de organizaciones como sistemas abiertos
- La principal opción organizacional
- La necesidad de optimizar y reconciliar los sistemas sociales y técnicos
- La importancia de grupos auto-directivos
- Los problemas de enajenación del trabajo

El diseño socio-técnico vino a ser asociado entonces con un principio ético que estaba llamado a aumentar la capacidad del individuo en el control del medio ambiente laboral inmediato. Aconsejaron que los gerentes dijeran a sus grupos de trabajo que hacer, pero no como hacerlo. A partir del Principio Socio-Técnico se extienden diseños que buscan llevar a cabo su implementación, siendo la *participación* el modelo más viable.

¿Qué es la participación? ¿Quiénes son los participantes? y ¿En que participan?

Puesto que la participación de usuarios en el diseño de sistemas de información es un aspecto relevante, resulta necesario aclarar lo que significa la participación y lo que ella implica. Una cosa es segura: la participación no es un concepto nuevo en el extenso plano de las ciencias de la información. En el entendido

de que toda unidad de información trabaja en función del beneficio social; el participante más importante de ella es, por lo tanto, el usuario. De esta manera, se puede definir a la participación como un proceso en el cual dos o más partes involucradas influyen sus decisiones, planes y políticas en torno a los sistemas, servicios y recursos que ofrece una determinada unidad. Sin embargo, debemos reconocer que es poco el debate que se hace acerca de las decisiones tomadas, es decir, de los distintos intereses que influyen en ellas. Si deseamos tener un acercamiento más amplio en el diseño de sistemas es necesario considerar las funciones, la estructura, los intereses y los problemas mismos que influyen en los participantes implicados.

Dentro de todo sistema, existen varios tipos de participantes que necesariamente intervienen en los sistemas que ellos mismos implementan para agilizar la recuperación de sus recursos. Todo ello, en respuesta a un primer planteamiento: ¿Cómo brindar la información correcta al usuario? La estructura operativa de toda unidad de información debe ser determinada y evaluada antes de integrar al usuario en su esquema de diseño de sistemas. Por esta razón, conviene hacer un breve análisis Pre Diseño Participativo. De acuerdo a Jackson y Keys (Jackson & Keys), no todos los problemas planteados son creados de la misma manera, por lo que los diferentes problemas deberían ser abordados, entonces, con diferentes métodos de análisis. Estos dos autores desarrollaron un sistema de clasificación llamado "System Of Systems Methodologies", que adecuándolo al ámbito de las ciencias de la información, consideran dos partes primordiales dentro de cada unidad de información: los sistemas de información y los participantes de ellos. SOSM basa su enfoque en la identificación de los participantes internos de cada unidad de información y su forma de interactuar entre ellos mismos y los sistemas. La descripción de este sistema de clasificación es como sigue: existen dos tipos de sistemas, los simples y los complejos. Los sistemas simples tienen la característica de tener pocos componentes o subsistemas, los cuales interactúan poco y de forma básica. Por lo general no cambian mucho con el paso del tiempo al no verse afectados por las acciones independientes de cada componente o por el medio ambiente. Los sistemas complejos, por el contrario, se caracterizan por tener un gran número de subsistemas que interactúan entre sí. Su estructura es tan compleja e inconstante que tienen que adaptarse y evolucionar en un medio ambiente turbulento. Así mismo, también

existen tres tipos de participantes, esto es, SOSM clasifica las relaciones entre las personas que tienen algo que ver con el sistema de información: participantes unitarios, pluralistas y coercitivos. La relación de tipo unitaria implica que sus participantes poseen valores, creencias e intereses idénticos; ellos comparten el mismo propósito. En las relaciones pluralistas, sus participantes comparten los mismos intereses, no así sus valores o creencias, lo que debe promover un espacio de intercambio de ideas, acuerdos e incluso conflictos. En total contraste, en las

relaciones coercitivas existen pocos o nulos intereses en común, la libertad de expresión es muy baja y los valores y creencias de cada individuo resultan ser muy conflictivas. Las decisiones son tomadas mediante enfoques absolutistas de gente con poder. Así, de la combinación de los tipos de sistemas y de los tipos de participantes se producen 6 contextos teóricos: simples-unitarios; simples-pluralistas; simples-coercitivos; complejos-unitarios; complejos-pluralistas y complejos-coercitivos. Esto se puede representar mediante el siguiente cuadro:

<i>Tipos de Sistemas</i>	Sistema Simple	Sistema Complejo
<i>Tipos de Participantes</i>		
Unitarios	Simple unitario	Complejo unitario
Pluralistas	Simple pluralista	Complejo pluralista
Coercitivos	Simple coercitivo	Complejo coercitivo

Líneas de investigación y Desarrollo. Problemas en el uso del diseño participativo.

Sin embargo, cuando un nuevo sistema de información es visto como parte fundamental de un plan de la organización para su porvenir, alcanzar un consenso en todos los niveles de la organización es un acto muy complicado. La mayoría de las unidades de información requieren solo de la negociación entre la dirección general y académicos importantes que poco hacen por asegurar la participación de sus empleados. Incluso existen esquemas axiológicos que consideran a la participación en todos sus niveles como un acto indócil y de proceso lento, indeseable por diversos motivos. Muchas veces los comités consultivos están diseñados con el fin de comprobar que la visión que tiene la dirección de la unidad de información no está en conflicto con los intereses de sus empleados, buscando asegurar de alguna manera que ciertas estrategias particulares reciban el apoyo de la mayoría.

Resultados Obtenidos/ Esperados.

Ningún sistema de información puede ser creado a expensas del usuario: *el usuario está en el sistema.* Esta investigación, conducirá a proponer recomendaciones practicas que conduzcan a la participación del usuario a lo largo de todo el proceso software. En este primer trabajo abordaremos la etapa de diseño pero intentaremos en futuros proyectos involucrar a

todos los stakeholders a lo largo de todo el proceso software.

Formación de Recursos Humanos.

Nuestra primera transferencia a la docencia, creemos que es fundamental que los docentes investigadores hagan su primera transferencia a los estudiantes. A esto lo consideramos la primera acción de RRHH. Una vez que el proyecto esté un poco más avanzado se incorporaran estudiantes al mismo como noveles investigadores.

Bibliografía.

- S.L.Pfleeger, Ingeniería de Software. Teoría y práctica. Ed. Prentice Hall, Bs.As. 2002.
- R. Pressman, Ingeniería del Software, Un enfoque práctico. 4ta edición. McGraw-Hill, 1998.
- I. Sommerville, Ingeniería de Software, Editorial Addison Wesley.
- Pensamiento de sistemas, practica de sistemas. Peter Checkland. Megabyte. Grupo Noriega Editores.1993.
- Exploring the Importance of Participation in the Post-Implementation Period of ES Project: A Neglected Area. Erica L. Wagner/Sue Newell. Journal of the Association for Information Systems. Volume 8, Issue 10 Article 10, pp.508-524, October 2007.

Calva González, Juan José (2004). Las necesidades de información: fundamentos teóricos y métodos. México: Universidad Nacional Autónoma de México, Centro Universitario de Investigación Bibliotecológica.

Cherns, Albert (1976). The principles of sociotechnical design. *Human Relations*, 29 (8): 783-792.

Eiglier, Pierre y Langeard, Éric (1992). *Servucción (El marketing de servicios)*. Madrid: MacGraw-Hill.

González Solano, José Alberto y González Nolasco, Christian Alejandro (2008). Prototipo A-1/Greenstone08 (1ra evolución). En *E-prints in Library and Information Science*. Disponible en:
<http://eprints.rclis.org/archive/00014091/>

González Suárez, Enrique (1995). "La actividad grupal con los usuarios; una vía para resolver el desbalance entre la producción y el consumo de información". *Revista Española de Documentación Científica*, 18 (4): 405-415.

González Suárez, Enrique (2000). "Sistema de factores en la conducta del usuario". *Revista Iberoamericana sobre usuarios de información*, (9): 6-17.

Jackson, M. C., y Keys. P. (1984). "Towards a system of system methodologies". *J. Oper. Res. Soc.*, 35 (6): 473-486.

Mumford, Enid (1983). *Designing human systems, the ETHICS approach* Manchester Business School. Manchester.

Mumford, Enid (1983). *Designing participatively: a participative approach to computer systems design*. Manchester: Manchester Business School.

Mumford, Enid (2000). *Socio-technical design: an unfulfilled promise or a future opportunity*. En Baskerville, R., Stage, J., y DeGross, J. I., *Organizational and social perspectives on information technology* (pp. 33-46). Boston: Kluwer Academic Publications.

Núñez Paula, C. Israel (2004). "Las necesidades de información y formación: perspectiva socio-psicológica e informacional". *ACIMED*, septiembre - octubre, 12 (5): 1-1.

Rey Martín, Carina (2000). "La satisfacción del usuario: un concepto en alza". *Anales de Documentación*, (3): 139-152.

Desarrollo de herramientas de evaluación para los procesos de gestión de PYMEs de software

Alicia Mon¹
Andrea Arancio¹
Eduardo De Maria¹
Graciela Romanelli¹
Marcelo Estayno²

¹ Grupo de Ingeniería de Software (G.I.S.) /Dto. de Ingeniería/Universidad Nacional de La Matanza
² Grupo de Ingeniería de Software (G.I.S.) /Fac. de Ingeniería/Universidad Nacional de Lomas de Zamora



CONTEXTO

El Grupo de Ingeniería de Software “G.I.S.” tiene varias líneas de trabajo en el área de calidad de software y enmarca estos trabajos en base a estudios realizados sobre la evolución del desarrollo de software a nivel internacional detectando que el mismo ha estado marcado en los últimos años por la tendencia en la consecución de productos de calidad a partir del desarrollo de procesos maduros y estables que permitan garantizar niveles de calidad competitivos.

En este sentido, una de las líneas de trabajo del grupo es que uno de los objetivos centrales del desarrollo de software en Argentina resida en la búsqueda de los niveles de calidad adecuados, de modo tal que permita desarrollar productos con calidad y posicionarse en el mercado nacional e internacional con ventajas competitivas.

Si bien existen diferentes Modelos de Proceso Software y Normas de Calidad con los que las empresas de desarrollo pueden certificar sus procesos, (ISO 90003, ISO15504 o CMMI) [2], [3], [4], por lo general resultan complejos, difícil de comprender, aplicar, y su implementación es altamente costosa, para las PyMES de la industria del software, ya que han sido elaborados para organizaciones grandes, tal es el caso de CMMI o para la gestión en general, como resulta la Norma ISO 9001 que al no ser específica para la industria del software, exige un esfuerzo adicional de interpretación. Por esta razón, las pequeñas empresas se encuentran con serias dificultades para adecuar las actividades en un proceso de mejora continua y certificación [5].

RESUMEN

En la actualidad, se encuentra finalizado el proyecto Competisoft [6], desarrollado en el marco del Programa CyTED [1] el cual ha generado un Modelo de Referencia de Procesos y un Modelo de Mejora de Procesos adaptado a las necesidades de calidad de las PyMES iberoamericanas, generando un modelo adaptable, comprensible y de costos accesibles en su implementación [12].

El presente artículo, expone la línea de investigación del grupo GIS en lo que implica la continuación de un proyecto de investigación que ha generado como resultado el desarrollo y la aplicación de un Instrumento de Diagnóstico basado en el Modelo Competisoft para realizar una evaluación inicial en pequeñas y medianas organizaciones de la industria del software [14], [15], [16], [17], [18]. Las prácticas del modelo están sintetizadas en un conjunto de procesos que abarcan las responsabilidades asociadas con la organización: Alta Dirección, Gerencia y Operación.

El instrumento de diagnóstico está basado en el conjunto de procesos de Operación [9], [10] y se pretende, por medio de esta línea de investigación, abordar la evaluación de los procesos de Gerencia y de Alta Dirección.

Palabras clave: Modelos de Madurez. PYMEs. Modelos de Calidad. Desarrollo de software

1. INTRODUCCION

El Modelo de referencia de Procesos Competisoft [7] tiene una estructura de 3 (tres) categorías que abarcan las responsabilidades asociadas con la organización: Alta Dirección, Gerencia y Operación. Dentro de cada categoría se definen un conjunto de procesos que incluyen prácticas y roles específicos [8], [11], [13].

La categoría 1- Alta Dirección, aborda las prácticas relacionadas con la Gestión del Negocio, proporciona los lineamientos a los procesos de la categoría de Gerencia y se retroalimenta con la información generada por ellos.

La categoría 2-Gerencia, aborda las prácticas de Gestión de Procesos, Proyectos y Recursos en función de los lineamientos establecidos en el nivel de Alta Dirección y proporciona los elementos para el funcionamiento de los procesos de la categoría de Operación, recibe y evalúa la información generada por éstos y comunica los resultados a la Alta Dirección.

La categoría 3- Operación se subdivide en tres procesos, Administración de Proyectos Específicos, Desarrollo de software y Mantenimiento de software. El primer proceso busca establecer y llevar a cabo sistemáticamente las actividades que permitan cumplir con los objetivos de un proyecto en tiempo y costo esperados. El segundo y tercero, apuntan a la realización sistemática de las actividades de análisis, diseño, construcción, integración y pruebas de productos de software nuevos o en etapas de mantenimiento, cumpliendo con los requerimientos especificados. En este nivel se realizan las actividades de acuerdo a los elementos proporcionados por el nivel de Gerencia y entrega a ésta la información y productos generados.

Estas tres categorías se encuentran relacionadas entre sí a través de los diferentes procesos y los productos de entrada que cada uno requiere y de salida que cada uno genera.

La siguiente Figura 1 presenta el esquema del conjunto de procesos por categoría que define el modelo Competisoft:

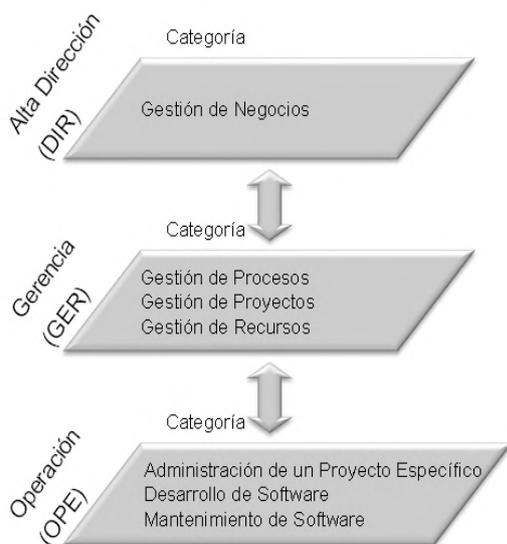


Figura 1. Estructura del Modelo Competisoft

2. LINEAS DE INVESTIGACION y DESARROLLO

El Instrumento de Diagnóstico ha sido desarrollado para realizar un análisis inicial de la madurez en una organización, en una instancia del Proceso de Mejora. Este instrumento, contiene un conjunto de cuestionarios orientado a los Procesos de Operación del Modelo Competisoft, que incluye el Proceso de Administración de Proyectos Específicos (OPE1) y el Proceso de Desarrollo y Mantenimiento de Software (OPE2) y un método definido para la implementación de los cuestionarios, de manera tal que permita realizar una evaluación inicial sobre las prácticas realizadas por una organización y el inicio de un ciclo de mejora. La implementación de esta herramienta permite realizar un diagnóstico inicial, para poder instalar un Proceso de Mejora, comenzando con la definición de la estructura de Roles de una organización, los Productos de trabajo que utiliza en el momento del diagnóstico y las herramientas que dispone como soporte a la gestión y al desarrollo.

El Método es iterativo incremental y debe aplicarse desde dos aspectos, uno cuantitativo y otro cualitativo. Comienza por la realización del análisis cuantitativo, luego el análisis cualitativo y en base al resultado del mismo propone Recomendaciones, aplica un ciclo de mejora y vuelve a iniciar el ciclo con el siguiente incremento de análisis. El método finaliza cuando se llega a cumplir el nivel de madurez deseado por la empresa que encara este proceso de mejora.

Como hasta ahora se han abordado las actividades de Administración de Proyectos Específicos y las actividades de Desarrollo y de Mantenimiento de software. A partir de ahora se abordarán las prácticas de la categoría de Gerencia que incluyen la Gestión de Procesos, Gestión de Proyectos y Gestión de Recursos en función de los lineamientos establecidos en el nivel de Alta Dirección.

3. RESULTADOS ESPERADOS

Como resultado de esta línea de investigación se obtendrá un Modelo de Evaluación aplicable a todo el Modelo Competisoft tomado como Modelo de Referencia. El desarrollo del mismo se realizará a través del diseño de un conjunto de cuestionarios y un método definido para la implementación de estos cuestionarios, de manera tal que permita realizar un diagnóstico inicial sobre el conjunto de prácticas que propone el Modelo de Referencia y que sean implementadas por una organización. Esto implica continuar con la investigación en los resultados obtenidos hasta el

APLICATIVO DE SOPORTE PARA LA VALORACIÓN DE MODELOS Y ESTÁNDARES DE CALIDAD PARA EL PROCESO DE SOFTWARE.

Autores: Raúl Omar Moralejo^{1,2}, Nerina Claudia Dumit Muñoz¹, Gustavo Mercado¹, Carlos Taffernaberry¹, Santiago Perez¹, Salvador Navarría², Franco Catena¹, Verena Vicencio¹, Andrea Micca Longo¹, María Jáuregui¹, Jessica Vidal¹

¹GRIDTICs (Grupo de Investigación y Desarrollo en Tecnologías de Información y las Comunicaciones). Universidad Tecnológica Nacional - Facultad Regional Mendoza. Rodríguez 273, Ciudad de Mendoza (5500), Argentina, Teléfono: 5244500, Fax: 5244531

{rmoralejo, ndumit, gmercado, carlos_taffe, santiagocp}@frm.utn.edu.ar; {francocatena, verevere}@gmail.com; andrea.miccalongo@gridtics.frm.utn.edu.ar; {jauregui_maria, jessicavidalcorrea}@yahoo.com.ar

²ICTI (Instituto de Calidad en Tecnologías de la Información). Universidad de Mendoza - Facultad de Ingeniería. Peatonal Emilio Descotte 750, Ciudad de Mendoza (5500), Argentina, Teléfono: 4201872. Fax: 4201100 {raul.moralejo, salvador.navarría}@um.edu.ar.

CONTEXTO

La línea de investigación es modelos y estándares de evaluación y mejora del proceso de software, la misma se desarrolla en el GRIDTICs (Grupo de Investigación y Desarrollo en Tecnologías de Información y las Comunicaciones) - Universidad Tecnológica Nacional - Facultad Regional Mendoza y en el ICTI (Instituto de Calidad en Tecnologías de la Información) - Universidad de Mendoza - Facultad de Ingeniería, en el marco del Proyecto Valoración de Modelos y Estándares de Evaluación y Mejora del Proceso de Software (MyEEMPS).

RESUMEN

En los últimos años han aparecido un gran número de estándares y propuestas internacionales y regionales relacionadas con mejora de procesos software (SPI) para micro y pequeñas empresas software (VSEs). Esto evidencia el creciente interés en la comunidad de Ingeniería de Software en abordar el tema de SPI para VSEs. Este creciente interés se suscita porque la industria del software en la mayoría de los países está formada en gran parte por este tipo de empresas. Es importante entonces fortalecer a las VSEs con prácticas y guías eficientes de Ingeniería de Software adaptadas a su tamaño y tipo de negocio. Así pues, para orientar a las VSEs en relación a la elección de modelos y estándares de calidad en el proceso de software disponibles cuando comiencen un proyecto de mejora de procesos, se presenta un software de soporte para la aplicación del instrumento VAMEPS que permite la valoración de los conocimientos sobre modelos y estándares de calidad en el proceso de software, y en base a estos resultados decidir cual podría ser el más conveniente para la utilización en la organización.

Palabras Claves

Valoración, arquitectura, método de evaluación, mejora del proceso, herramientas, modelos y estándares.

1. INTRODUCCIÓN

A principios del siglo XXI la comunidad de Ingeniería del Software (industria e investigadores) ha expresado especial interés en la mejora de procesos software (conocida por las siglas SPI, del término en inglés Software Process Improvement) en micro y pequeñas empresas desarrolladoras de software con menos de 50 empleados (a las cuales se hace referencia en el presente artículo por la sigla VSEs del término en inglés Very Small Software Enterprises). Esto se evidencia por el creciente número de artículos que tratan el tema según el análisis de la tendencia de las publicaciones de mejora de procesos en micro, pequeñas y medianas empresas, así como por la

aparición de un gran número de estándares y propuestas internacionales y regionales relacionadas con SPI para VSEs.

Este creciente interés acerca de SPI en VSEs, se suscita porque este tipo de empresas son una pieza muy importante en el engranaje de la economía mundial, además la industria del software en la mayoría de los países está formada por un tejido industrial compuesto en gran parte por VSEs. En Europa el 85% de las compañías del sector de las tecnologías de la información son muy pequeñas, entre 1 y 10 empleados. En Iberoamérica el 75% de las empresas software tienen menos de 50 empleados. Además aproximadamente el 94% de empresas que desarrollan software son pequeñas.

Este tipo de organizaciones desarrollan productos significativos que, para su construcción, necesitan estrategias, prácticas y/o guías eficientes de ingeniería del software adaptadas a su tamaño y tipo de negocio. Siguiendo esta premisa es importante entonces ayudar a estas organizaciones a entender los estándares/modelos y propuestas internacionales, tales como Bootstrap, ISO 9001, ISO/IEC 15504 o CMMI, y regionales relacionadas con SPI para VSEs, tales como Moprosoft.

El objetivo es ofrecer una estrategia, a través del uso del instrumento VAMEPS y luego del aplicativo, que facilite a las VSEs la elección de aquellos modelos y estándares más convenientes.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

La línea de investigación y desarrollo se encuentra en el marco del proyecto valoración de modelos y estándares de evaluación y mejora del proceso de software (MyEEMPS).

El presente trabajo forma parte de un proyecto mayor que está siendo ejecutado en forma conjunta por la Universidad Tecnológica Nacional – Facultad Regional Mendoza y la Universidad de Mendoza – Facultad de Ingeniería, y tiene el aval de las mismas Universidades según el siguiente detalle: Resolución 7/06 del Departamento de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional – Facultad Regional Mendoza. Resolución 06PI/06 del Departamento de Electrónica de la Universidad Tecnológica Nacional – Facultad Regional Mendoza. Resolución 14/06 del Consejo Académico de la Facultad de Ingeniería – Universidad de Mendoza. Resolución N° 416/2007 del Consejo Académico de la Universidad Tecnológica Nacional – Facultad Regional Mendoza.

3. RESULTADOS OBTENIDOS

Se ha iniciado el desarrollo del aplicativo que permite la carga y procesamiento de datos relevados con el instrumento VAMEPS. En forma general las principales funcionalidades son: Abrir un Proyecto existente o Crear un Nuevo Proyecto. Por Proyecto definir: Descripción, Año. Agregar datos demográficos al instrumento: Apellido y Nombre, Universidad, Edad, Sexo (M / F), Trabaja (Si / No). ABM de Modelos y Estándares, de Aspectos, ABM de Valoraciones de aspectos, ABM de Valoración Global. Emisión de Cuestionarios en función de los parámetros. Carga de datos de los cuestionarios en función de los parámetros. Carga de datos “interactiva” de los cuestionarios en función de los parámetros. Validaciones y controles sobre la carga de datos. Informes del sistema.

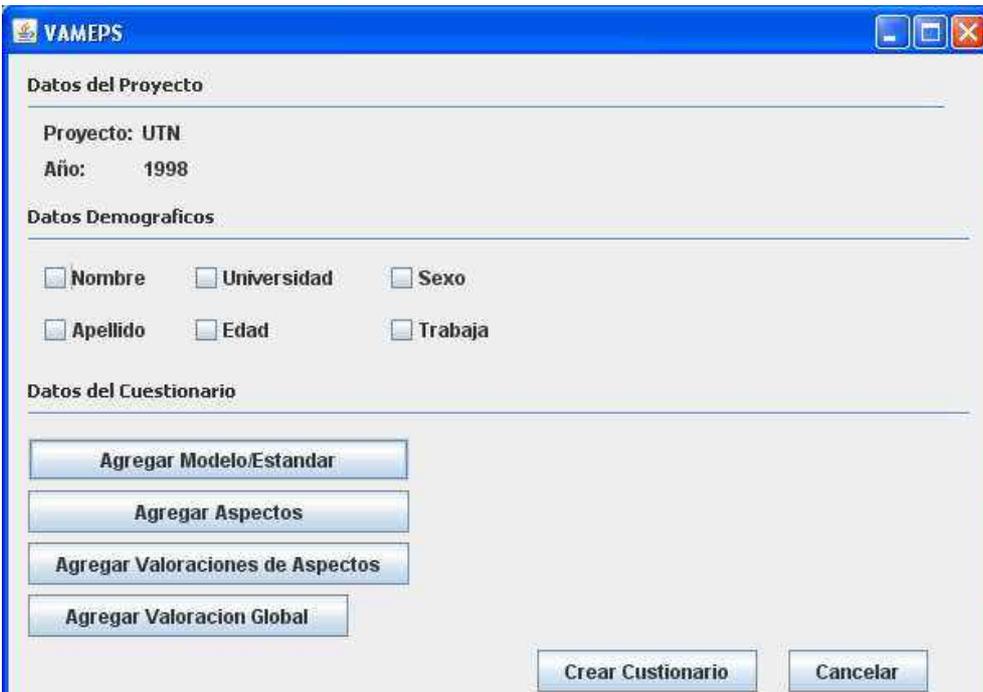
A continuación se muestran algunas de las pantallas del aplicativo:



Cuando hacemos clic en “Nuevo” se abre otra pantalla donde podemos colocar los datos del proyecto como: nombre, año, y qué tipo de proyecto.



Quando hacemos clic en “Aceptar” de la pantalla anterior, vemos a continuación las características del proyecto:



Los botones para “agregar” abren pantallas para poder ingresar nuevos modelos, así como también nuevos aspectos y valoraciones.



Quando hacemos clic en “Crear Cuestionario” aparece esta pantalla, donde podemos ver los modelos que existen para poder cargarlos al proyecto. Una vez terminada la selección se crea el cuestionario y aparece el archivo en formato pdf, listo para imprimir o guardar.

VAMEPS

Datos del Proyecto
 Proyecto: UTN
 Año: 1998

Datos Demograficos
 Nombre Universidad Sexo
 Apellido Edad Trabaja

Modelos y Estandares

DISPONIBLES
 BOOTSTRAP
 ISO/IEC 15504
 CMMI-SW

SELECCIONADOS
 BOOTSTRAP
 CMMI-SW
 ISO/IEC 15504

Nombre:
 Agregar Modificar Eliminar

Aspectos

DISPONIBLES
 Esta de acuerdo con la Arquitectura y/o Estandares
 Esta de acuerdo con el Método de Evaluación
 Esta de acuerdo con el Proceso de Mejora

SELECCIONADOS
 Esta de acuerdo con el Método de Evaluación
 Esta de acuerdo con el Proceso de Mejora
 Esta de acuerdo con la Arquitectura y/o Estandares

Nombre:
 Agregar Modificar Eliminar

Valoraciones

DISPONIBLES
 1 - Muy de Acuerdo
 2 - En Desacuerdo
 3 - Ni Acuerdo Ni Desacuerdo

SELECCIONADOS
 1 - Muy de Acuerdo
 2 - En Desacuerdo
 3 - Ni Acuerdo Ni Desacuerdo

Nombre:
 Agregar Modificar Eliminar

Crear Cuestionario Cancelar

Creacion Encuesta

 Se Creo correctamente!

Aceptar

JasperViewer

ENTREVISTA - VALORACION INDIVIDUAL DE MODELOS Y ESTANDARES DE EVALUACION Y MEJORA DEL PROCESO DE SOFTWARE

Datos del Proyecto
 Nombre: UTN
 Año: 1998

Datos Personales
 Nombre:

BOOTSTRAP

	1 - Muy de Acuerdo	2 - En Desacuerdo	3 - Ni Acuerdo Ni Desacuerdo	4 - De Acuerdo
Esta de acuerdo con el Método de Evaluación				
Esta de acuerdo con el Proceso de Mejora				
Esta de acuerdo con la Arquitectura y/o Estandares				

Valoracion Global

Página 1 de 3

4. FORMACIÓN DE RECURSOS HUMANOS

Durante la ejecución del proyecto se han formado profesores, graduados y alumnos, y se han iniciado dos maestrías relacionadas con la temática.

Formación de profesores: Ing. Nerina Dumit – UTN Regional Mendoza.

Formación de graduados y estudiantes: Ing. Verena Vicencio – UTN Regional Mendoza; Franco Catena – UTN Regional Mendoza; Andrea Micca Longo – UTN Regional Mendoza; María Jáuregui – UTN Regional Mendoza; Jessica Vidal – UTN Regional Mendoza.

Maestrías relacionadas con la temática CTS:

Maestría en Administración de Negocios (MBA) - Universidad Tecnológica Nacional Facultad Regional Mendoza (Ing. Nerina Dumit).

Maestría en Ingeniería en Calidad – Universidad Tecnológica Nacional Facultad Regional Mendoza (Ing. Verena Vicencio).

5. BIBLIOGRAFÍA

BOOTSTRAP (Software Engineering Body of Knowledge). 2005. <http://www.bootstrap.org>

CMMI (Capability Maturity Model Integration). 2005. <http://www.sei.cmu.edu/cmml/>

ESI. Europe Software Institute. 2007, <http://www.esi.es/en/main/iitmark.html>.

Fayad, M.E., M. Laitinen, and R.P. Ward, Software Engineering in the Small. Communications of the ACM, 2000. Vol. 43(3) March pp. 115-118.

ISO (Organización Internacional para la Estandarización), <http://www.iso.org>; IEC (Comisión Electrónica Internacional). 2005. <http://www.iec.ch>

Mayer&Bunge. Panorama de la Industria del Software en Latinoamérica. Mayer & Bunge Informática LTDA. Brasil. 2004, www.mbi.com.br/200409_panorama_industria_software_america_latina.pdf

MoProSoft (Modelo de Procesos para la Industria de Software – México). 2005. <http://www.software.net.mx/>

Pino, F., F. Garcia, and M. Piattini, Revisión sistemática de mejora de procesos software en micro, pequeñas y medianas empresas. Revista Española de Innovación, Calidad e Ingeniería del Software (REICIS), 2006. Vol. 2(1) Abril pp. 6-23.

VAMEPS (Valoración de Modelos y Estándares del Proceso de Software). 2004 – 2008. Diseñado y desarrollado en el marco de la Tesis Doctoral “Enseñanza de Modelos y Estándares de Evaluación y Mejora del Proceso de Software desde la perspectiva CTS (Ciencia, Tecnología y Sociedad)” - Programa de Doctorado en Enseñanza de las Ciencias y la Tecnología - Departamento de Didáctica de las Ciencias Experimentales de la Universidad de Granada, España. 2008, correspondiente a Moralejo, Raúl O.

TICs. Aplicaciones en E-Government y Procesos Productivos

Patricia Pesado^(1,2), Marcelo Naiouf⁽¹⁾, Pablo Thomas⁽¹⁾, Ariel Pasini⁽¹⁾, Franco Chichizola⁽¹⁾, César Estrebou⁽¹⁾, Guillermo Feierherd⁽³⁾, Nicolás Galdamez⁽¹⁾, Ismael Rodríguez⁽¹⁾, Adrián Pousa⁽¹⁾, José Pettorutti⁽¹⁾, Armando De Giusti⁽¹⁾

⁽¹⁾Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

⁽²⁾Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

⁽³⁾Grupo de Investigación en Tecnología Informática Aplicada (GITIA)
Facultad de Ingeniería - UNPSJB (Sede Ushuaia)

{ppesado, mnaiouf, pthomas, apasini, francoch, cesarest}@lidi.info.unlp.edu.ar, feierherdge@speedy.com.ar,
{ngaldamez, ismael, apousa, josep, degiusti}@lidi.info.unlp.edu.ar

CONTEXTO

Esta línea de investigación forma parte del Proyecto “Sistemas de Software Distribuidos. Aplicaciones en procesos industriales, E-government y E-learning” del Instituto de Investigación en Informática LIDI acreditado por la UNLP y de proyectos específicos apoyados por la Comisión de Investigaciones Científicas de la Pcia. de Bs. As. (CIC), ALTEC SE, Telefónica, ESI Center Cono Sur, la Red Interamericana de Formación en Gobierno Electrónico y diferentes organismos públicos de la Pcia. de Buenos Aires.

En este proyecto el Instituto forma parte del PAE EICAR (Electrónica, Informática, Comunicaciones, Automática y Robótica para la producción de Bienes y Servicios) integrado por una red de Universidades (UNLP, UNSJ, UNLM, UTN, UCC, UNPA, UNCPBA), institutos y cámaras empresarias.

También el III-LIDI forma parte del proyecto EELA2 (E-infraestructura compartida entre Europa y Latinoamérica) que preve instalar un Grid de alta capacidad y propósito general con aplicaciones en E-Gov, E-Learning e E-Health.

RESUMEN

Esta línea de investigación incluye por un lado investigación y desarrollo de soluciones de Gobierno Electrónico, utilizando Tecnologías de la Información y la Comunicación y por el otro focaliza en temas específicos de Planificación,

supervisión y optimización de la producción en Pymes incluyendo tecnología de gestión de stocks distribuidos y móviles, en conjunto con las Universidades de la RedEICAR. Esta Red tiene entre sus objetivos coordinar centros regionales (por áreas productivas) que combinan investigación, desarrollo, asesoramiento, capacitación y consultoría especializada en TICs para PYMES.

En el área de Gobierno electrónico tienen especial interés las experiencias concretas (tales como el voto electrónico, la certificación de calidad en sistemas y organizaciones, el estudio de sistemas distribuidos de tiempo real en aplicaciones de E-Gov sobre VLDBs y la auditoría de equipamiento específico) realizadas con organismos públicos concretos (las Provincias de Río Negro y Buenos Aires, Universidades Nacionales, el Ministerio de Desarrollo Humano de la Pcia. de Buenos Aires y los Municipios de La Plata y Ushuaia).

En el área de Planeamiento de la Producción se ha presentado un PME de Mejora del Equipamiento para la RedEICAR, un PICT de Planificación, Supervisión y Optimización de la Producción en PYMES y un PID de Tecnología Gestión de Stocks (fijos y en movimiento) En el marco de la formación de recursos humanos se trabaja en Tesis de Magister y Tesinas de Grado relacionadas directamente con los temas de esta línea de I/D.

Keywords: *E-Government, E-Citizen, E-Health, Civismo digital, Tecnologías de la Información y la Comunicación, Voto electrónico, Planeamiento de la Producción, Gestión de stocks distribuidos.*

1. INTRODUCCION

El desarrollo de las TICs ha permitido la generalización de aplicaciones que potencian la vinculación del ciudadano con el Estado, de modo de mejorar la gestión de la información de interés social. [1] [2] [3]

Al concepto clásico de E-Government que se ha enfocado en poner los servicios tradicionales del Estado al alcance del ciudadano (Consultas en línea, Gestión de trámites, Expedientes digitales, Voto Electrónico, Consultas populares, etc) [4] [5] [6] se agrega la concepción de un E-Citizen, es decir un ciudadano capacitado para interactuar con el Estado, empleando Tecnología. [7] [8].

Esta concepción requiere un gran esfuerzo en disminuir la brecha digital, mediante esfuerzos de capacitación en los cuales deben participar las Universidades [9] [10].

A su vez el Estado puede utilizar las TICs para integrar información (los casos típicos son las fichas de pacientes de Hospitales y los contenidos pedagógicos en Escuelas) y facilitar el proceso de toma de decisiones. [11] [12] [13].

Por último todos estos procesos requieren una cuidadosa auditoría de calidad y seguridades referidas a los derechos individuales y también a la inviolabilidad de la información crítica que maneje el Estado. [14] [15].

En síntesis con el E-Gov se abren varias líneas de Investigación y Desarrollo, asociadas con la integración de las TICs en la sociedad y que requieren un esfuerzo importante en la formación de recursos humanos.

En esta línea se ha trabajado sobre casos específicos, entre los cuales podemos mencionar:

- Desarrollo de equipos específicos para Voto Electrónico.
- Desarrollo de software para diferentes modelos de elecciones.
- Desarrollo de equipos de consulta popular, aplicables en diferentes áreas.

- Estudio y aplicación de sistemas de Identificación biométrica.
- Análisis de performance en transacciones de E-Gov en tiempo real, contra grandes Bases de Datos.
- Auditoría de sistemas de voto electrónico (hardware y software).
- Evolución de la urna electrónica hacia el E-vote utilizando InterNet. Seguridad.
- Certificación de calidad en procesos de gestión pública.
- Modelos de integración de datos heterogéneos, en particular en el ámbito de la salud (E-Health).
- Modelos de integración de contenidos en redes de Escuelas.
- Redes de bibliotecas y documentación digital.
- Utilización de middleware de GRID en aplicaciones de E-Gov.
- Migración de procesos y sistemas para el “ciudadano presencial” por otros para el “ciudadano virtual”, utilizando TICs.

En cuanto a los sistemas de producción, la toma de decisiones relacionadas con el Planeamiento de la Producción en Plantas Industriales PYME requiere el análisis de datos complejos que involucran desde las proyecciones de venta hasta el control en tiempo real de las máquinas que realizan la producción. Este análisis incluye algunos elementos de mayor dificultad cuando las plantas pueden estar físicamente distribuidas, lo que obliga a trabajar sobre aspectos de la logística de materias primas y productos terminados [34][19].

Optimizar esta toma de decisiones tiene un alto significado económico para las empresas (reducción de stocks ociosos, minimización de los tiempos de respuesta, optimización en la asignación de recursos, reducción de los costos de la logística) que requiere el desarrollo de Sistemas de Software que integren modelos e información de mercado con la planificación de la producción [31].

Por otra parte la utilización de tecnología en la gestión de stocks fijos (almacenes) y móviles (durante su transporte) para el seguimiento de productos y materias primas en plantas industriales PYME representa un problema que afecta a la logística del planeamiento de la producción. La minimización y efectiva planificación de la distribución tiene un impacto directo en

los costos de producción y el beneficio a obtener. Este análisis incluye elementos de mayor dificultad cuando las plantas se encuentran físicamente distribuidas. En particular la incorporación de tecnología (tal como la de RFID) a los almacenes de stock (únicos o distribuidos) de una organización permite un control exacto de existencias en tiempo real [18].

Por otra parte el seguimiento de los stocks en movimiento (por ejemplo utilizando tecnología de georeferenciación) permite un control en tiempo real de las transferencias de materias primas y los movimientos de productos terminados entre centros de distribución.

En este tema se ha trabajado sobre casos específicos, entre los cuales podemos mencionar un sistema integrado de Planeamiento de la Producción para PYMES parametrizable y un sistema para mejoramiento y control de la producción, carga de máquinas, abastecimiento de materias primas y servicios, trazabilidad, almacenamiento y distribución de Productos Terminados para una empresa geográficamente distribuida [33].

2. LINEAS DE INVESTIGACION y DESARROLLO

- Sistemas distribuidos. Sistemas basados en tecnología GRID. Modelos y caracterización de performance. Enfoque a las aplicaciones de E-Citizen.
- Voto Electrónico y Voto por InterNet. Arquitecturas adaptadas a la legislación vigente.
- Identificación biométrica [25]. Sistemas y performance. Aplicaciones.
- Análisis de Sistemas de Tiempo Real y estudio de tiempos de respuesta.
- Seguridad y Tolerancia a Fallas en Sistemas de E-Gov.
- Certificación de software y hardware para sistemas de E-Gov. Normas.
- Formación de Tutores en Civismo Digital.
- Formación de especialistas en Certificación de Calidad para sistemas de E-Gov.
- E-Health e E-Learning como herramientas complementarias de E-Gov.
- Modelos de planeamiento de la producción para diferentes clases de plantas industriales.
- Integración de información para la toma de decisiones en organizaciones industriales distribuidas.

- Tecnologías para control inteligente de stocks (en particular RFID).
- Optimización de movimientos en stocks distribuidos.

3. RESULTADOS OBTENIDOS /ESPERADOS

- ✓ Desarrollo y experimentación de una urna electrónica y su software parametrizable para elecciones en Argentina (Municipales, Provinciales, Nacionales). Estudio de Casos en Bs. As. y Tierra del Fuego. [16] [17].
- ✓ Desarrollo de una terminal integral de consulta, utilizable también como puesto de Voto Electrónico, para lugares fijos. Utilización de la misma en la UNLP. [30].
- ✓ Auditoría y certificación de los equipos de Voto Electrónico aprobados en la Pcia. de Río Negro y desarrollados por ALTEC SE [20] [32].
- ✓ Auditoría del software de recuento e integración de datos en las elecciones de Río Negro.
- ✓ Desarrollo e implementación del E-Vote sobre Internet, utilizado en la Facultad de Informática y las Sedes de Tres Arroyos y Las Flores en las elecciones estudiantiles 2008 [21] [30].
- ✓ Curso de "TIC'S aplicadas a E-Governments" dictado en la Escuela de CACIC 2008
- ✓ Investigación en el empleo de tecnología GRID para la vinculación de hospitales e historias clínicas en la Pcia. de Buenos Aires. Investigación en la digitalización de estudios por imágenes. (Proyecto en curso con IOMA). [22].
- ✓ Estudio sistemático del modelo de los Planes Sociales vigentes en la Pcia de Buenos Aires (más de 20 sistemas) y el modo de utilizar TICs para su simplificación y auditoría. (con el Ministerio de Desarrollo Humano de la Pcia. de Bs. As.). [23].
- ✓ Proyecto de desarrollo de contenidos e integración de redes de Escuelas (a nivel mu-

nicipal) para mejorar los procesos de E-Learning. [24].

- ✓ Estudio de diferentes equipos de identificación biométrica, en particular para huellas digitales. Experiencia de sistemas de trámites WEB basados en identificación segura con estos equipos y contra BD de tamaño creciente, centralizadas y distribuidas. [26].
- ✓ Convenio para la capacitación de Tutores en Civismo Digital y desarrollo de un primer curso en la UNLP en Marzo 2008. [27].
- ✓ Convenio para constituir el ESI Center Argentina, y dentro del mismo proyecto de certificación de los procesos de Concursos, Manejo de Resoluciones y Expediente Digital en la Facultad de Informática de la UNLP. [28].
- ✓ Investigar la especificación e implementación de WEB services eficientes sobre Grid, aplicados en E-Gov.
- ✓ Acuerdo en la Asociación de Universidades del Grupo Montevideo (AUGM) para integrar una red de Bibliotecas que abarque todas las Universidades miembro. [29].
- ✓ Desarrollo de un “Sistema integrado de Planeamiento de la Producción para PY-MES” subsidiado por la CIC, cuyo objetivo fue analizar un conjunto de modelos de planta con distintas características (mono y multimáquina, producción a pedido, por análisis de stock, por pronósticos de ventas, etc.) y desarrollar un sistema básico y parametrizable para resolverlo.
- ✓ “Desarrollo de herramientas digitales para mejorar las estructuras de la producción, control de la producción, carga de máquinas, abastecimiento de materias primas y servicios, trazabilidad, almacenamiento y distribución de Productos Terminados” para la Empresa GRAFEX (Fábrica de Tintas y Barnices), subsidiados por la Agencia Nacional de Promoción de Ciencia y Tecnología. Se trató de un sistema que contempla la distribución geográfica de los lugares de producción (Villa Mercedes-San Luis y Buenos Aires)

4. FORMACION DE RECURSOS HUMANOS

Hay 6 Investigadores realizando su Posgrado y 3 alumnos avanzados están trabajando en su Tesina de Grado de Licenciatura.

5. BIBLIOGRAFIA

- [1] Center for democracy and technology E-Government Handbook. 2002.
<http://www.cdt.org/egov/handbook/>
- [2] Jones A., Williams L. “Public Services and ICT - FINAL REPORT. How can ICT help improve quality, choice and efficiency in public services?”. London: The Work Foundation. 2005.
- [3] London: National Audit Office. “Better Public Services through e-government”. Report HC 704-I Session 2001-2002.
- [4] Washington DC: Office of Management and Budget. “E-Government Strategy: Simplified Delivery of Services to Citizens”. OMB. 2002.
- [5] “Citizen Centric Government: Global Best Practice in Delivering Agile Public Services to Citizens and Businesses”. London: Gov3 Ltd. 2006.
- [6] Brussels: European Commission. “The Role of eGovernment for Europe's Future”. Report COM(2003) 567 Final. Europe. 2003.
- [7] Wang L., Bretschneider S., Gant J. “Evaluating Web-based e-government services with a citizen-centric approach”. Proceedings of the 38th Hawaii International Conference on System Sciences. 2005.
- [8] “Transforming Public Services: The Next Phase of Reform. Edinburgh: Scottish Executive”. Scotland 2006.
- [9] Clift S. “e-Government and Democracy: Representation and Citizen Engagement in the Information Age”. 2004.
- [10] O'Donnell S., McQuillan H., Malina A. “eInclusion: expanding the Information Society in Ireland. Dublin: Government of Ireland. Information Society Commission”. 2003.
- [11] Juma C., Yee-Cheong L. “Reinventing global health: the role of science, technology and innovation”. Lancet 2005.
- [12] Laguna A., Ferri Tormo R., Hernandez V., Peñarrubia J. “gCitizen: uso de tecnologías Grid para la interoperabilidad entre Administraciones Públicas”. IX Jornadas

- sobre Tecnología de la Información para la Modernización de las Administraciones Públicas. Sevilla. 2006.
- [13] Brunner J. "Educación: escenarios de futuro. Nuevas Tecnologías y sociedad de la información". PREAL, Santiago de Chile. 2000.
- [14] Bhutta Z. "Ethics in international health research: a perspective from the developing world". Bulletin of the World Health Organization. 2002.
- [15] Alakeson, Vidhya, Aldrich T., Goodman J., Jorgensen B., Mill P. "Social Responsibility in the Information Society. Brussels: DEESD - Digital Europe: e-business and sustainable development". Report Deliverable 7 (D7) DEESD IST-2000-28606. 2003.
- [16] Feierherd G., De Giusti A., Pesado P., Depetris B. "Una aproximación a los requerimientos del software de voto electrónico de Argentina". CACIC 2004.
- [17] Pesado P., Feierherd G., Pasini A.. "Especificación de Requerimientos para Sistemas de Voto Electrónico". CACIC 2005.
- [18] Rehg J., Kraebberg H. "Computer-Integrated Manufacturing" (Third Edition). Prentice Hall. 2004.
- [19] Everdingen, Y., Hillegersberg, J., Waarts, E. "ERP Adoption by European Midsize Companies". Communications of the ACM, vol. 43, nº. 4, pp. 27-31. 2000.
- [20] Pasini A., Ibáñez E., Galdamez N., Estrebou C., Rodríguez I., Pousa A., De Giusti A. "Análisis Urna Electrónica ALTEC S.E.". Informe Técnico III-LIDI. 2007.
- [21] Feierherd G. "Voto por Internet". Reporte Técnico UNPSJB.
- [22] "Protocolo Específico entre IOMA y la Facultad de Informática de la UNLP". 2007.
- [23] "Acuerdo entre el Ministerio de Desarrollo Humano de la Pcia. de Bs. As. y la Facultad de Informática de la UNLP". 2007.
- [24] De Giusti A. "Integración de contenidos para Redes de Escuelas en la Pcia. De Bs. As.". Informe Técnico. 2007.
- [25] Reid P. "Biometrics for Network Security". Prentice Hall. 2004.
- [26] Carri J., Pasini A., Pesado P., De Giusti A. "Reconocimiento biométrico en aplicaciones de E-Government. Análisis de confiabilidad / tiempo de respuesta." CACIC 2007.
- [27] "Acuerdo de colaboración y cooperación entre la Facultad de Informática y la Red Interamericana de Formación en Gobierno Electrónico del COLAM-OUI". 2007.
- [28] "Acuerdo de creación de un ESICENTER en Argentina, entre European Software Institute, Grupo Tekne, Universidad Nacional de La Plata, Universidad Nacional de San Martín". 2007.
- [29] "Red de Bibliotecas de las Universidades de AUGM". Proyecto aprobado en el Núcleo Disciplinario "Redes Académicas" de AUGM. 2007.
- [30] Pesado P., Pasini A., Ibáñez E., Galdamez N., Chichizola F., Rodríguez I., Estrebou C., De Giusti A. "E-Government-El voto electrónico sobre Internet". CACIC 2008.
- [31] Sprott D. "Enterprise Resource Planning: Componentizing the Enterprise Application Packages". Communications of the ACM, vol. 43, nº. 4, pp. 63-69. 2000.
- [32] Estrebou C., Galdamez N., Pasini A., Pousa A., De Giusti A. "Análisis Urna Electrónica ALTEC S.E. Actualización 2008". Informe Técnico. III-LIDI. 2008.
- [33] Estrebou C., Romero A., Galdamez N., Moralejo L. "Sistema Web para Planeamiento de la Producción de una empresa con Plantas distribuidas". Jornadas AUGM. Campinas (Brasil). 2006.
- [34] Bernroider E., Hampel A. "Enterprise resource planning and it governance in perspective: strategic planning and alignment, value delivery and controlling". Academic Publishers. Viena (Austria). 2005.

Calidad en el desarrollo de Sistemas de Software

Pesado Patricia^(1,2), Bertone Rodolfo⁽¹⁾, Esponda Silvia⁽¹⁾, Pasini Ariel⁽¹⁾,
Martorelli, Sabrina⁽¹⁾

⁽¹⁾Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

⁽²⁾Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

(ppesado, pbertone, sesponda, apasini, smartorelli) @lidi.info.unlp.edu.ar

1. CONTEXTO

Esta línea de Investigación forma parte del Proyecto “Sistemas de Software Distribuidos. Aplicaciones en procesos industriales, E-government y E-learning” del Instituto de Investigación en Informática LIDI acreditado por la UNLP y de proyectos específicos apoyados por CyTED (Red Iberoamericana integrada por 16 Universidades de Argentina, México, Brasil, Colombia, España, Uruguay, Perú, Ecuador, entre otras), CIC y la Facultad de Informática de la UNLP.

2. RESUMEN

El objetivo del subproyecto “Calidad en el desarrollo de Sistemas de Software” es analizar, investigar y desarrollar propuestas en temas relacionados con el aseguramiento de la calidad en cada una de las etapas del desarrollo de Sistemas de Software. En particular se ha trabajado y estudiado las normas de calidad más ampliamente difundidas, líneas ISO y CMMI, y su incidencia en entornos como la pequeña y mediana industria de nuestro país.

Desde el año 2006, se ha trabajado en la adecuación de las normas de calidad a empresas desarrolladoras de software de pequeña o mediana envergadura. En particular el proyecto COMPETISOFT, del cual el grupo de trabajo es parte desde su origen, ha trabajado en tal sentido, desarrollando una propuesta de Mejora de Procesos Software para Pequeñas y

Medianas Empresas y Proyectos y durante 2008 ha publicado sus resultados definitivos [PIA08]. El modelo desarrollado puede catalogarse como descriptivo, puesto que no prescribe métodos o prácticas específicas para realizar las diversas actividades propuestas. Con este modelo se están evaluando empresas de la zona de La Plata y definiendo un plan de mejoras.

El III-LIDI a través de la Facultad de Informática, participó en la creación del ESI Center Cono Sur (European Software Institute), cuyos objetivos son:

- Investigación conjunta en temas de Calidad en Ingeniería de Software
- Estudio y aplicación de Normas de calidad
- Formación de Recursos Humanos orientado a la certificación de software.

Existen convenios de cooperación con diversos organismos del estado para la evaluación de distintos áreas del proceso y la definición de un plan de mejoras. En particular en la Facultad de Informática se ha creado un Área de Calidad, para analizar, definir y establecer un plan a ser aplicado a distintos procesos de la Gestión Universitaria, en coordinación con este proyecto.

Palabras Claves

Ingeniería de Software – Calidad – Normas de Calidad – CMM, ISO, COMPETISOFT

3. INTRODUCCION

La demostración de la calidad y a posteriori su validación y certificación son uno de los grandes retos que debe afrontar la industria de la región Iberoamericana para obtener frutos del trabajo realizado en esta área tecnológica. La mejora de procesos software es un esfuerzo planeado, gestionado y controlado que tiene como objetivo incrementar la capacidad de los procesos de desarrollo de software de una empresa.

Una empresa madura tiene un alto nivel de capacidad de sus procesos, lo cual le permite gestionar exitosamente los procesos asociados con el desarrollo, mantenimiento y soporte de productos software. Esta mejora en el proceso software es una práctica de interés para las PyMES que se dedican a construir software. Estas organizaciones pretenden asegurar la calidad de sus productos a través de la evaluación y mejora de sus procesos, acreditándose en modelos reconocidos por la industria del software (CMMI, SCAMPI, IDEAL, ISO/IEC 15504). [ISO95] [ISO04a] [Lan05] [Pau93] [Pau95] [SEI06].

En algunos países iberoamericanos se abordaron estos problemas con algunas iniciativas como el modelo MOPROSOFT de México [Okt05], el modelo MR MPS de Brasil [Web04], o el modelo SIMEP-SW de Colombia [Hur03], incluso la metodología Métrica v.3 propugnada por el MAP en España también intenta conseguir la mejora de los procesos y productos software.[Pin05]

Los modelos antes mencionados plantean características comunes:

- Presentar de forma efectiva los componentes fundamentales del marco metodológico desarrollado.
- Exponer los aspectos tecnológicos y metodológicos más significativos,

relacionados con la mejora de procesos software para PyMES.

En consecuencia surge el modelo COMPETISOFT con el fin de unificar el criterio de estos países.

En particular su versión COMPETISOFT Perfil Básico que surge de la retroalimentación obtenida de la aplicación de COMPETISOFT a PyMEs. Esta versión aplica solamente los procesos de la categoría Operación [CYT07-2].

4. LINEAS DE INVESTIGACION Y DESARROLLO

- Evaluar en forma continua el cuestionario de Administración de Proyectos Específicos y el correspondiente a la Etapa de Desarrollo, generado oportunamente por el grupo de trabajo, basado en COMPETISOFT Perfil Básico.
- Generar un entorno WEB que permita asistir a una PyME en los pasos necesarios para la aplicación de COMPETISOFT Perfil Básico en sus procesos de desarrollo de software. Proponiendo un ejemplo de documentación necesaria y brindando una ayuda en línea que indique no solamente los aspectos definidos por COMPETISOFT sino que, además, presente soluciones y ejemplos que hayan sido utilizados en otros entornos y puedan ser de interés en encontrar una alternativa de trabajo.
- Evaluar la Norma ISO/IEC 20000 para su aplicación en el proceso de certificación del curso de PRE Ingreso a Distancia y el proceso de Concursos, en la Facultad de Informática de la UNLP.

5. RESULTADOS OBTENIDOS/ ESPERADOS

- Se aplicó el Modelo COMPETISOFT Perfil Básico a una PyME de la ciudad de La Plata, obteniendo mejoras en los procesos de desarrollo. En la actualidad

se asiste a la empresa para alcanzar el siguiente nivel de capacidad. .

- Se evaluó la calidad en los procesos de desarrollo de los equipos de Voto Electrónico aprobados en la Pcia. de Río Negro y desarrollados por ALTEC SE.
- Se evaluó la calidad en los procesos de desarrollo del software de recuento e integración de datos en las elecciones de Río Negro.
- Se participó en el Taller de Capacitación COMPETISOFT: Mejora de Procesos Software para Pequeñas Empresas, celebrado en Lima – Perú, en abril del 2008.
- Se realizó un método de cuantificación de los cuestionarios elaborados para el proceso de Administración de un Proyecto Especifico en el marco del Modelo de Competisoft Perfil Básico.
- Se desarrolla un sistema de plantillas para la asistencia, generación y seguimiento de documentos asociados con calidad.
- Se evalúa la Norma ISO/IEC 20000 para su aplicación en el proceso de certificación del curso de PRE Ingreso a Distancia.
- Se evalúa el proceso de Concursos, en la Facultad de Informática de la UNLP, a fin de realizar su mejora.
- Se he definido un plan de trabajo, con el fin de aplicar los conocimientos adquiridos para evaluar y sugerir cambios en la Empresa ALTEC, Sociedad del Estado de la Pcia de Río Negro. Aquí, la empresa considera que para mejorar su proceso de construcción de software es necesario evaluar los planes vigentes, analizando ventajas y desventajas presentes y generando un plan de mejora que permita alcanzar niveles de madurez óptimos para su desarrollo.

6. FORMACION DE RECURSOS HUMANOS

En el año 2008 se aprobaron dos Tesinas de Grado de Licenciatura, dirigidas por miembros del proyecto y se encuentran en desarrollo 1 Tesinas de Grado de Licenciatura y 2 Tesis de Magíster, en temas relacionados con el Proyecto.

Los integrantes de esta línea de Investigación participan en el dictado de Asignaturas/cursos de grado/postgrado de la Facultad de Informática de la UNLP.

7. BIBLIOGRAFIA

[ACM] Colección de Communications of the ACM

[ACM] Colección de ACM SIGSOFT (Special Interest Group on Software Engineering)

[Ber06] Bertone, Pasini, Ramon, Esponda, Pesado, De María, Mon, Gigante, Estayno Gestión de Calidad en la Construcción del Software. Un enfoque para PyME's. Cacic 2006. San Luis.

[CYT07]Competisoft Modelo de Proceso para PyMEs de Íbero América, Proyecto CYTED (Código 3789) Agosto 2007

[CYT07-2]Competisoft Perfil Básico Modelo de Proceso para PyMEs de Ibero América, Proyecto CYTED (Código 3789) Agosto 2007

[IEEE] Colecciones de Transaction on Software Engineering

[IEEE] Colecciones de Computer

[ISO95] ISO/IEC .12207: 1995 Information Tecnology – Software life cycle processes.ISO/IEC.1995

[ISO04a] ISO/IEC .15504-1: 2004 Information Tecnology – Process assessment – Part 1: Concepts and vocabulary.ISO/IEC,2004

[Lan05] Susan K. Land.Jumpstart CMM/CMMI Software Process Improvements: Using IEEE Software Engineering Standards.Wiley-IEEE Computer Society Press.2005

[Okt05] Modelo de procesos para la industria del software. MoProSoft. Por

niveles de Capacidad de Procesos. Versión 1.3, Agosto 2005.

[Pas08] Pasini, A, Esponda S, Bertone R, Pesado P. Aseguramiento de Calidad en PYMES que desarrollan software. Una experiencia desde el proyecto COMPETISOFT. CACIC 2008. Chilecito La Rioja Octubre 2008

[Pau93]Paulk, M.C.; Curtis, B et al. Capability Maturity Model, Versión 1.1. IEEE Software, 10(4), 18-27. (1993)

[Pau95]Paulk, M.C.; Weber, C.V. et al. The Capability Maturity Model: Guidelines for Improving the Software Process. Reading, MA: Addison-Wesley. (1995)

[Pia08] Piattini, M; Oktaba, H; Pino, F; Orozco, M; Alquicira, C. COMPETISOFT. Mejora de Procesos Software para Pequeñas y Medianas Empresas y Proyectos. Editorial RaMa. ISBN; 978-84-7897-901-1. 2008

[Pin05]Pino, F.; García, F; Piattini, M.: Adaptación de las normas ISO/IEC 12207:2002 e ISO/IEC 15504: 2003 para la evaluación de la madurez de procesos de software en países en desarrollo.

[Ple02] Pleegeer. Ingeniería de Software: Teoría y Práctica. Prentice-Hall. 2002

[SEI06] Proceedings of the First International Research Workshop for Process Improvement in Small Settings, 2005. January 2006 . SPECIAL REPORT CMU/SEI-2006-SR-001

[Som05] Sommerville Ian. .Requeriments Engineering, A good practice guide. .John Wiley. 2005

[Web04] Modelo de Referência e Método de Avaliação para Melhoria de Processo de Software - versão 1.0 (MR-MPS e MA-MPS)"

Sistemas de Software Distribuido

P. Pesado^(1,2), H. Ramón⁽¹⁾, P. Thomas⁽¹⁾, M. Boracchia⁽¹⁾,
I. Rodríguez⁽¹⁾, E. Ibañez⁽¹⁾, L. Marrero⁽¹⁾, L. Delia⁽¹⁾,
C. De Vito⁽¹⁾, N. Restelli⁽¹⁾, G. Caseres⁽¹⁾, J. Pettoruti⁽¹⁾

⁽¹⁾Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

⁽²⁾Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

{ppesado, hramon, pthomas, marcosb, ismael, eibanez, lmarrero, ldelia, cdevito, nrestelli,
gcaseres,josep}@lidi.info.unlp.edu.ar

CONTEXTO

Esta línea de Investigación forma parte del Proyecto “Sistemas de Software Distribuidos. Aplicaciones en procesos industriales, E-government y E-learning” del Instituto de Investigación en Informática LIDI acreditado por la UNLP y de proyectos específicos apoyados por CIC, Agencia y Telefónica, ALTEC e INASE.

RESUMEN

El objetivo de este subproyecto es realizar investigación y desarrollo en temas relacionados con los aspectos de Ingeniería de Software que se orientan al desarrollo e implementación de proyectos concretos de Sistemas Distribuidos, manejo de datos físicamente distribuidos y soluciones de hardware y software para Sistemas Distribuidos de Tiempo Real.

En este contexto se trabaja en aspectos de la Ingeniería de Software de diferentes sistemas con inteligencia distribuida (computadoras, robots, teléfonos móviles).

Palabras claves: *Sistemas Distribuidos – Ingeniería de Requerimientos – Planificación – Metodologías de Desarrollo – Sistemas Web – Bases de Datos Distribuidas - Sistemas Distribuidos de Tiempo Real.- Robótica*

1 INTRODUCCION

Un sistema distribuido consiste en un conjunto de computadoras autónomas conectadas por una red y con soporte de software distribuido. Permite que las computadoras coordinen sus actividades y compartan los recursos de hardware, software y datos, de manera tal que el usuario percibe una única facilidad de cómputo integrada aunque ésta pueda estar implementada por varias máquinas en distintas ubicaciones [1].

El desarrollo de sistemas distribuidos es una necesidad a partir de la utilización de redes de computadoras y de computadores personales de alta performance.

Algunas ventajas del procesamiento distribuido son:

- Mejora de la disponibilidad: la operación es factible en una configuración reducida cuando algunos nodos están temporalmente no disponibles.
- Configuración más flexible: una aplicación puede configurarse de distintas maneras, seleccionando el número apropiado de nodos para una instancia dada.
- Control y administración más localizada: un subsistema distribuido, ejecutando en su propio nodo, puede diseñarse para ser autónomo, de modo que puede ejecutar en

relativa independencia de otros subsistemas en otros nodos.

- Expansión incremental del sistema: si existe sobrecarga, el sistema puede expandirse agregando más nodos.
- Costo reducido: con frecuencia una solución distribuida es más barata que una centralizada.
- Balance de carga: en algunas aplicaciones la carga total del sistema puede ser compartida entre varios nodos.
- Manejo eficiente de datos distribuidos físicamente.
- Mejora en el tiempo de respuesta: los usuarios locales en nodos locales pueden obtener respuestas más rápidas a sus requerimientos.

Las características de los Sistemas Distribuidos conducen a la utilidad de desarrollar prácticas de Ingeniería de Software que apunten a los distintos aspectos del desarrollo de sistemas desde la captura de requerimientos y la planificación, pasando por las metodologías de desarrollo, hasta la verificación y simulación de procesos distribuidos, incluyendo el aseguramiento de calidad.

En particular un sistema distribuido de tiempo real debe interactuar con el mundo real, en puntos físicamente distantes y no necesariamente fijos, en períodos de tiempo que vienen determinados por el contexto o las restricciones de la especificación (en muchos casos a partir de una activación asincrónica).

Algunas de las dificultades principales del desarrollo de software para sistemas distribuidos de tiempo real son:

- Modelizar condiciones de concurrencia y paralelismo.
- Manejar las comunicaciones inter-procesos e inter-procesadores.
- Tratamiento de señales en tiempo real.
- Tratamiento de interrupciones y mensajes asincrónicos con diferente prioridad.
- Detectar y controlar condiciones de falla, a nivel de software, de procesadores y de comunicaciones. Prever diferentes grados de recuperación del sistema.

- Asegurar la confiabilidad de los datos y analizar su migración en condiciones de funcionamiento normal o de falla.
- Organizar y despachar la atención de procesos, manejando las restricciones de tiempo especificadas.
- Testear y poner a punto un sistema físicamente distribuido.

Todas estas dificultades conducen a la utilidad de desarrollar herramientas de Ingeniería de Software orientadas a STR y SDTR, de modo de trabajar en la modelización, especificación y verificación del software considerando las restricciones temporales.

La Ingeniería de Software comprende la aplicación de principios científicos para realizar la transformación ordenada de un problema en una solución elaborada de software, y el mantenimiento subsecuente de ese software hasta el final de su vida útil [2]. La utilización de estas prácticas para resolver sistemas distribuidos y de tiempo real hacen necesaria su adaptación en función de las características de dichos sistemas.

La adopción de un enfoque ingenieril para el desarrollo de software, genera una serie de fases o estados conformando un ciclo de vida. Este ciclo de vida está guiado por una planificación que incluye el conjunto de acciones a realizar, y los productos generados por la aplicación del plan (inclusive el mismo plan) están administrados por diferentes Metodologías de Gestión y Desarrollo [3].

En el recorrido del ciclo de vida del desarrollo del software, la fase inicial comprende a la Ingeniería de Requerimientos que permite comprender, documentar y acordar sobre el alcance del problema, teniendo esto impacto directo sobre la Planificación y la Gestión del Proyecto de acuerdo a la Metodología de desarrollo seleccionada para el mismo [4]. Este no es el único impacto que justifica el énfasis en la Ingeniería de Requerimientos, ya que esta disciplina permite establecer claramente *que* se debe hacer posponiendo el *cómo* a etapas posteriores en el ciclo de vida,

constituyendo el fundamento de la construcción de un Sistema de Software [5].

El modelo de procesos elegido para el desarrollo de software define las actividades a realizar para la generación de productos de acuerdo a los objetivos planteados [6].

Asociado con la Ingeniería de Software de Sistemas Distribuidos, está el problema de utilizar un entorno WEB para los servicios que ofrece el Sistema. La tendencia creciente al desarrollo de arquitecturas centradas en un servidor (o un conjunto de servidores distribuidos) que ofrecen una interfaz WEB a los usuarios, ha generado un importante desarrollo de la Investigación en metodologías y herramientas orientadas a Sistemas WEB, así como ha obligado a establecer nuevas métricas y parámetros de aseguramiento de la Calidad para tales Sistemas. [7] [8] [9]

Toda aplicación esta dividida, al menos, en dos subsistemas, front-end y back-end. El front-end es una aplicación que interactúa en forma transaccional con los usuarios (clientes, empleados, proveedores, entre otros) de una Organización y requiere del apoyo del subsistema back-end que posee características analíticas y de workflow. Un ECA-rules (Event-Condition-Action) representan la tecnología para BD activas y el método natural para soportar dicha funcionalidad. ECA rules puede ser utilizado para actividades de chequeo automático de restricciones, o mensajes. Una cuestión importante, asociada a un ECA es cómo predecir el comportamiento en tiempo de ejecución. ECA rules posee una sintaxis declarativa de alto nivel, que le permite llevar a cabo acciones automáticas en respuesta a condiciones previamente definidas. Un ECA-rules sobre un subsistema Back-End permite controlar de manera transparente para el usuario un conjunto de reglas que hacen a su negocio. Estos controles se realizan directamente en el servidor de la aplicación y, ante una acción, disparan los mensajes correspondientes sobre el usuario en su aplicación front-end. [10] [11]

Por último, el modelo distribuido de datos hace posible la integración de BD heterogéneas proveyendo una independencia global del administrador de bases de datos respecto del esquema conceptual. [12].

2 LINEAS DE INVESTIGACION Y DESARROLLO

- Conceptos de procesamiento distribuido. Arquitectura, comunicaciones y software.
- Metodologías de especificación, validación y desarrollo de SSD y SSDTR.
- Metodologías ágiles de desarrollo utilizando frameworks propios y disponibles de uso libre con diferentes tecnologías.
- Ingeniería de Requerimientos, en particular de sistemas distribuidos.
- Planificación de tareas para desarrollo de sistemas distribuidos.
- Bases de Datos Distribuidas [13]
- Herramientas de integración y mantenimiento de proyectos distribuidos.
- Lenguajes y ambientes para procesamiento distribuido.
- Reingeniería de sistemas complejos que migran por down-sizing a esquemas cliente-servidor distribuidos.
- Sistemas de Tiempo Real. Sistemas Distribuidos de Tiempo Real.
- Tratamiento de señales en tiempo real.
- Sistemas de control de robots en tiempo real. Algoritmos colaborativos entre máquinas móviles.
- Sistemas con identificación segura en tiempo real.
- ECA-rules sobre un subsistema Back-End permite controlar de manera transparente para el usuario un conjunto de reglas que definen su comportamiento [14]

3 RESULTADOS ESPERADOS/OBTENIDOS

- Avanzar en la capacitación continua de los miembros de la línea de investigación.
- Desarrollar soluciones a problemas concretos de software de sistemas distribuidos, poniendo énfasis en el desarrollo de metodologías y herramientas específicas para clases de aplicaciones.

- Framework NET4DB desarrollada en el III-LIDI que permite generación automática de código .NET, interactuando con diferentes motores de BDD [14] [15]. Actualmente se estudian herramientas que permiten tomar un modelo UML de un sistema y derivar código a partir de él o realizar una verificación de consistencia lógica de dicho modelo.
- El motor simpleECA para resolver el back-end de PHP4DB [16] y NET4DB.

Se ha realizado transferencia a los siguientes Organismos.

3.1 Instituto Nacional de Semillas (INASE):

El Instituto Nacional de Semillas dependiente de la Secretaría de Agricultura de la Nación fiscaliza la comercialización de semillas para siembra, con el propósito de certificar la calidad de estas semillas.

Para realizar este control, el organismo utiliza un Sistema Web donde se tramitan las presentaciones realizadas por las empresas vendedoras de semillas.

El III-LIDI tiene a cargo el proceso de reingeniería de este Sistema, con el objetivo de que las empresas mencionadas previamente puedan realizar trámites ante el INASE utilizando este software.

Con este proyecto se pretende descentralizar la distribución de numeración de rótulos utilizados para identificar a las semillas con “control de calidad realizado”, así como facilitar la gestión de trámites, permitiendo que varias etapas del proceso generado con estos trámites, sean cumplidas directamente por las empresas.

El resultado esperado es el fortalecimiento del proceso de fiscalización nacional de semillas, dado que las empresas podrán gestionar sus trámites a través de Internet, pudiendo controlar la identidad de la semilla en el mercado a través de la información del rótulo que la identifica, así como mantener el seguimiento de todo trámite iniciado en tiempo real.

3.2 Proyecto de Portal de la Innovación en TICs – Fase Primera

El portal de innovación es una herramienta de comunicación comunitaria y social cuyo fin es soportar las actividades de innovación de las Instituciones de I+D y las empresas TIC de la Provincia de Bs. As. Al mismo tiempo se instaura como espacio por excelencia para información, capacitación, networking y colaboración en temas de innovación TIC.

En la primera fase, el proyecto propone desarrollar la tecnología base que permita publicar contenidos ricos, poner en marcha las áreas funcionales de desarrollo de contenidos y consultoría/capacitación, e investigar funcionalidad de redes sociales y colaboración para el contexto de la innovación en TICs.

La demanda local de necesidades de los proyectos de innovación es amplia pero dispersa y no hay organismo o programas públicos no privados que se ocupen sistemáticamente, de organizarla, y buscar soluciones en la oferta en el sector científico y privado, las necesidades de un portal para:

- Promover el intercambio de información sobre oferta y demanda de proyectos de innovación y experiencias de I+D entre los actores de la comunidad TIC.
- Sacar del aislamiento a los focos de innovación e interconectarlos entre sí.
- Crear herramientas que sistematicen la comunicación y vinculación a una escala local-regional-nacional.
- Proveer servicios de transferencia de capacidades para el nivel local (capacitación en gerenciamiento tecnológico, financiamiento a empresas de base tecnológicas e internacionalización).

El portal está alineado con las tendencias de Web 2.0 dando libertad para la generación de contenido para la comunicación entre actores.

3.3 Sistema de Billing para Locutorios VoIP

Este proyecto forma parte del acuerdo entre el III-LIDI y el Grupo Angras. Se realizó la

implementación de un sistema de billing para el manejo de locutorios implementados sobre sistemas de Voz sobre IP (VoIP).

Se implementó un emulador del sistema POSNet para realizar testing y un prototipo del sistema evitando detalles de implementación en el hardware.

En el marco de la arquitectura VoIP se decidió utilizar Asterisk como central telefónica (PBX VoIP) para el desarrollo del proyecto, haciendo uso de la distribución de software libre AsteriskNow.

La implementación del proyecto se realizaron en C utilizando, sockets, librerías Standard ANSI y pthreads; esta última para el manejo de concurrencia en la aplicación.

La comunicación entre el servidor de billing y los clientes POSNet, requirieron la definición de un protocolo de comunicación adecuado al marco del problema, que trabaje sobre el protocolo de Internet TCP/IP. Se utilizaron los beneficios que brinda TCP pues proporciona una cantidad considerablemente mayor de servicios a las aplicaciones que UDP, como ser: la recuperación de errores, control de flujo y fiabilidad.

Como una evolución natural del desarrollo, se extendió la implementación del servidor de billing para posibilitar la atención a múltiples POSNets, verificando cada método de protección de la ejecución concurrente.

3.4 Convenio de Cooperación con Alta Tecnología Sociedad del Estado (ALTEC SE)

El proyecto Intranet Pública Provincial se basa en una Única Red de Datos, este moderno sistema permitirá que los diferentes organismos públicos, empresas públicas, y beneficiarios de servicios públicos reales y potenciales distribuidos territorialmente a grandes distancias se conecten a través de una Intranet basada en estándares predefinidos y administrada en forma centralizada. El objetivo principal es el de brindar comunicaciones de datos a los Organismos rionegrinos optimizando los servicios que presta el Estado en las áreas de educación,

salud y seguridad. Las tareas a desarrollar son:

- Análisis de la Performance de los sistemas de software utilizados en la Pcia. de Río Negro (Dirección General de Rentas, Policía de Río Negro, Dirección de Catastro, Registro de la Propiedad del Inmueble, Ministerio de Educación, Ministerio de Salud), sobre los cuales ALTEC tiene responsabilidad.
- Diseño del manual de procedimientos de prueba de la intranet Provincia. Esto debe incluir comunicaciones puras y sobrecarga con los sistemas anteriores
- Análisis de los resultados obtenidos con las sugerencias que surjan de dicho análisis para la infraestructura y arquitecturas de sistemas.

4 FORMACION DE RECURSOS HUMANOS

Los integrantes de esta línea de investigación dirigen Tesinas de Grado en la Facultad de Informática, 1 Tesis Doctoral, 2 Tesinas de Grado y Becarios III-LIDI en temas relacionados con el proyecto y participan en el dictado de asignaturas/cursos de postgrado de la Facultad de Informática de la UNLP.

5 BIBLIOGRAFIA

- [1] G. Coulouris. Distributed Systems – Concepts and Design. Addison-Wesley. 1994.
- [2] R. Pressman. Ingeniería de Software: Un Enfoque Práctico. McGraw-Hill. 2002
- [3] R. Wysocki. Effective Project Management: Traditional, Adaptive, Extreme, .Wiley .2003
- [4] Loucopoulos, P; Karakosas, V.. Systems Requirements Engineering. .McGraw Hill. Book Company. 1995
- [5] G. Kotonya and I. Sommerville,. Requirements Engineering: Processes and Techniques, Wiley. 1998
- [6] Pleegeer. Ingeniería de Software: Teoría y Práctica. Prentice-Hall. 2002

[7] Stephen Kan. .Metrics and Models in Software Quality Engineering (2nd Edition).Addison Wesley. 2003

[8] Offutt J., “Quality Attributes of Web Software Applications”. IEEE Software: Special, Issue on Software Engineering of Internet Software 19 (2):25-32, Marzo/Abril 2002.

[9] Wu, Y. y Offutt, J. “Modeling and testing web-based Applications”.
<https://citeseer.ist.psu.edu/551504.html>: 1-12, Julio 2004

[10] Automatic Control of Workflow Processes Using ECA Rules, Joonsoo Bae, Hyerim Bae, Suk-Ho Kang, Yeongho Kim, IEEE Transactions on Knowledge and Data Engineering , Volume 16 , Issue 8 (August 2004), 2004, ISSN:1041-4347

[11] An Event-Condition-Action Language for XML, J. Bailey, A. Poulouvasilis, P. Word, WWW2002, May 7-11, 2001, Honolulu, Hawaii, USA. ACM 1-58113-449-5/02/0005.

[12] Silberschatz A et all: “Fundamentos de Bases de Datos”, Tercera Edicion Mc Graw Hill 1998

[12] Ozsu M. Valduriez, P. : “Principles of Distributed Database Systems”, Segunda Edicion. Prentice Hall 1999

[13] Mello S. J., “Executable UML”, Addison-Wesley, 2002.

[14] S. Ceri, P.Fraternalli,A. Bongio, “Web Modeling Language (WebML): a modeling language for designing web sites”. Computer Networks, vol 33, 2000.

[15] Tesis Stefan Böttger, Distributed Composite Event Monitoring. June, 2006 Ruprecht-Karls-Universität Heidelberg (Kirchhoff-Institut für Physik) – Universität Leipzig (Fakultät für Mathematik und Informatik – Institut für Informatik).

[16] Delia, Caseres, Ramón, Thomas, Bertone; Framework para el Desarrollo Ágil

de Aplicaciones Web, CACIC 2006, San Luis, 17 al 21 de octubre de 2006.

TRANSFORMANDO DATOS EN CONOCIMIENTO: NUEVAS PRÁCTICAS

Msc. Raúl Oscar Klenzi

Departamento e Instituto de Informática F.C.E.F. y N. Universidad Nacional de San Juan. Ignacio de la Roza y Meglioli (5400). rauloscarklenzi@gmail.com

Msc. María Alejandra Malberti

Departamento e Instituto de Informática F.C.E.F. y N. Universidad Nacional de San Juan. Ignacio de la Roza y Meglioli (5400). amalberti@gmail.com

Msc. Graciela Elida Beguerí

Departamento e Instituto de Informática F.C.E.F. y N. Universidad Nacional de San Juan. Ignacio de la Roza y Meglioli (5400). grabeda@gmail.com

Lic. Silvina Migani

Departamento e Instituto de Informática F.C.E.F. y N. Universidad Nacional de San Juan. Ignacio de la Roza y Meglioli (5400). silvina.migani@gmail.com

CONTEXTO

En la actualidad, la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de San Juan, FCEFNU-SJ, y la Universidad toda, está abocada a la etapa de auto evaluación institucional, destinada a mejorar la calidad de la institución.

Con el propósito de acompañar este proceso, en esta propuesta se realiza la gestión de conocimiento que, entre otros datos, se plasma sobre los pertenecientes a los Sistemas de Biblioteca y de Alumnos de la mencionada facultad. Esto permite, no solo lograr una efectiva transferencia de conocimiento a diferentes cátedras de las carreras Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información pertenecientes al Departamento de Informática, sino también ofrecer a la comunidad distintos cursos de actualización, perfeccionamiento y posgrado.

RESUMEN

La construcción de almacenes de datos – Data Warehouses- DW, la búsqueda de conocimiento inmerso en los datos por medio de OLAP, de estrategias de minería de datos – Data Mining- DM, y su validación mediante herramientas estadísticas, es el área donde el proyecto “Búsqueda estratégica de conocimiento en los datos de Biblioteca y Alumnos de la FCEFNU”, se desarrolla.

Este proyecto es continuación del denominado "Descubrimiento de conocimiento a través de Data Warehousing y Data Mining, en los datos de la Biblioteca de la FCEFNU-SJ" - 2006/2007- y de anteriores proyectos: "Búsqueda inteligente de Información no relacionada en Grandes Bases de Datos" y "Medición de la biblioteca de la FCEFNU-SJ", los cuales abordaron temáticas relacionadas con la actividad de la biblioteca de la FCEFNU. Actualmente, se avanza sobre los resultados alcanzados, al incorporar los datos correspondientes a los alumnos de la FCEFNU.

Otro tema tratado, es el análisis de encuestas de satisfacción de usuarios de la biblioteca de la FCEFNU, por medio de la aplicación de estrategias de DM como: segmentación, clasificación, y descubrimiento de reglas de asociación, que permiten determinar inductores influyentes en la imagen de los usuarios respecto de su biblioteca. En este proceso, se usan y comparan distintas herramientas de software libre.

Palabras clave: Gestión del Conocimiento - Data Mining - Data Warehouse - Satisfacción de Usuarios.

1. INTRODUCCION

Existen organizaciones que poseen una enorme cantidad de datos generados en el día a día, cuya recolección y organización

insumieron tiempo e importantes sumas de dinero y, sin embargo, están como dormidos en distintos depósitos.

Actualmente la tecnología computacional permite no sólo que se capturen y almacenen esas inmensas cantidades de datos, sino también encontrar patrones, tendencias o anomalías en ellos. Resumirlos a través de modelos cuantitativos simples, no es una tarea trivial, ya que deben convertirse los datos en información, y convertir la información en conocimiento, lo que representa uno de los grandes desafíos de la era de la información (Larose, D. 2006), (Hernández Orallo, J. y otros, 2008).

La construcción de DW, la utilización de herramientas OLAP y de DM se presentan como elementos esenciales y complementarios que hacen posible contar por primera vez, con una memoria corporativa que provee información accesible, consistente, adaptable y segura, además de transformarla en conocimiento. Dicha transformación es el cimiento en la toma de decisiones de toda organización. (Kimball y otros, 2004), (Inmon, 2002). Esto provoca considerables beneficios a las organizaciones que la aplican, pues tienen la oportunidad de descubrir patrones rentables, y tendencias novedosas para su negocio.

A partir de ello, la construcción de DW y DM se presenta como una tecnología emergente con varias ventajas:

- Ahorra grandes cantidades de dinero a las empresas y organizaciones en general.
- Abre nuevas oportunidades de negocios.
- Mejora los servicios brindados a los usuarios de cualquier organización ya que, aún en aquellas aplicaciones en las que no se vislumbran importantes beneficios monetarios, existe la posibilidad de incorporar la tecnología de minería de datos.
- Resulta un buen punto de encuentro entre los investigadores y las personas de negocios.

En esta nueva área, relacionada a los Sistemas de Soporte de Decisión, confluyen distintas áreas de conocimiento, tales como: Bases de Datos, Estadística, Inteligencia Artificial, Aprendizaje Automático, Técnicas de Visualización, etc.

Específicamente hablando de DM, el propósito es encontrar modelos matemáticos obtenidos mediante técnicas de aprendizaje, supervisadas o no, que desde los datos rescaten patrones, relaciones, y conocimiento anteriormente no percibido por otras técnicas, y desde allí estimar predicciones sobre las variables analizadas.

DM aplicado a datos inherentes a una biblioteca recibe el nombre Bibliomining (Nicholson, S. y otro, 2003). En el marco de este proyecto el caso de estudio central ha sido la biblioteca de la FCEF-UNSA, desde los datos del movimiento bibliográfico y estructura que la caracterizan, como así también desde la perspectiva que sus usuarios tienen de aquella.

La búsqueda de la satisfacción de usuarios de bibliotecas y la aplicación de técnicas de minería de datos, abarca una primer etapa exploratoria y descriptiva realizada sobre las respuestas obtenidas de una encuesta practicada a los alumnos y usuarios de la biblioteca de la FCEF-UNSA, validada oportunamente en (Beguerí, G., 2006) (Beguerí, G. 2007) y analizadas mediante técnicas de Segmentación, Clasificación, y Reglas de Asociación en (Klenzi, R., 2008) (Malberti, M., 2008). En este proyecto esencialmente se han realizado tareas de descripción, encontrando atributos inductores entre los que conforman la encuesta (Klenzi, R. y otros, 2008) que, conjuntamente con el procesamiento de los datos de circulación de material bibliográfico de diferentes años ha permitido proponer políticas a futuro (redistribuciones presupuestarias, y de material en estanterías, refacciones edilicias, etc.). Similar procedimiento, en cuanto a la satisfacción de usuarios, se está aplicando a encuestas realizadas a usuarios de software libre, intentando encontrar los inductores que permitan inferir ventajas y desventajas de tales softwares según la perspectiva de sus usuarios.

2. LINEAS DE INVESTIGACION y DESARROLLO

En el proceso de desarrollo del proyecto, se realizan tareas de DM, Bibliomining, Text Mining y Web Mining:

- En el DW construido a partir de los datos del departamento alumnos de la FCEFN y de los datos de circulación de material bibliográfico (Migani, y otro, 2006) (Migani, y otros, 2007).
- En encuestas, de satisfacción de usuarios de la Biblioteca de la citada facultad y de usuarios de herramientas código libre.
- En los logs de acceso desde Internet a consulta bibliográfica.

Para ello, se hace uso de las siguientes herramientas:

- R, WEKA-Waikato Environment Knowledge Algorithms y RapidMiner, bajo licencias GNU.
- ORACLE Warehouse Builder y ORACLE Data Miner., del administrador de Base de datos ORACLE , bajo licencia OTN y
- Microsoft SQL Server 2005, en particular SQL Server Business Intelligence Development Studio.

Conjunto a la indagación teórica inherente a cada una de las estrategias de DM abordadas, se realizan análisis de los algoritmos que las integran, y se promueve el análisis comparativo entre las distintas herramientas usadas, tanto en lo referente a la capacidad de las mismas en la búsqueda de conocimiento, como de los algoritmos que proveen. En este aspecto se ha realizado un sucinto análisis comparativo entre herramientas de software libre, aplicables en el área de la minería de datos (Gonzalez, y otros, 2008). Las herramientas analizadas han sido **WEKA - Waikato Environment for Knowledge Analysis**- versión 3.5.8, Universidad de Waikato en Nueva Zelanda con licencia GNU General Public License (GPL) (Witten I. H. & Frank E. 2005), **RAPIDMINER** (Versión Original YALE) 4.2.000 con licencia AGPL versión 3. **R Project** que es un lenguaje y entorno de programación para análisis estadístico y gráfico (R Development Core

Team. 2006). R se distribuye bajo la licencia GNU GPL. Las tres herramientas están disponibles para los sistemas Operativos Windows, Macintosh, Unix, y GNU/Linux.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados alcanzados hasta este momento, han sido publicados y divulgados en diferentes eventos, tal es el caso de:

- **Divulgación De Las Actividades Del Proyecto 21/E824.** Beguerí; Malberti; Klenzi,. Seminario de Bibliotecas Universitarias, Los Cambios de Paradigmas en su Gestión. Universidad Nacional de San Juan- Junta de Bibliotecarios. 29 y 30 de Agosto de 2008.
- **Problemas de Asignación de Aulas, Docentes y Materias en Instituciones Educativas.** Aballay; Klenzi. CACIC 2008.
- **Análisis comparativo entre estrategias de clasificación y reglas de asociación, en la determinación de inductores, en encuesta de satisfacción de usuarios de una biblioteca universitaria.** Klenzi, Malberti, Beguerí. Jornadas Chilenas de Computación 2008, XX Encuentro Chileno de Computación 2008.
- **Minería de datos con herramientas de software libre.** González; Aciar; Martínez. 1° Congreso Nacional de Software Libre-San Juan
- **JOOMLA en la implementación de un prototipo de autoevaluación, en una cátedra universitaria.** Echegaray, Klenzi, Malberti. 5° Jornada de Informática y Educación. U. N. V. M. Argentina.

En particular, en el ámbito de la FCEFN, conjuntamente con otro proyecto de la facultad, se organizó un Workshop de actualización en Telecomunicaciones y Gestión del Conocimiento en Bibliotecas, en el marco del cual se presentaron los siguientes trabajos:

- **Datamarts y Olap en Oracle 11g.** Migani, Lépez, González.
- **Minería de texto. Conceptos y Herramientas.** Gutiérrez, Klenzi
- **Computación Paralela y sus Aplicaciones.** Dra. Piccoli (Invitada)
- **IsisMarc Cliente-Servidor. I+D aplicado a la reingeniería de un producto.** Lic. Marmonti. (Invitado).
- **Catalogo de Bibliotecas en el marco de la web 2.0 y las redes sociales.** Lic. Marmonti.

Actualmente, desde un constructo generado y validado destinado a usuarios de software libre, se realiza el análisis y procesamiento con el propósito de determinar las necesidades implícitas y explícitas de aquellos usuarios.

Asimismo, se espera ampliar el ámbito de aplicación de la tecnología de minería de datos, al análisis de datos sísmicos perteneciente al área de ciencias de la tierra, temática totalmente pertinente a nuestra facultad.

4. FORMACION DE RECURSOS HUMANOS

En lo referente a la formación de recursos humanos, en el marco del proyecto se han desarrollado diferentes tesis de posgrado, de grado y dirección de becarios.

Se realizó la defensa de dos Tesis de Maestría, correspondientes a la Maestría en Informática dependiente de la Universidad Nacional de La Matanza. Ellas son:

- **Aplicación de minería de reglas de asociación en una biblioteca universitaria: una tecnología apropiada.** Lic. A. Malberti
- **Aplicación de minería de datos a la gestión bibliotecaria.** Ing. R. Klenzi

Se dirigió una Tesis de Maestría, correspondiente a la Maestría en Logística dependiente de la Universidad Nacional de Cuyo:

- **Propuesta de Autoevaluación de una Biblioteca Universitaria en un marco de Gestión Logística.** Prof.V. Márquez.

Se dirigieron dos trabajos finales, de la carrera Licenciatura en Ciencias de la Información:

- **Implementación de Autoevaluaciones en entorno WEB.** F. Echegaray.
- **Minería de Datos Aplicada a un DW. Utilización de la Herramienta ORACLE para el desarrollo del DW y la extracción de información oculta.** H. Lépez, A. González.
- **Búsqueda de satisfacción en usuarios de software libre.** Martinez, Aciar (en ejecución)

En la actualidad se dirigen dos Becas de Investigación para alumnos avanzados, de la UNSJ:

- **Proceso de Extracción, Transformación, Carga y Refresco de un DW con datos de provenientes del sistema de alumnos y de biblioteca.** Becario H. Lépez.
- **Aplicación de Minería de Datos a Encuesta de Satisfacción de Usuarios de una Biblioteca Universitaria.** Becario A. González.

En el bienio 2006-2007, se dirigió una beca de innovación tecnológica en las ingenierías, "Digitalización de Archivos históricos" que fue financiada con \$10000 por la fundación Banco San Juan y estuvo a cargo del becario Ing. Cristian de los Ríos.

5. BIBLIOGRAFIA

Beguerí, G., Olguín, L. (2006). *Estudio sobre la Percepción del Usuario en una Biblioteca Universitaria. Normas ISO 11620, IRAM – ISO 11620.* Publicado en: <http://www.uniram.com.ar/jornadas/XXV/TC-14.pdf>

Beguerí, G. (2007) *Logística como garantía de satisfacción del usuario.* Tesis de Maestría-Universidad Nacional de Cuyo.

Gonzalez A., Aciar J., Martinez S. (2008) *Minería de Datos con Herramientas de Software Libre (Análisis Comparativo de*

Herramientas de Software Libre Aplicadas al Área de la Minería De Datos) 1° Congreso Nacional de Software libre. San Juan 2008

Hernández Orallo J., Ramirez Quintana, J, Ferri Ramirez, C. (2008) **Introducción a la Minería de Datos**. Pearson-Prentice Hall.

Inmon (2002). **Building the Data Warehouse Third Edition**. Wiley Computer Publishing

Kimball R., Reeves L., Ross M., Thornthwhite W. **The Data Warehouse Life Cycle Toolkit Expert Methods for Designing, Developing and Deploying Data Warehouses**. John Wiley & Sons.405 pag.

Klenzi, Raúl O. (2008). **Aplicación de minería de datos a la gestión bibliotecaria**. Tesis de Maestría. Maestría en Informática. Universidad Nacional de la Matanza.

Klenzi, R.; Malberti, M. A.; Beguerí, G. (2008) **Análisis comparativo entre estrategias de clasificación y reglas de asociación, en la determinación de inductores, en encuesta de satisfacción de usuarios de una biblioteca universitaria**. Jornadas Chilenas de Computación 2008 (JCC '08) en el marco del XX Encuentro Chileno de Computación 2008 (ECC '08). Punta Arenas- Chile

Larose, D. (2006). **Data Mining Methods and Model**. Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Malberti, María A. (2008). **Aplicación de Minería de Reglas de Asociación en una Biblioteca Universitaria**. Tesis de Maestría. Maestría en Informática. Universidad Nacional de la Matanza.

Migani, S.; Chavez, S. (2006) **Diseño Conceptual de un Data Warehouse**. Revista Ciencias. N°1 2006 Pag. 83 a 90.

Migani, S.; Chavez, S. Klenzi, R; (2007) **Construcción de un Data Warehouse-Parte I**. Revista Ciencias. N°1 2007 Pag. 45 a 56.

Nicholson, S. and Stanton, J. (2003). **Gaining Strategic Advantage through Bibliomining: Data Mining for Management Decisions in Corporate, Special, Digital, and Traditional Libraries**. Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance, H. Nemati and C. Barko, eds. (Hershey, Penn.: Idea Group Pub.).

R Development Core Team. (2006). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.

Witten I. H. & Frank E. (2005). **Data Mining, practical Machine Learning Tools and techniques**. Second Edition. Morgan Kaufmann Publisher.

Enseñanza de Ingeniería de Software en un Escenario Distribuido y Colaborativo

Estela Torres¹, Laura Aballay², Cintia Ferrarini², Sergio Zapata², Cesar Collazos³, Fáber Guiraldo⁴, Sergio Ochoa⁵

¹ Departamento de Informática, Universidad Nacional de San Juan
etorres@iinfo.unsj.edu.ar

² Instituto de Informática, Universidad Nacional de San Juan
{laballay,ferrarini,szapata}@iinfo.unsj.edu.ar

³ Grupo IDIS, Universidad del Cauca, Colombia
ccollazo@unicauca.edu.co

⁴ Grupo SINFOCI, Universidad del Quindío, Colombia
fdgiraldo@uniquindio.edu.co

⁵ Departamento de Ciencias de la Computación, Universidad de Chile
sochoa@dcc.uchile.cl

CONTEXTO

Relación de la línea de I/D presentada con los proyectos del grupo/institución.

El Instituto de Informática (IdeI) alberga un gabinete de ingeniería de software en el cual se vienen ejecutando líneas de I+D desde hace 6 años aproximadamente especialmente orientadas a la mejora de la calidad de software. Se han ejecutado distintos proyectos y actividades, muchas de ellas vinculadas al medio local, y con distinto grado de éxito.

Las líneas de investigación de este gabinete involucraron la participación de varios docentes del IdeI, como también del Departamento de Informática, especialmente del área Estadística de esa unidad. En este marco y en los últimos años se ejecutaron los siguientes proyectos:

- Adaptación de Modelos Internacionales de Procesos de Software a Empresas Locales Productoras de Software (AMIProSoft, 21/E638).
- Red de Investigación Aplicada en Ingeniería de Software Experimental (SPU).
- Mejora de la Capacidad del Proceso de Software en empresas locales (MECAProSoft, 21/E829).

RESUMEN

No más de 200 palabras exponiendo sintéticamente los alcances de la línea de I/D

Palabras clave: ingeniería de software global, aprendizaje colaborativo apoyado por computador, ingeniería de software experimental.

1. INTRODUCCION

Una exposición general sobre el tema de I/D con las referencias bibliográficas relacionadas. Se sugiere no más de 3 páginas.

El crecimiento vertiginoso de las Tecnologías de la Información y Comunicaciones (TICs) está generando nuevas formas de trabajo y modificando diversas prácticas en la vida cotidiana de las personas. En esta transformación tecnológica se observa una tendencia progresiva hacia la colaboración entre personas para alcanzar un objetivo común, donde el trabajo se organiza en equipos y cada integrante interactúa con el resto del grupo para obtener una mejor productividad.

A partir de esta perspectiva, cobra mayor importancia el área de investigación de Trabajo Cooperativo Asistido por Computador (CSCW, por su sigla en inglés), encaminada al estudio del ser humano dentro del contexto de trabajo, así como del diseño de herramientas (groupware) que den soporte al trabajo en grupo [7]. El objetivo no es sólo la mejora de la comunicación, sino también promueve la generación de nuevos paradigmas de interacción, en los que sea posible contar con estrategias que fomenten los componentes

esenciales del trabajo colaborativo, en entornos donde impera la necesidad del trabajo en grupo.

El aprendizaje colaborativo soportado por computador (CSCL, por su sigla en inglés) permite integrar un proceso de aprendizaje de tal manera que el mismo tenga lugar en ambientes mediados por computadores. Para recibir el beneficio completo del aprendizaje social, los alumnos deben interactuar unos con otros, compartir información y coordinar acciones. Desafortunadamente, las investigaciones indican que la mediación computacional crea potenciales obstáculos para la interacción alumno-alumno. Específicamente, los miembros de los equipos de trabajo tienden a experimentar un progreso más lento para alcanzar confianza, cohesión, eficacia y conocimiento compartido, los cuales impactan en la interacción efectiva de los estudiantes [3].

Una preocupación para los desarrolladores y diseñadores de herramientas de apoyo al CSCL es cómo fomentar interacciones efectivas alumno-alumno en entornos colaborativos [11].

El desarrollo de software no es ajeno a la perspectiva del trabajo en equipo, y la colaboración es uno de los aspectos fundamentales y destacados. Para que dichos equipos puedan lograr prácticas de trabajo colaborativo adecuadas, que les permitan alcanzar eficazmente sus objetivos, existen diversas metodologías de desarrollo de software. Estas metodologías establecen el marco de referencia para coordinar el trabajo individual de los integrantes del equipo de desarrollo.

Otro rasgo distintivo de nuestra época, especialmente debido a la expansión de Internet, es el desarrollo de software en escenarios geográficamente distribuidos, Ingeniería de Software Global [12]. Aquí los actores del proceso de software están dispersos geográficamente y deben utilizar modelos comunicacionales basados en computadora (chat, videoconferencia, email, gestores de documentos compartidos, etc.) para lograr sus objetivos.

Más allá de lo expuesto, una de las principales tareas del sistema de educación superior moderno es preparar a los alumnos para la

participación en una sociedad de información, donde el conocimiento es un recurso crítico para el desarrollo social y económico, donde la creación de redes sociales, para intercambio de conocimientos es una nueva habilidad. Comunicación efectiva, capacidad de negociación y creación de nuevos conocimientos para evaluar críticamente recursos de información o productos, se encuentran entre las competencias transferibles que la educación superior tiene como objetivo desarrollar en los estudiantes

El aprendizaje colaborativo involucra un trabajo intelectual conjunto que persigue ciertos resultados del aprendizaje, los cuales pueden ser mejores en la medida que los compañeros de aprendizaje otorguen diferentes perspectivas a un problema o tema [1].

Las tecnologías de la comunicación basadas en Internet conceden a los estudiantes la oportunidad de “hablar” o interactuar con sus compañeros de diferentes países, y desarrollar capacidades como las anteriormente citadas.

Una parte importante del proceso es una reflexión sobre el aprendizaje de un concepto, habilidad, o tema, mediante una discusión con otro. El Aprender Juntos es un modelo utilizado en la educación superior para fomentar la reflexión sobre el aprendizaje, ya sea mediante la realización de proyectos conjuntos o por ayudar a los demás a entender el material de aprendizaje. En el diálogo entre los alumnos se pueden obtener múltiples perspectivas, provocando un conflicto cognitivo, se fomenta el desarrollo de habilidades críticas y la capacidad para el debate profesional, la objetividad, y reflexión discursiva [5].

Desde esta perspectiva es válido proponer un modelo que permita soportar el proceso de enseñanza-aprendizaje de temáticas de Ingeniería de Software, usando estrategias colaborativas entre grupos de alumnos geográficamente dispersos. Siendo esta la línea de investigación que se presenta en este trabajo.

Actualmente existen propuestas para la inclusión de modelos colaborativos en la enseñanza de otras áreas del conocimiento, como por ejemplo Inteligencia Artificial [9], Programación [13], y Sistemas Expertos [4], entre otros. De igual forma existen experiencias

tendientes a incorporar nuevos esquemas de enseñanza-aprendizaje para la orientación de cursos relacionados con Ingeniería del Software.

Manjarres y otros, han planteado una estrategia participativa en la impartición del curso “Análisis, diseño y mantenimiento de software”, en la Escuela Técnica Superior de Ingeniería Informática de la UNED (España) [9]. La actividad práctica consiste en la participación en un proyecto real basado en Software Libre, que implica el análisis, diseño y desarrollo de una aplicación de gestión de socios y voluntarios para la organización Ingeniería Sin Fronteras (ISF). Allí los estudiantes se integran en el equipo de desarrollo de la organización, pudiendo realizar su colaboración a distancia. Esta colaboración supone la práctica en diferentes técnicas de ingeniería, a la vez que la interiorización de los valores intrínsecos al paradigma de desarrollo del Software Libre [8].

Mesa y otros, han planteado una estrategia para la enseñanza de Ingeniería de Software desde la perspectiva de PBL (Problem-Based Learning), realizando un trabajo coordinado para la ejecución adecuada del proyecto [10].

Sin embargo ninguna de estas iniciativas considera el trabajo con personas geográficamente dispersas usando modelos colaborativos, que es la base de nuestra investigación.

2. LINEAS DE INVESTIGACION y DESARROLLO

Detallar sintéticamente los ejes del tema que se están investigando.

El objetivo de esta línea de investigación es encontrar un modelo de enseñanza-aprendizaje para Ingeniería de Software en un ambiente colaborativo y distribuido. Este modelo será alcanzado luego de realizar una serie de experimentaciones controladas que nos permitan tomar conocimiento de las principales variables de control que operan en ambientes de esta naturaleza.

Estas experimentaciones se diseñarán de tal forma que también serán útiles para obtener datos preliminares respecto de atributos de técnicas, métodos o herramientas utilizados en Ingeniería de Software. En este sentido un área

de interés, para aplicar estos procesos empíricos, es la elicitación distribuida de requisitos.

Los casos experimentales que buscarán lograr resultados válidos que puedan ser trasladables, aunque sea parcialmente, a la industria se ejecutarán basados en los conceptos de la Ingeniería de Software Experimental [2], la cual provee de un conjunto de herramientas para alcanzar fiabilidad estadística de estos procesos empíricos.

3. RESULTADOS OBTENIDOS/ESPERADOS

Exponer los resultados alcanzados y los objetivos en curso.

Se diseñó y ejecutó un experimento controlado en conjunto con la Universidad de Chile y Universidad del Cauca (Colombia). En el mismo participaron alumnos de grado de las carreras de informática de las tres universidades y se utilizaron herramientas de comunicación IP estándares (Skype, Conference XP, emails, etc.) para la enseñanza de la temática “Estimación de Tamaño de Software”. Si bien preliminarmente se puede decir que el experimento fue exitoso, todavía se están procesando los datos estadísticos para arribar a conclusiones más fundadas.

Dentro del marco de un trabajo final de carrera, se diseñó y ejecutó un experimento controlado, donde participaron alumnos de segundo y quinto año de la carrera de informática de la Universidad Nacional de San Juan, para comparar dos técnicas de testing funcional. De este experimento, que se realizó siguiendo la metodología de colaboración LuCoA [6], se obtuvieron resultados (indicios, hallazgos) tanto respecto de las técnicas comparadas como del proceso de aprendizaje colaborativo.

Para el presente año se planifica continuar realizando experimentos controlados que permitan adquirir mayor conocimiento en las áreas de investigación abordadas. Se espera realizar una experiencia colaborativa conjunta con la universidad de Chile, Universidad del Cauca (Colombia), Universidad del Quindío (Colombia) y Universidad Tecnológica de Panamá. Esta actividad está dentro del marco del proyecto Fortalecimiento de Red de Investigación Aplicada en Ingeniería de Software Experimental que coordina el Instituto de Informática de la UNSJ y financia la

Secretaría de Políticas Universitarias de Gobierno Nacional.

Se encuentra en proceso de diseño un experimento a ejecutarse próximamente sobre técnicas de elicitación de requisitos en un escenario geográficamente distribuido.

En todos los casos se espera aplicar la metodología de aprendizaje colaborativo LuCoA con el fin de realizar ajustes perfectivos.

4. FORMACION DE RECURSOS HUMANOS

Exponer los resultados obtenidos/esperados y la relación con otros grupos del país y/o del exterior en la formación de recursos humanos.

Se han concluido exitosamente dos trabajos finales de graduación en la carrera de Licenciatura en Sistemas de Información de la UNSJ. Actualmente están en proceso de desarrollo dos trabajos finales que se estiman concluirán en el presente año.

Dentro del marco de la presente línea de investigación se están desarrollando dos tesis de maestrías (Universidad de la Matanza) en las áreas de Elicitación Distribuida de Requisitos y Metodologías Colaborativas Aplicadas a la Educación Superior.

Están avanzadas las gestiones administrativas para llamar a convocatoria de postulantes a becas doctorales dentro del marco del proyecto PRH 2007 (Proyecto de Radicación de Investigadores y Formación de Doctores en Áreas Estratégicas).

5. BIBLIOGRAFIA

Un número razonable de citas relacionadas con el proyecto. (se sugiere no más de 30 referencias).

[1] Alavi, M. Computer-mediated collaborative learning: An empirical evaluation. *MIS Quarterly*, 18(2), 150-174. 1994.

[2] Basili, V.R., Rombach, D., Schneider, K., Kitchenham, B., Pfahl, D. and Selby, R. *Experimental Software Engineering Issues: Assessment and Future*, Dagstuhl seminar, Germany, LNCS 4336, Springer-Verlag, 2007

[3] Collazos, C., Ochoa, S., Mendoza, J., Mejorando los esquemas de evaluación a través de procesos de colaboración. *Revista Educación y Educadores* 10(1), 2007.

[4] Cuneo, C., Mariño, M., Entorno colaborativo en la enseñanza de Sistemas

Expertos. *Comunicaciones Científicas y Tecnológicas*, 2005.

[5] Falchikov, N. *Learning together: Peer tutoring in higher education*. London: Routledge Falmer. 2001.

[6] Ferrarini, C., Collazos, C. *Metodología LuCoA para el Aprendizaje Lúdico Colaborativo en la Educación Superior*. Congreso Colombiano de Computación, Medellín, Abril de 2008.

[7] Grudin, J. *Computer-Supported Cooperative Work: History and focus*. *IEEE Computer*, 27(5), 19-26. 1994.

[8] Manjarres, A., Arias, M., Gaudio, E. *Competencias transversales en la enseñanza de la ingeniería del software* Actas de las VI Jornadas de Redes de Investigación en Docencia Universitaria, Universidad de Alicante, 9-10 Junio 2008.

[9] Mariño, S., *Diseño de un entorno virtual de enseñanza-aprendizaje para la asignatura de Inteligencia Artificial*, *Quaderns Digitals*. Revista electrónica No.3, 2008.

[10] Mesa, J., Alvarez, J., Villanueva, J., Cos, F. *Actualización de Métodos de Enseñanza-Aprendizaje en Asignaturas de Dirección de Proyectos de Ingeniería*. *Formación Universitaria*, 1(4), pp.23-28, 2008.

[11] Orvis, K., Lassiter, A. *Computer-Supported Collaborative Learning: Best Practices and Principles for Instructors*. Information Science Publishing, Hershey, New York. 2007.

[12] Petkovic, D., Thompson G., Todtenhoefer, R. *Teaching practical software engineering and global software engineering: evaluation and comparison*. *Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education*. ACM, Bologna, Italy, 2006

[13] Redondo, M., Mendes, A., Ortega, M., *Planificación colaborativa del diseño para el aprendizaje de la programación*, Actas del TISE 2001.

SISTEMA DE INFORMACIÓN GEOGRÁFICA APLICADO A TURISMO Y PATRIMONIO HISTÓRICO Y CULTURAL

de San Pedro Maria Eugenia, Serón Natalia, Cristian Montenegro

Laboratorio de Tecnologías Emergentes (LabTEem)

Unidad Académica Caleta Olivia

Universidad Nacional de La Patagonia Austral

(9011) – Ruta 3 s/n Acceso Norte - Caleta Olivia – Santa Cruz

e-mail: { edesanpedro, nseron, cmontenegro }@uaco.unpa.edu.ar

CONTEXTO

Esta línea de Investigación se inicia con el proyecto de vinculación tecnológica que la Universidad desarrolló para la Municipalidad de Caleta Olivia, “*Sistema de Información Geográfica de la ciudad de Caleta Olivia*”, y que continuará su investigación, desarrollo e implementación, a través de un nuevo proyecto que abarcará el estudio del Patrimonio Histórico y Cultural de la zona norte de la provincia de Santa Cruz. Dicho proyecto se desarrollará en el marco del Laboratorio de Tecnologías Emergente (LabTEem).

RESUMEN

El Proyecto tiene como objetivo aplicar las nuevas tecnologías de la información al estudio del Patrimonio Histórico y Cultural y obedece tanto a un interés científico como cultural.

Este Sistema pretende brindar información fehaciente y certera acerca de las actividades Turísticas y Culturales que se pueden realizar en el Departamento Deseado de nuestra Provincia, actividades que reconforten al turista en un todo, ya sea física como intelectualmente.

Palabras clave: *Patrimonio Histórico y Cultural, Turismo, Sistema de Información Geográfica, GIS, georeferenciación.*

1. INTRODUCCION

Un Sistema de Información Geográfica (SIG o GIS, en su acrónimo inglés), es una integración organizada de hardware y software, datos geográficos y personal, diseñado para capturar, almacenar, manipular, analizar y desplegar en todas sus formas la información geográficamente referenciada con el fin de resolver problemas complejos de planificación y gestión. También puede definirse como un modelo de una parte de la realidad referido a un sistema de coordenadas terrestre y construido para satisfacer unas necesidades concretas de información.

El Sistema de Información Geográfica separa la información en diferentes capas temáticas y las

almacena independientemente, permitiendo trabajar con ellas de manera rápida y sencilla, y facilitando al profesional la posibilidad de relacionar la información existente a través de la topología de los objetos, con el fin de generar otra nueva que no podríamos obtener de otra forma.

El SIG se utiliza para resolver investigaciones complejas, para los problemas de manejo, y para la planeación.

Los componentes principales del SIG son: una interfase del usuario, la capacidad para el manejo de la base de datos del sistema, la creación de base de datos, la capacidad para la captura de datos, el paquete de la manipulación y el análisis de datos espaciales y la demostración de la función de generación. [1], [2].

Las áreas de aplicación de un SIG, son entre otras: Cartografía automatizada, Medio ambiente, Equipamiento social, Recursos mineros, Ingeniería de Tránsito, Demografía, GeoMarketing, Banca, Planimetría, Cartografía Digital 3D, Infraestructura, Gestión territorial y en nuestro caso Turismo y Gestión de Patrimonio Histórico y Cultural. [4]

En general, los SIG se diferencian dos componentes fundamentales:

- Un modelo de datos en el que se almacenan las características de los objetos geográficos, de manera similar a como se almacenan en una base de datos convencional (de hecho, muchos sistemas emplean sistemas de gestión de base de datos convencionales), junto con información posicional (coordenadas) y las relaciones entre los distintos objetos (qué está conectado a qué, o junto a).

- Una colección de funciones que nos permiten interrogar a la base de datos y obtener respuestas, ya sea en base a listados o a imágenes (mapas).

Una característica esencial de los sistemas de información geográficos, es que intentan capturar en su modelo datos de la realidad, y no una imagen determinada de ésta. [5]

Los SIG, cuyos antecedentes datan de varias décadas, se han posicionado como una tecnología

básica, imprescindible y poderosa para capturar, almacenar, manipular, analizar, modelar y presentar datos espacialmente referenciados. Se trata por lo tanto, de una categoría dentro de los sistemas de información que se especializa en manejar datos espaciales, con las particularidades y requerimientos que ello conlleva. Cabe aclarar que un SIG no es un programa de PC, ni un software tipo CAD. Lo específico de un SIG reside en rasgos tales como su capacidad para almacenar grandes masas de información georeferenciada (operación de obtener y asignar coordenadas geográficas a una información que carece de ella. Suele referirse para situar imágenes de la Tierra o eventos asociados a direcciones postales) o su potencia para el análisis de la misma, que le hacen idóneo para abordar problemas de planificación y gestión, es decir para la toma de decisiones. [3]

El sistema permite separar la información en diferentes capas temáticas y las almacena independientemente, permitiendo trabajar con ellas de manera rápida y sencilla, y facilitando la posibilidad de relacionar la información existente a través de la topología de los objetos, con el fin de generar otra nueva capa que no podríamos obtener de otra forma. [6], [7]. Un Sistema de Información Geográfica puede mostrar la información en capas temáticas para realizar análisis complejos.

Entender los recursos patrimoniales desde una perspectiva actual, comprehensiva, implica el desplazamiento desde una visión tradicional del patrimonio, fundamentalmente ligada al campo estético y limitada a la obra de arte singular o al excepcional producto de la naturaleza y la cultura, hacia un enfoque más amplio, en el que confluyen también razones de carácter ético, científico, social y pedagógico.

Ello supone ir transitando desde la consideración única y exclusiva de los monumentos como producto histórico sobresaliente, a considerar sus entornos y posteriormente a incorporar los conjuntos urbanos, valorando significativamente más la dimensión histórica que la artística en la apreciación del patrimonio cultural.

Asimismo, hay un paulatino reencuentro de las dimensiones natural y cultural del patrimonio, que durante mucho tiempo han discurrido por sendas paralelas. Se trata de una aproximación de doble sentido: por un lado, mediante la consideración de los paisajes culturales como expresión valiosa patrimonial en el que el modelado del medio físico natural, a través fundamentalmente de las actividades rurales, da lugar a unos productos históricos de gran valor social, cultural y ambiental; por otro lado está la constatación, de que la presencia humana ha modificado en menor o mayor grado la inmensa mayoría de los espacios naturales existentes, contribuyendo a modelar e introducir variaciones sobre las características y fisonomía de dichos espacios.

La visión tradicional del patrimonio cultural o natural ha sido, reconocerlo como tal, establecer las correspondientes tutelas, y posteriormente desarrollar las condiciones de su utilización, fundamentalmente las turísticas y recreativas. En la actualidad supone una visión más amplia e integrada de los valores patrimoniales, aportar una visión del funcionamiento conjunto de los recursos naturales y culturales en el territorio, desplazando el punto de mira desde la óptica de la protección a la de su valoración adecuada y sostenible.

Esto hace necesario profundizar en el conocimiento de esta cuestión y desarrollar metodologías adecuadas para reconocer y evaluar aquellos elementos del patrimonio que presentan una potencialidad para su puesta en valor, especialmente como activos para el desarrollo de las comunidades. [9].

Los sistemas de información geográfica constituyen un potente instrumento metodológico que sirve no sólo para abordar un adecuado tratamiento teórico-empírico de la cuestión, sino que es posteriormente de enorme utilidad para la gestión y puesta en valor de los recursos patrimoniales, por parte de las comunidades implicadas.

El agotamiento e impacto socio-ambiental negativo de los modelos turísticos convencionales caracterizados por la masificación y por una oferta reducida de "sol y playa", junto con el desarrollo creciente de una nueva demanda sensible al medio ambiente y a la diversidad Cultural, convierten al turismo Cultural en una actividad en crecimiento.

Hoy en día, el objetivo Cultural y el disfrute de la naturaleza dentro de un viaje, ha cobrado gran importancia al momento de tomar la decisión de concretarlo.

Por tales motivos, este proyecto se plantea como desafío, brindar una mayor importancia a nuestro Patrimonio Histórico y Cultural con el objetivo de conseguir que éstos lleguen a ser los atractivos que busca el turista y de esta manera favorecer a la creación de nuevos empleos en aquellas zonas con una riqueza importante de recursos naturales y Culturales, que logren aprovechar sus potencialidades, adecuando su territorio como un espacio turístico capaz de llamar la atención al visitante, con una oferta de calidad, y con una identidad propia y diferenciada.

2. LINEAS DE INVESTIGACION y DESARROLLO

El trabajo que viene llevando a cabo este grupo de trabajo, se desarrolla en el ámbito de la vinculación tecnológica dentro de la Universidad. Esto lleva una actividad de investigación importante, para poder desarrollar las actividades planteadas dentro del

proyecto. Fue necesario realizar un estudio que permitió conocer las bondades y defectos de las aplicaciones existentes de SIG, para poder definir la que se usaría como base de la nueva aplicación. [8]

A raíz de este desarrollo, el grupo participará de un proyecto en conjunto con un grupo de la Unidad Académica Río Gallegos y la Subsecretaría de Planeamiento de la Provincia de Santa Cruz, con el objetivo de desarrollar una herramienta útil para la gestión de información catastral en los diferentes municipios de la provincia, así como en la propia dirección provincial de catastro, mediante alguna variante de tipo concentradora de informaciones catastrales locales.

Este proyecto estará relacionado con un proyecto de investigación que viene trabajando hace tiempo en el estudio del Patrimonio Turístico, Histórico y Cultural de la zona. La participación con ese proyecto, se abordará desde el aporte tecnológico brindado para poder visualizar a través de un SIG, la información recolectada.

En primera instancia, se realizará un relevamiento de los potenciales atractivos que se buscarán incluir en la oferta turística del Departamento Deseado con la información recolectada de datos existentes. Para dicho fin, se implementaran las siguientes tareas:

- Investigación secundaria de la literatura publicada para comprender la historia humana del Departamento, y referir dicha información a todas las áreas del Proyecto.

- Redacción de una lista de recursos Culturales para ser incluida en el inventario. Cuando nos referimos a darle un mayor valor a los atractivos el Patrimonio Histórico y Cultural no sólo hablamos de museos, obras de arte, artesanías o paisajes., hablamos de sucesos que pueden ser importantes para la localidad: industrias que se encuentren en la actualidad funcionando o que en algún momento participaron en el desarrollo local. También se deben considerar como parte del Patrimonio Cultural y como posibles atractivos turísticos a las comidas, fiestas populares del lugar o la región entre otras muchas cosas. Con respeto al paisaje, constituye el atractivo natural de la región por sí solo, pero sus componentes, árboles, plantas, musgos, flores, aves y animales en general pueden constituirse en los elementos que atraigan al turista.

- Entrevistas con profesionales idóneos y con residentes que aporten más y mejor información sobre los relevamientos existentes y recursos conocidos.

- Identificación de relevamientos, inventarios y otras listas existentes, con información sobre los bienes del Patrimonio Cultural dentro del área de estudio.

- Estructurar la base de datos del Proyecto en un Sistema de Información Geográfica (SIG) con la información de los relevamientos existentes y la información que se obtenga de futuros trabajos de campo.

- Utilización de un Sistema de Información Geográfica.

Con la gran variedad de datos que se recolectarán y luego de la estructuración de la Base de Datos del Proyecto, todos los usuarios del Sistema podrán tener una aproximación al paisaje que da contexto a un sitio a través de las fotografías, las imágenes panorámicas, los videos, las fotografías aéreas y toda la información textual almacenada.

La implementación de este Sistema servirá como referencia, por ejemplo, para proyectos de restauración y conservación, para administración y mantenimiento de los sitios, para servir como base de estudios sobre impactos ambientales o planificación urbana, y como registro digital permanente para el futuro. Esta base de datos también representa una manera de introducir este rico Patrimonio Cultural en la educación primaria y secundaria.

Cabe destacar que la información del Sistema, se podrá actualizar regularmente o agregar resultados de investigaciones recientes.

Concluyendo, este proyecto supone un importante desafío técnico y metodológico. En muchos aspectos se trata de un proyecto pionero al aplicar tecnología de gran importancia para el estudio y difusión del Turismo y Patrimonio Histórico y Cultural. Si bien el objetivo de estudio es el Departamento Deseado de la Provincia de Santa Cruz, la aplicación desarrollada servirá también para el análisis y la gestión de otros lugares de interés.

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto se llevará a cabo cubriendo actividades para las siguientes etapas, en las cuales se pretende lograr:

- *Análisis y Diseño:* Luego de la investigación y estudio, se pretende en esta etapa, poseer un relevamiento e inventario con datos certeros del Patrimonio Histórico y Cultural con el que cuenta el Departamento Deseado de la Provincia de Santa Cruz. Identificando a su vez, los lugares Turísticos de la zona. Luego, con el material obtenido de dicha investigación, se trabajará sobre las bases de datos correspondientes.

- *Desarrollo:* Generar todas las interfases de usuario que se hayan estimado en la etapa de recolección de Requerimientos Funcionales del Producto. Generar una base de datos robusta y eficiente. Obtener un mapa de la zona de trabajo.

- *Implementación:* Generar un Sistema de Información Geográfica, de gran utilidad, rico en información y que logre satisfacer todas las necesidades del usuario final.

Este proyecto puede ser utilizado por todos aquellos organismos que tengan relación directa con la Información Turística, Patrimonio Histórico y Cultural de la región, etc.

Además esta herramienta puede ser adaptada para luego incorporarse a estudios de Investigación de Patrimonio de otras regiones de la provincia, e ir ampliando la información a todo el territorio nacional.

4. FORMACION DE RECURSOS HUMANOS

Este proyecto permitirá brindarle a un integrante, la posibilidad de ir delineando la temática a abordar, para el desarrollo del Proyecto Final que le permitirá obtener el título de Ingeniero en Sistemas.

Dos alumnos avanzados de la Ingeniería en Sistemas se incorporarán al proyecto, a través de las pasantías internas que ofrece la Universidad para que los alumnos puedan iniciarse en la investigación de diferentes áreas de interés.

5. BIBLIOGRAFIA

- [1] Frassia Mercedes "El GIS en acción " Curso online pág. 9 – 10
- [2] George B. Korte P, "The GIS book", Fifth Edition Onword Press
- [3] Moreno Jiménez A., Cañada Torrecilla R., Cervera Cruañes B., Fernandez García F., Gómez García N., Martínez Suárez P., Prieto Flores M., Rodríguez Esteban J., Vidal Domínguez M. "Sistemas y Análisis de la Información Geográfica. Manual de autoaprendizaje con ArcGIS" Alfaomega Grupo Editor, México, Julio 2006
- [4] www.geotools.org: portal dedicado a la librería geotools1 y geotools2
- [5] Rubio Barroso "Los Sistemas de Información Geográficos: Origen y perspectivas" Revista general de información y documentación, ISSN 1132-1873, Vol. 7, N° 1, 1997, Págs. 93-106
- [6] Sistemas Abiertos de Información Geográfica – Manual de Usuario "Conceptos". Pág.1. SAIG S.L. Gines, Sevilla, 2006
- [7] <http://es.wikipedia.org/>: Portal Web dedicado a brindar información de diversos tópicos.
- [8] Serón N., Montenegro C., Vidal P., Villagra S., Orozco S., Valdéz J., Díaz F., de San Pedro M. – Implementación de un Sistema de Información Geográfica en Municipios de la Zona Norte de la Provincia de Santa Cruz – 10° Workshop de Investigadores de Ciencias de la Computación WICC'08, pp. 530-533 ISBN 978-950-863-101-5, Universidad Nacional de La Pampa, General Pico, La Pampa – Mayo 2008.
- [9] Perota S. – Propuesta de un SIG para la gestión del Patrimonio Natural y Cultural – Valles Calchaquíes en la provincia de Salta - VII Jornada "Técnicas de Restauración y Conservación del Patrimonio", LEMIT , La Plata, 2008.

MODELO DE CASOS DE USO - UN EJE PARA EL PROCESO DE DESARROLLO DE SOFTWARE

María Inés Lund, Laura Aballay, Cintia Ferrarini, María Romagnano
Instituto de Informática – Fac. Cs. Exactas F. y Naturales – Universidad Nacional de San Juan

CONTEXTO

El proyecto se encuentra inserto dentro del Gabinete de Ingeniería de Software del Instituto de Informática y está relacionado con las cátedras Diseño de Software y Sistemas de Información II de las carreras de informática de la FCEF N de la Universidad Nacional de San Juan.

RESUMEN

El propósito de este proyecto es efectuar una recopilación, estudio y análisis de los usos existentes de los casos de uso, de autores reconocidos en el ambiente internacional de la Ingeniería de Software y del desarrollo orientado a objetos. Se analizarán los usos o extensiones de uso que se les dan a los modelos de casos de uso, para lograr una descripción completa y bien definida (formalización) de cada uno de ellos, permitiendo que los modelos de casos de uso sean un eje que acompañe y ayude a todo el proceso de desarrollo de software.

Un modelo de caso de uso es una notación que permite especificar y documentar requisitos, planificar el proyecto de software, tales como fechas de entregas, tamaño del software, también ayudan al equipo de diseño, al diseño de interfaz de usuario y de pruebas del sistema.

(No más de 200 palabras exponiendo sintéticamente los alcances de la línea de I/D)

Palabras clave: casos de uso, análisis, modelos de casos de uso, requisitos, proceso de desarrollo de software.

1. INTRODUCCION

Los proyectos de desarrollo de software suelen fallar, por entregas tardías, mayor presupuesto del original, a causa de la falta de concordancia entre las necesidades reales de los usuarios finales del sistema y los requisitos elicitados [Kotonya&Sommerville 2002]. En general esto no se debe a que el grupo que trabaja sea ineficiente o aplique escasas técnicas de ingeniería, sino como consecuencia de que los requisitos definidos para el sistema no son lo suficientemente claros.

Los requisitos del sistema definen lo que se requiere que haga el sistema y las circunstancias bajo las cuales debe operar. Es decir, definen lo que el sistema se compromete a proveer y el conjunto de restricciones operativas.

Existe una audiencia variada para los documentos de requisitos; esta audiencia o lectores de los documentos son todos los interesados (stakeholders) del sistema, que tienen distintas intenciones y diferentes necesidades con respecto al documento de requisitos, entre los cuales se pueden encontrar:

- Los clientes del sistema: son quienes proporcionan las especificaciones del sistema y quienes chequean si sus necesidades se encuentran correctamente plasmadas. Ellos son los que solicitan cambios en los requisitos.

- Los administradores del proyecto de software: usan el documento de requisitos para planear el tamaño del sistema y planificar el proceso de desarrollo de software.
- Los ingenieros de sistemas: usan los requisitos para entender las necesidades del cliente/usuario y así desarrollar un sistema acorde a esas necesidades.
- Los ingenieros de pruebas del sistema: usan los requisitos para elaborar y llevar a cabo las pruebas del sistema.
- Los ingenieros del mantenimiento del sistema: usan los requisitos para entender el sistema y las relaciones entre las partes. [Kotonya&Sommerville 2002 pp. 6-16].

Una buena especificación debe incluir requisitos que sean técnicamente factibles, medibles, rastreables, no ambiguos y consistentes.

Muchas personas escriben casos de uso, para especificar requisitos, tanto para sistemas de software como para procesos de negocio. Pareciera que escribir casos de uso es una tarea sencilla, pero ¿Cuánto escribir?, ¿Qué detallar?, ¿Cómo escribirlo?. No es lo mismo decir que un caso de uso refleja correctamente la realidad que intenta representar, a decir que se ve bien.

En forma simple, se puede decir que un caso de uso es una porción de funcionalidad del sistema, desde la perspectiva del usuario, para obtener un resultado de valor. Tiene una notación gráfica y está acompañado de una descripción de lo que hace.

El conjunto de casos de uso especifican todas las formas existentes de usar el sistema, y reciben el nombre de Modelo de Casos de Uso.

En la World Wide Web, se encuentra gran cantidad de información relacionada a los casos de uso, y autores con diferentes perspectivas de aplicar o de emplear los casos de uso.

Estas diferentes 'vistas' o 'perspectivas' de los casos de uso pueden confundir al lector ya que, a pesar de tener el mismo origen (Jacobson,1967), se les fue dando diferentes usos, aplicaciones y extensiones de acuerdo a las necesidades y criterios. Esto también surge, tal como lo afirman Kotonya & Sommerville, y explicado anteriormente porque los lectores (stakeholders) de la especificación de requisitos son diferentes y con necesidades de información distinta.

Haciendo una breve reseña, los casos de uso surgen de los *escenarios* propuestos por Jacobson (1967). Estos representan distintos caminos posibles de ejecución de una funcionalidad, y de las propuestas de *conversaciones* de Rebecca Wirf-Brock (1988), en las que se distingue el diálogo entre el actor primario y el sistema, reflejando -en dos columnas- las acciones de cada uno de ellos, generalmente escritas para preparar el diseño de la interfaz del usuario, con gran detalle. Constantine y Lockwood, impulsores de los métodos ágiles -centrados en el usuario- para el desarrollo de software, adoptan este formato de conversación en lo que ellos llaman casos de uso esenciales, en donde todos los detalles de movimientos del usuario relacionados a la interfaz son omitidos, resultando descripciones mucho más cortas [Constantine&Lockwood, 2000].

En los 80, los casos de uso fueron incorporados en la comunidad de programación orientada a objetos, en donde llegaron para llenar un vacío existente en el proceso de requisitos. Alistair Cockburn (1994) introdujo el modelo conceptual de 'actores' y 'objetivos'. También en 1994, Booch, Rumbaugh y Jacobson, crean el Lenguaje de Modelado Unificado (UML) para los procesos de desarrollo de sistemas de información, el que en noviembre de 1997 se transforma en un estándar. Ivar Jacobson y otros en el libro *Object-Oriented Software Engineering: A Use Case Driven Approach*, refieren a los casos de uso para *representar lo que los usuarios deberían poder hacer con el sistema. Cada caso de uso es un curso de eventos completo, en el sistema, desde la perspectiva del usuario e iniciado por un actor. Es una forma específica de usar el sistema para realizar alguna porción de funcionalidad. Un caso de uso es entonces una secuencia especial de transacciones relacionadas, realizadas por un actor y el sistema, en un diálogo* [Jacobson et al.1998, pp157-159].

Algunos autores, reconocidos en el ambiente de la ingeniería de software orientada a objetos, también dan sus propias definiciones de casos de uso, siempre en consenso con la definición de Jacobson.

Si bien los casos de uso tienen una notación visual gráfica, deben estar acompañados de una descripción, generalmente en formato de texto u oraciones estructuradas, aunque también pueden usarse diagramas de flujo, diagramas de secuencia, Redes de Petri, o lenguajes de programación. Normalmente, los casos de uso sirven como medio de comunicación entre personas, sin que ellas deban tener algún entrenamiento en especial sobre 'cómo leerlos'. Por lo tanto, el texto simple suele ser la mejor opción.

Aprender a leer un caso de uso es aparentemente sencillo, aprender a construir adecuadamente casos de uso es una tarea ardua, y de ello depende una buena especificación de requisitos de software.

El modelo de casos de uso emplea el concepto de actores, para representar los roles que los usuarios pueden jugar en el sistema y los casos de uso para representar lo que los usuarios pueden hacer con el sistema.

Los casos de uso son el único concepto incluido en el enfoque de Jacobson en su libro de Ingeniería de Software Orientada a Objetos (OOSE) para modelar requisitos, pero cualquier persona sin conocimientos sobre orientación de objetos podría leer y entender el modelo.

El modelo de requisitos [Jacobson et al.1998, pp154-155] es la especificación de requisitos con la perspectiva del usuario, cuyo objetivo es delimitar el sistema y definir su funcionalidad. Para esto se desarrolla un gráfico conceptual del sistema, usando objetos del dominio del problema y también descripciones específicas de interfaz del sistema, si éstas son significativas para él. Se describe al sistema como un conjunto de casos de uso que son ejecutados por un conjunto de actores. Los actores se constituyen fuera del sistema y los casos de uso toman lugar dentro del sistema.

El modelo de requisitos debería funcionar como un contrato entre los desarrolladores y el cliente que solicita el sistema. Es la imagen para el desarrollador, de lo que quiere el cliente.

El modelo de requisitos luego es estructurado a través del modelo de análisis, realizado por un modelo de diseño, implementado por un modelo de implementación y probado por un modelo de pruebas. Estos modelos no solo son verificados sobre el modelo de requisitos, sino que son desarrollados directamente desde él.

2. LINEAS DE INVESTIGACION y DESARROLLO

- Identificar las características o elementos mínimos que debiera contener la descripción de los casos de uso, para poder dar trazabilidad de los requisitos del Documento de Requisitos del Sistema (SRD) o Documento de Especificación de Requisitos (ERS) a los casos de uso. Esto es, poder identificar cada requisito del usuario y saber qué caso de uso solucionará su necesidad. , aún en proceso de estudio.
- Encontrar indicadores de evaluación, en el modelo y la descripción de casos de uso, que permitan determinar si el diseño resultante (diagrama de clases) cumple con las necesidades reales ó da solución al problema planteado, y si está correctamente definido. En esta línea se está trabajando con información proveniente de las producciones de los alumnos avanzados de la cátedra Diseño de Software y Sistemas de Información II, de Licenciatura en sistemas de Información y Licenciatura en Ciencias de la Computación de la FCEF.N.

3. RESULTADOS PARCIALES OBTENIDOS

- Recopilación en Internet sobre el Modelo de Casos de Uso y perspectivas de uso en el desarrollo de Software.

- Análisis de la información relevada sobre casos de uso y su aplicación en distintas etapas del desarrollo de software.
- Estudio comparativo entre la concepción del Modelo de Casos de Uso y sus objetos entre los autores Larman, Amour-Miller, Weitzenfeld, Jacobson, Booch y Rumbaugh.
- Análisis de los ítems de evaluación relevantes en la construcción de un Modelo de Casos de Uso.
- Extracción e identificación de Indicadores de valoración del Modelo de Casos de Uso.
- Pruebas de valoración con esos indicadores.

Divulgación:

2008. Conferencia sobre “UML – Modelo de Casos de Uso”, destinada a alumnos y profesores de la Universidad del Cauca, Colombia, en el marco del proyecto Fortalecimiento de Redes Interuniversitarias II – de SPU. Disertante: María Inés Lund.

2008. “Experiencia Preliminar Incubación de Empresas Locales desarrolladoras de Software”, trabajo aceptado en el III Congreso Internacional: "El Aporte de la Universidad al Desarrollo Sostenible" – Univ. Católica de Cuyo – Instituto de Desarrollo Sostenible. San Juan. Autores: Pedro Zárate, Sergio Zapata, María Inés Lund, Emilio Ormeño.

2009. “Indicadores para la Evaluación de Diseños de Sistemas a través del Modelo de Casos de Uso en la Educación Superior”. Trabajo aceptado en el ICECE'2009 - International Conference on Engineering and Computer Education. Autores: María Inés Lund, Cintia Ferrarini, Emilio Ormeño y Sergio Zapata.

2009. “A CMT-based Modeling Language for Courseware Design” en el CSCWD 2009. esponsorado por la IEEE. Autores: Emilio Ormeño, Sergio Ochoa, Francisco Ibañez, M. Inés Lund, Susana Ruiz, Laura Aballay, Víctor Rosales.

4. FORMACION DE RECURSOS HUMANOS

Año 2008: Aprobación del Curso de Posgrado –“USO DE LA ESTADISTICA EN LA INGENIERIA DE SOFTWARE”. Instituto de Informática. Res N° 56/07-CD-FCEFN. Destinado a profesionales egresados en carreras relacionadas a la Ingeniería de Software. Laura Aballay, María Romagnano y María Inés Lund.

Año 2009: Aprobación del Curso de Capacitación en lenguaje Java. Cintia Ferrarini.

En ejecución Tesis de Maestría en Informática:

Metodología LuCoA para el Aprendizaje Lúdico Colaborativo en la Educación Superior. Cintia Ferrarini Oliver. Directores: Cesar Collazos, Mónica González de Doña.

Especificación de Requisitos y Uso de los Casos de Uso. María Inés Lund. Directores: Emilio Gustavo Ormeño y Fernando Pinciroli.

Mtdología para Modelar Procesos de Negocio. Maria G. Domagnano. Directores: Adriana De Luca y Emilio G. Ormeño

Asesoramiento de Tesis de Licenciatura:

2008. “Sistema de Gestión de solicitudes y proceso de adquisición de Obras Bibliográficas” – Desiderata. Autores: J.M.Blanco, A.Orellana Vassallo.

2008. “Propuesta de Desarrollo de Software adaptado a Pequeñas Organizaciones Locales de Software” - Aplicación a un caso de estudio: SIGEPRI. Autores: M.J. Sarmiento, V.E.Lucero.

2009. “Evaluación de Técnicas de Testing Funcional aplicando Ingeniería de Software Experimental”. Autores: A.Soto, Sánchez.

En ejecución: “Evaluación experimental de Técnicas de Elicitación Distribuida de Requisitos”. Alumnos: L.Montaña, M.Reus.

“Ciudad Digital: Sistema Web de Administración de Programas Sociales para el municipio de 25 de Mayo”. Alumno: Héctor Aravena.

Incorporación de un alumno con cargo auxiliar de 2da categoría, al cual se lo está formando en la temática general del proyecto, y capacitando en la búsqueda y evaluación de información en Internet, en adhesión con la competencia informacional “Aprender a buscar y evaluar información” propuesta en [Pinto 2009], en donde se pretende básicamente definir las principales secuencias y estrategias para llevar a cabo una búsqueda precisa y pertinente, para evaluarlos.

5. BIBLIOGRAFIA

Referencias:

[Armour&Miller 2001] Advanced use case modeling. Addison-Wesley © 2001. Frank Armour, Granville Miller.

[Booch et al. 1999] El lenguaje unificado de modelado. Addison-Wesley. 1999. Grady Booch, James Rumbaugh, Ivar Jacobson

[Cockburn 2005] Writing Effective Use Cases. Addison-Wesley. © 2001, Ed. 2005. Alistair Cockburn

[Constantine&Lockwood 2000] Artículo: “Structure and Style in Use Case for User Interface Design”. Larry L. Constantine, Lucy A.D. Lockwood. © 2000. <http://foruse.com/articles/structurestyle2.htm>

[Jacobson et al.1998] Object-Oriented Software Engineering: A Use Case Driven Approach. Addison-Wesley. © 1992, Ed.1998. Ivar Jacobson, Magnus Christerson, Patrik Jonsson, Gunnar Overgaard.

[Jacobson et al. 2000] El proceso unificado de desarrollo de software. Pearson. 2000. Ivar Jacobson, Grady Booch, James Rumbaugh

[Kotonya&Sommerville 2002] Requirements engineering: processes and techniques. John Wiley. © 1998, Ed. 2002. Gerald Kotonya, Ian Sommerville.

Bibliografía de referencia:

[Larman 1999] UML y Patrones. Introducción al análisis y diseño orientado a objetos. Prentice-Hall & Pearson. 1999. Craig Larman

[Leffingwell&Widrig 2003] Managing Software Requirements. Second Edition. A Use Case Approach. Ed. 2003. Dean Leffingwel y Don Widrig.

[Muller 1997] Modelado de Objetos con UML. Gestión 2000. 1997. Pierre-Alain Muller.

[Pinto, 2009] Pinto, María. Proyecto Alfin-EEES. Año 2009. www.mariapinto.es/alfineees/buscar.htm

[Pressman 2002] Ingeniería del software: un enfoque práctico. McGraw-Hill. 2002. Roger S. Pressman

[SWEBOK 2004] "SWEBOK" - Guide to the Software Engineering Body of Knowledge. 2004 version. - IEEE Computer Society -. www.swebok.org

Bibliografía de Internet:

1. "Use cases, ten years later". Alistair Cockburn. 2002. <http://alistair.cockburn.us/index.php/>

2. Métodos Ágiles. <http://www.foruse.com>

3. "Use cases - yesterday, today, tomorrow". Ivar Jacobson. 2003.

http://www.therationaledge.com/content/mar_03/f_useCases_ij.jsp

4. “UML – Resource Page”. <http://www.uml.org/>

5. “Casos de Uso. Un método práctico para explorar requerimientos”. Santiago Ceria. 2002. http://www-2.dc.uba.ar/materias/isoft1/2002_2/apuntes/CasosDeUso.pdf

6. “Relaciones entre Casos de Uso en el Unified Modeling Language”. Roxana S. Giandin, Claudia F. Pons. 2000. http://www.unab.edu.co/editorialunab/revistas/rcc/pdfs/r11_art4_c.pdf

7. “Estudio de la Necesidad y Propuestas para Formalizar Esquemas de Trazabilidad”. Angelina Espinoza Limón. 2004. <http://lucio.ls.fi.upm.es/doctorado/Trabajos20042005/Espinoza.pdf>
8. “Indicadores organizacionales para comparación de herramientas case en venezuela”. Teresita Rojas, Luis E. Mendoza, María A. Pérez. 2000.
[http://www.lisi.usb.ve/prof/lmendoza/Documentos/PS-6117%20\(Laboratorio\)/Indicadores%20Organizacionales%20para%20Comparaci%C3%B3n%20de%20Herramientas%20CASE%20en%20Venezuela.pdf](http://www.lisi.usb.ve/prof/lmendoza/Documentos/PS-6117%20(Laboratorio)/Indicadores%20Organizacionales%20para%20Comparaci%C3%B3n%20de%20Herramientas%20CASE%20en%20Venezuela.pdf)

Prometeo: Una Herramienta Para El Aprovechamiento De Metadatos De Base De Datos Relacionales

Marciszack, Marcelo
Maldonado, Calixto
Martinez Spessot, Cesar
Muñoz, Roberto
Navarro, Adrián
Peretti, Juan Pablo
Roggero, Luis

**Laboratorio de Investigación de Software/Dep. Ing. en Sistemas de Información/
Facultad Regional Córdoba
Universidad Tecnológica Nacional**

CONTEXTO

El presente trabajo es un informe de avance y describe estado del proyecto Prometeo según los logros alcanzados en el año 2008 por los integrantes del proyecto, iniciado en el año 2004 y que ha sido distinguido con la acreditación de las Secretarías de Ciencia y Técnica de la UTN en enero del 2009. Es un proyecto relacionado con TecnoDB, que es un motor de Base de datos relacional para ser ejecutado en dispositivos móviles, proyecto acreditado en sus dos fases, la última con finalización en el presente año. El objetivo del proyecto Prometeo es definir una heurística de explotación de meta datos y un software capaz de implementar esta metodología. Con este nombre fue presentado y aprobado en el programa de Grupos de Reciente Formación de la Agencia Córdoba Ciencia en el año 2007.

RESUMEN

Se presenta una síntesis del proyecto Prometeo y los avances logrados durante el año 2008. Se enuncian los antecedentes, características de la heurística usada y los objetivos del software, logros recientes y avances previstos en este año.

Palabras clave: *SQL, Generador de Consultas, QueryByExample, SQL autogenerado*

1. INTRODUCCIÓN

En presentaciones en WICC2005 [GAS05] y 35JAIIO [GAS06] se describía el proyecto Prometeo como “un sistema de software que se ejecute en múltiples plataformas de sistemas operativos, en principio, Windows 2000 y Red Hat Linux 7.1 y que sea capaz de generar sentencias SQL basados en los metadatos de los diccionarios de datos de Oracle y PostgreSQL: Consultas simples con SELECT, Consultas multi tablas de las tablas relacionadas con Constraints Referenciales, DML para insertar, borrar y actualizar columnas, DDL para creación de objetos

como vistas y tablas accesorias, modificación y borrado de objetos como Vistas, OQL o lenguaje de consulta de Objetos cuando la base de datos estudiada contenga esa opción”

Hubo cambios de integrantes por el egreso de los desarrolladores originales y por el ingreso de nuevos integrantes, provenientes del alumnado de la carrera de Ingeniería en Sistemas de Información.

El proyecto, que responde al nombre extenso de “Desarrollo de un método y una herramienta para el aprovechamiento de Metadatos de Base de Datos Relacionales” fue acreditado este año por la SeCyt del UTN, y ha logrado sólidos avances. El nombre extenso es habitualmente reemplazado por el nombre abreviado de Prometeo, que según Wikipedia [WIK09] “es el Titán amigo de los mortales, honrado principalmente por robar el fuego de los dioses en el tallo de una cañaheja, darlo a los humanos para su uso y ser castigado por este motivo.” La metáfora del nombre es por que este software ayuda a conocer y aprovechar la información almacenada en los metadatos y con la ayuda de unos criterios o Heurística permite construir consultas, ejecutarlas y guardarlas en tablas propias.

La propuesta inicial fue presentada en un paper que fue aceptado en las 34 JAIIO desarrolladas en Rosario en el mes de Septiembre del 2005, en el simposio ASIS [GAS05]

El hecho que permite obtener provecho de los metadatos es la capacidad de la sentencia SELECT de poder combinar el resultado de seleccionar el contenido de las columnas de las vistas del Diccionario con literales para poder escribir sentencias SELECT validas. Ejemplo:

Dada la vista del diccionario de datos Oracle [CON05] USER_TABLES, que contiene todas las

tablas del usuario conectado, se podría generar una sentencia select de todas las columnas de las tablas existentes con la siguiente sentencia:

```
SELECT 'select * from '||table_name||';'
FROM USER_TABLES;
```

La ejecución de esta consulta generará en la salida estándar de la herramienta usada:

```
select * from tabla1;
select * from tabla2;
```

Una sentencia SELECT por cada registro consultado de la vista del diccionario USER_TABLES.

Agregándole una línea de texto podríamos crear una sentencia SQL para crear una vista que guarde esta consulta como un objeto de la base. Ejemplo:

```
SELECT 'create or replace view '||table_name ||'_v
as select * from '||table_name||';'
FROM USER_TABLES;
```

Lograríamos esta sentencia valida SQL, basada en todas las columnas de la tabla 'PEPE' perteneciente al usuario actual:

```
create or replace view as PEPE_v as select * from
PEPE ;
```

Este funcionamiento del lenguaje SQL es la base del generador de sentencias SQL, basadas en los metadatos.

Sobre la Heurística tiene como objetivo el aprovechamiento de los metadatos residentes en motores de bases de datos comerciales como Postgresql y Oracle y en desarrollo, como TecnoDB. Este método o heurística guía a la generación de las sentencias útiles dentro de todas las combinaciones posibles y contiene 21 puntos o criterios a seguir y los resultados a obtener.

Logrando esto se espera poder generar sentencias SQL con algun grado de optimización [Cis02] que permitan ahorrar trabajo y errores comunes a los desarrolladores de aplicaciones en estas bases de datos. Estamos buscando mejorar y completar el método que viene desarrollandose desde su inicio y aumentar las funcionalidades de su interfaz, para que pueda definir qué consultas se deben generar para que asista al desarrollador a construir reportes, generar conjuntos de filas de pruebas de integridad y desarrollar una herramienta que la aplique par lograr que con el solo hecho de ejecutarla, obtener un subconjunto de sentencias útiles sin necesidad de conocer profundamente el modelo de datos objetivo. Una de las formas de mejora en el proyecto es diseñar el método para generar Java Beans, es decir,

extender la accion a construir codigo basado en JAVA, generando codigo reusable [WIK09b] [SUN09]

Prometeo tiene como objetivo también aportar al proyecto TecnoDB sirviendo como aporte al diseño del diccionario de datos, aún no finalizado en el desarrollo del motor de Bases de datos relacional.

Las sentencias que genera Prometeo son inicialmente SELECT, y CREATE VIEW. Se completará con las sentencias INSERT y CREATE TABLE, pudiendo aumentar esta capacidad a otras sentencias mas. Las sentencias INSERT van ser generadas para insertar conjuntos de datos de prueba de las tablas elegidas, respetando las restricciones de integridad referencial detectadas en el modelo[EMS08a] [EMS08b] [Dat97].

En imágenes una explicación de cómo este software, genera, muestra, ejecuta, muestra los resultados y grababa en sus propias tablas, las sentencias SQL obtenidas con las indicaciones del usuario. En la figura 1 se muestra un esquema del funcionamiento.



Fig. 1 Pasos en la generación de Sentencias

En la figura 2 el modelo de datos inicial, modificado durante 2008 al requerir almacenar mas información.

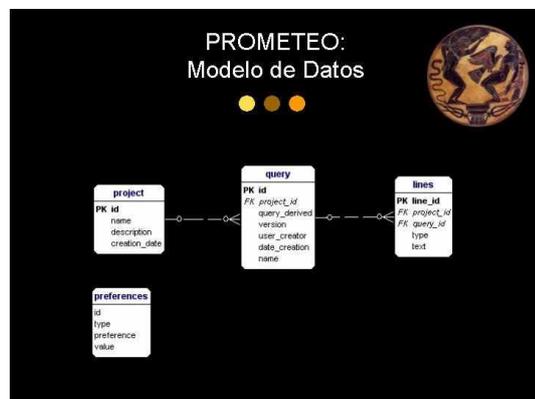


Fig. 2

La interfaz se conecta al diccionario de datos de la base de datos objetivo y guarda en la base de datos repositorio (que puede ser la misma) la información acerca de los objetos analizados en tablas propias. Permite ejecutar las consultas generadas o almacenadas y mostrar el resultado en la pantalla, para luego guardarla en el repositorio. Los lenguajes a utilizar en la construcción son PHP y JAVA para la interfaz y procedimientos y SQL para la comunicación con las bases de datos objetivo. Se realizó la comunicación con PostgreSQL y se está trabajando para realizar la conexión con MySQL.

También se avanzará en que la interfaz pueda recopilar información de documentación con los contenidos de Comentarios sobre tablas y columnas posibles de ser guardados en las vistas del diccionario de las bases de datos objetivo, por ejemplo, en Oracle [CON05] la vista ALL_TAB_COMMENTS y la ALL_COL_COMMENTS permitirán mejorar la documentación de los desarrollos que utilicen a Prometeo como herramienta.

Se están definiendo los casos de usos correspondientes a cada punto enunciado para poder diseñar la aplicación que lo resuelva y que sirva como base al proceso de testing [COC00] [LAR00] [JAC00].

2. LINEAS DE INVESTIGACION y DESARROLLO

Desarrollo de avances en la Heurística [SPE05] para el uso de los Metadatos en los diccionarios de las Bases de Datos Oracle, PostgreSQL, MySQL. [TOA08]

Diseño del modelo de datos de Metadatos para TecnoDB.

Diseño de Vistas para las distintas versiones de RDBMS usadas para lograr independencia en las distintas instalaciones.

Creación de Java Beans con la información de los Metadatos para construir automáticamente métodos de acceso, actualización a los datos de las tablas analizadas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Hasta el 2007 se programó un sistema web, desarrollado en PHP 5 y que obtuviera información de los metadatos de PostgreSQL [GAS05].

Durante el 2008, se trabajó sobre MySQL lográndose lo mismo que hasta el 2007 y además se pudo construir sentencias SELECT multitaslas, es decir relacionaba tablas con constraints de clave foránea para escribir los Joins sobre las mismas tablas. Esto estaba previsto en la Heurística original pero no se había concretado aún [SPE07] [ALT08].

Durante el trabajo con MySQL se observó que el modelo de Datos objetivo de los metadatos de MySQL son manejados por el motor a través de la base de datos INFORMATION_SCHEMA, la misma está caracterizada por no contener tablas, sino que la información es manejada a través de vistas. En este Diagrama de Entidad-Relación se pueden observar las vistas más importantes (sobre todo las utilizadas por Prometeo) en forma de tablas y relaciones entre las mismas. [MYS08]

Se modificó esta pantalla con la idea de brindar la posibilidad al usuario de seleccionar el tipo de consulta a realizar. Se vio que sería importante lograr la pantalla de inicio debería mostrar una leyenda de bienvenida al usuario y darle la posibilidad de seleccionar una tarea entre varias, como por ejemplo, generar consultas, seleccionar consultas guardadas de acuerdo a un proyecto particular y ejecutar dicha consulta y/o ver todas las consultas guardadas junto a los proyectos a las cuales pertenecen. (Fig. 3)

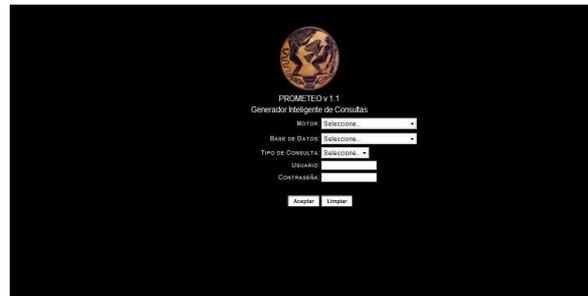


Fig 3 Pantalla de Inicio de la V1.1

La siguiente pantalla (Fig. 4) ofrece dos criterios para lograr los emparejamientos de columnas para las consultas multitaslas, pero solo soporta uno solo de los dos criterios (Tablas referenciadas por esta tabla). Es decir, la funcionalidad está implementada para ese criterio solamente, por lo que debería desarrollarse la funcionalidad para el criterio restante.

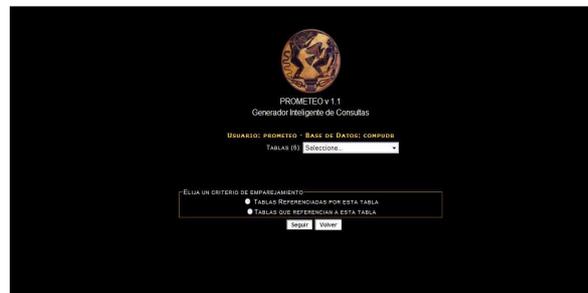


Fig. 4 Pantalla para generar consultas multitaslas.

En la siguiente imagen (Fig 3) se muestra página con la consulta generada a partir de los datos seleccionados anteriormente. La funcionalidad de

los botones “Guardar Consulta” y “Ejecutar Consulta” esta en desarrollo.



Fig 3. Resultado de la Generación de consultas multitabla.

Son objetivos del trabajo durante este año:

- Extender la aplicación a otros motores comerciales como SQL Server y DB2 [DB208].
- Avanzar con nuevos productos del análisis de los metadatos como generar automáticamente Java Beans que contengan los métodos get y set de una tabla. [ISQ,08]
- Migrar la programación de la aplicación de PHP a Java. [QUE08]
- Generar en las tablas estudiadas set de datos de prueba de acuerdo a los criterios que se podran registrar como preferencias en la herramienta.

También se ha previsto para la próxima etapa del proyecto la inclusión de elementos de diseño, tales como estilos web para lograr una interfaz amigable con el usuario y desarrollar un asistente o Wizard para tener un punto de inicio y donde el usuario pueda seleccionar la/s tarea/s que interesan realizar y que se vayan desarrollando en la secuencia correcta.

Cabe señalar que del desarrollo anterior, solo se hizo una adaptación de lo existente para ser utilizado con MySQL, no se desarrolló de nuevo la funcionalidad para consultas simples, y que deben completarse ciertos aspectos que son heredados del desarrollo anterior.

4. FORMACION DE RECURSOS HUMANOS

Los participantes son tres alumnos de la UTN-FRC que están cursando los últimos años de la carrera de Ingeniería y su trabajo ha sido considerado dentro del ámbito de la materia Práctica Profesional Supervisada o PPS, durante el presente año.

Durante el año 2008 estuvieron trabajando como Becarios de la Secretaria de Alumnos y Extensión al desarrollar la tarea descripta.

Se hará transferencia de los aprendizajes adquiridos en algoritmos de mejora de performance, almacenamiento y procesamiento de consulta de datos a las cátedras de la Universidad Tecnológica Nacional, especialmente a Sistemas de Soporte de Decisión (5to año de la carrera de Ingeniería en

Sistemas de Información) donde se tratan temas avanzados de base de datos.

El trabajo esta relacionado con el grupo de desarrollo de TecnoDB de la UTN, con el objetivo de que Prometeo sea una herramienta de TecnoDB para mejorar sus prestaciones al contar con una interfaz amigable para realizar consultas SQL.

5. BIBLIOGRAFIA

[ALT08] www.Altova.com/DatabaseSpy visitado Mayo2008

[BUR99] Burgos M. (1999) “Tutorial de SQL”, <http://members.tripod.com/~MoisesRBB/sql.html>
Fecha de consulta: mayo 2005

[CAV00] Cavadini S. (2000) “Compiladores e Intérpretes”, <http://www.ucse.edu.ar/fina/compiladores/>
Fecha de consulta: agosto 2005

[CIS02] Cisterna M.(2002) “Métodos de Optimización de Consultas para el Lenguaje SQL”, <http://macine.epublish.cl/tesis/index.html>
Fecha de consulta: agosto 2005

[COC00] Alistair Cockburn “Writing Effective Use Cases, Addison-Wesley

[CON05] Documentación oficial Oracle Database, Concepts 10g Release 2- June 2005

[DAT97] Date C. J. - Darwen H. (1997) “A guide to the SQL Standard”, Addison Wesley, Reading – MA, Fourth Edition

[DB208] DB2 edicion express C <http://www-01.ibm.com/software/data/db2imstools/db2tools/db2wqt/> visitado mayo 2008

[DEE02] Deepak P. - Sandeep S. (24 de agosto 2002) "Valgrind HOWTO", <http://www.tldp.org/HOWTO/Valgrind-HOWTO/>
Fecha de consulta: desde enero a agosto de 2005

[EMS08a]<http://www.sqlmanager.net/products/mssql/datagenerator>

[EMS08b] <http://www.sqlmanager.net/?gclid=CNeW07SouZUCFQyenAod9gQCQA> visitado junio 2008

[GAS05] Gastañaga, Iris et al TecnoDB: Desarrollo de una metodología de aprovechamiento de Metadatos de los diccionarios de Datos de Bases de Datos Relacionales para lograr un generador de sentencias SQL JAIIO34 ASIS 2005 Argentine

Symposium on Information Systems Rosario,
Argentina - September 29-30, 2005

[GAS06] Gastañaga, Iris et al TecnoDB una Base de
Datos Relacional y Prometeo un metodo de
aprovechamiento de Metadatos y Generador de
Consultas WICC 2006

[ISQ,08] <http://java-source.net/open-source/sql-clients/isql-viewer> visitado Julio de 2008

[JAC00] Jacobson I. - Booch G. - Rumbaugh J.
(2000) “El Proceso unificado de Desarrollo de
Software”, Addison Wesley, Madrid – España

[Lar00] Craig Larman “Applying UML and
Patterns”, Addison- Wesley

[MYS,08]
[http://www.mysql.com/products/tools/query-
browser/](http://www.mysql.com/products/tools/query-browser/) visitado Mayo 2008

[ORA08]
[http://www.oracle.com/global/kr/download/seminar/
2008/otn/s01.pdf](http://www.oracle.com/global/kr/download/seminar/2008/otn/s01.pdf) visitado Junio 2008

[QUE08] <http://www.quest.com/> visitado en Julio
2008

[QTO08] <http://www.querytool.com/> visitado en
julio 2008

[SPE07] Martinez Spessot et al, PROMETEO Un
método de explotación de Metadatos. JIDIS 07

[TOA08] <http://www.toadsoft.com/>

[UNI05a][http://www.universia.com.ar/contenidos/un
iversidades/index.html](http://www.universia.com.ar/contenidos/universidades/index.html)
Fecha de consulta: agosto 2005

[UNI05b]“Universidades.org”,
<http://www.universidades.org/> - Fecha de consulta:
agosto 2005

[SUN09]
[http://java.sun.com/javase/technologies/desktop/javabe
ans/index.jsp](http://java.sun.com/javase/technologies/desktop/javabeans/index.jsp)

[WIK09a]<http://es.wikipedia.org/wiki/Prometeo>
visitado Abril 2009

[WIK09b] Java Beans
<http://es.wikipedia.org/wiki/JavaBean> - Fecha de
consulta: desde diciembre 2008

TITULO

Gestión de los Sistemas de Información (GSI)

Sebastián José García*
sgarcia@iinfo.unsj.edu.ar

Mirta Elizabeth Navarro*
mirtaenavarro@yahoo.com.ar

Maria Claudia Gómez*
cgomez@unsj-cuim.edu.ar
cacugomez@yahoo.com.ar

CONTEXTO

La línea de investigación incluye a los proyectos “Adaptación de las Técnicas de Resolución de Conflictos a la Gestión de los Sistemas de Información” e “Integración de Tecnologías Informáticas como Soporte para los Sistemas de Información” acreditados con códigos 21-E-819 y 21-E-821 en la Convocatoria 2007 de la Universidad Nacional de San Juan.

La línea se desarrolla como continuación de los proyectos “Apoyo a Organizaciones Locales en la Caracterización de sus Sistemas de Información” de la Convocatoria 2003, “Análisis e Integración de los Sistemas de Información en función del avance de las Tecnologías de Información” y “Contribuciones al Desarrollo de Sistemas de Información e implementación de TICs en organizaciones locales” de la Convocatoria 2005.

RESUMEN

En esta línea de investigación tenemos como objetivo principal profundizar en los componentes de la función de gestión de los Sistemas de Información y sus relaciones con aspectos relevantes de los Sistemas de Información, tales como: su evolución, la integración tanto internamente como interorganizacional, los conflictos personales y de grupo que se generan y el desarrollo organizacional.

Palabras Clave: Gestión, Evolución e Integración de los Sistemas de Información.

1. INTRODUCCION

Los Sistemas de Información (SI) han evolucionado y tomado una importancia estratégica en las organizaciones a tal punto que se ha generado gran necesidad de hacer una gestión efectiva y eficiente de ellos. La gestión exitosa de los SI es crucial debido a varios factores, entre ellos:

- La Tecnología de la Información (TI) es ahora un activo estratégico [3] que se necesita para realizar una gestión estratégica y habilitar el cambio de los procesos organizacionales.
- Las aplicaciones de TI han incrementado su complejidad incluyendo ambientes interorganizacionales.
- Las capacidades y complejidades en el uso de la TI están creciendo a tasas muy altas.
- La Globalización, por la cual las organizaciones buscan ofrecer y procurar sus bienes y servicios en cualquier lugar del mundo.
- La E-habilitación, que significa que la Internet permite hacer negocios electrónicamente donde clicks y ladrillos existen uno al lado del otro.
- La gestión y compartimiento del conocimiento a través de la Inteligencia de negocio. El gran tema es cómo manejarse con todo el conocimiento del mundo actual.

Es necesario entonces, disponer de herramientas de los aspectos estratégicos, tácticos y operativos para la gestión del uso de la TI. El trabajo y la experiencia obtenida en organizaciones

de primera línea debe ser la que sirva de base para estas definiciones.

Si bien diferentes autores muestran algunos aspectos de la GSI, se hace necesario una recopilación de los conceptos y las visiones presentadas por dichos autores con aplicabilidad al medio nacional y local.

1.1 Antecedentes

Desde la creación de la Carrera de Licenciatura en Informática con orientación a los SI en la UNSJ en 1985, este grupo de investigadores del área de SI se ha mantenido apoyando y transformando la carrera en Sistemas de Información. Primero a través de estudios de posgrado y luego a través de proyectos de investigación. Ahora, con la transformación específica en el 2005 en una carrera de sistemas de información la problemática se hace más necesaria e importante.

El libro [14] en todas sus versiones (la primera en 1986 y la última en el 2009) ha sido y sigue siendo la guía para ésta línea de investigación. Se ha generado material escrito en los proyectos y tesis, a modo de ejemplo cito los libros [5], [6] y [7].

1.2 Objetivos

Seguir trabajando en equipo y apuntando el trabajo docente, de investigación y de extensión relacionados con la Carrera de Licenciatura en Sistemas de Información.

1.3 Relevancia

La TI está teniendo la clase de impacto revolucionario y de reestructuración que ha sido prometido y esperado durante años. Los avances rápidos en la velocidad y capacidad de los dispositivos computacionales, unido a la penetración de la Internet, almacenamiento digital, dispositivos portátiles e inalámbricos y

el contenido de la multimedia están causando los principales cambios en la forma en que vivimos y trabajamos.

Aunque la TI afecta a casi todos los aspectos del comportamiento humano, esta línea enfatiza su uso en la gestión y operación de las organizaciones, incluyendo empresas comerciales, organizaciones de gobierno y organizaciones caritativas y sociales. Toda vez que las personas trabajan juntas para obtener objetivos, la TI está cambiando la forma en que ellas trabajan.

La gestión y operación de la TI para esos propósitos ha sido “un campo de práctica” cerca de 50 años. Primero conocida como procesamiento de datos del negocio y luego como sistemas de información de gestión (MIS), ahora el campo es conocido como tecnologías de información (TI). Ahora nos enfocamos en la gestión de los SI. Los SI combinan las tecnologías, personas, procesos y mecanismos organizacionales para promover el uso de la TI y mejorar el rendimiento organizacional.

2. LINEA DE INVESTIGACION y DESARROLLO

Este trabajo apunta a caracterizar la función de GSI mediante un modelo y obtener experiencias relacionadas con la aplicación de éste modelo en la GSI en las empresas del medio local. El modelo fue extraído de [14] y se introdujo en [5] y se aplicó a la estructuración de la función de GSI de la UNSJ. Ese modelo es interesante ya que muestra cómo las diferentes actividades relacionadas con SI pueden ser incluidas en dicha función.

En [7] se muestra una metodología de evaluación de la GSI en una escala de evolución de los SI, que se seguirá mejorando y contrastando en el medio ambiente local.

Una nueva dimensión en la problemática se ha introducido con el proyecto “Adaptación de las Técnicas de Resolución de Conflictos a la Gestión de los Sistemas de Información”. La nueva dimensión es el tratamiento de los conflictos en la GSI. Es interesante estudiar y proponer herramientas que permitan que las personas o grupos en conflicto se comuniquen y faciliten la resolución de conflictos.

3. RESULTADOS OBTENIDOS / ESPERADOS

Se espera que el trabajo en ésta línea de investigación ofrezca resultados al nivel teórico y de aplicación. En el nivel teórico se espera caracterizar mediante un modelo la GSI y detallar algunas dimensiones como el caso de los conflictos. En el nivel de aplicación se espera fortalecer las relaciones obtenidas con empresas del medio local en la aplicación de la GSI.

4. FORMACION DE RECURSOS HUMANOS

Se espera comenzar relaciones con grupos de otras Universidades del país y del exterior en la formación de recursos humanos.

5. BIBLIOGRAFIA

- 1- APPLGATE, Lynda M., MC FARLAN, F. Warren y MCKENNY, James L., *Corporate Information Systems Management: Text and Cases*, McGraw-Hill Companies Inc, United States of América, 1996.
- 2- BASCHAB, John y Piot, Jon, *The Executive's Guide to Information Technology*, John Wiley & Sons, Inc, EEUU, 2003. (Digital)

- 3- BOAR, Bernard H., *Strategic Thinking for Information Technology*, John Wiley & Sons Inc, United States of América, 1997.
- 4- CARR, Nicholas G., *IT Doesn't Matter*, HBR, May 2003.
- 5- GARCIA, Sebastián José, *Estructuración de la Función de Gestión de los Sistemas de Información*, Multicopy S.R.L., Julio 2002.
- 6- GARCIA, Sebastián José, *Gestión de los Sistemas de Información*, Valentina S.R.L., Febrero 2006.
- 7- GARCIA, Sebastián José, *Evolución de los Sistemas de Información*, Valentina S.R.L., Febrero de 2006.
- 8- HAX, Arnoldo y MAJLUF, Nicolás, *The Strategy Concept and Process: A Pragmatic Approach*, Prentice Hall, New Jersey, EEUU, 1996.
- 9- HARTMAN, Amir and SIFONIS, John with KADOR, John, *Net Strategies for Success in the Economy*, McGraw-Hill, New York, EEUU, 2000.
- 10- KAPLAN, Robert S. and NORTON, David P., *The Balanced Scorecard*, Harvard Business School Press, Boston, Massachusetts, EEUU, 1996.
- 11- KAPLAN, Robert S. and NORTON, David P., *The Strategy – Focused Organization, How Balanced Scorecard Companies Thrive in the New Business Environment* Harvard Business School Press, Boston, Massachusetts, EEUU, 2001.
- 12- LAUDON, C. Kenneth and LAUDON, Jane Price, *Management Information Systems: New*

- Aproaches and Tecnology*, Sexta edición, Prentice Hall College Div., New Jersey, EE.UU, 2000.
- 13- LAUDON, C. Kenneth and LAUDON, Jane Price, *Management Information Systems: Managing The Digital Firm*, Octava edición, Prentice Hall College Div., New Jersey, EE.UU, 2004.
- 14- MCNURLIN, Barbara y SPRAGUE, Ralph, *Information Systems Management in Practice*, Prentice Hall, New Jersey, EEUU, 1998, 2002, 2006 y 2009.
- 15- MENDELSON, Haim y Zieger Johannes, *Los 4 principios de la Empresa Inteligente*, Deusto, España, 2004.
- 16- MERCHANT, Kenneth A., *Modern Management Control Systems, Text and Cases*, Prentice Hall, New Jersey, EEUU, 1998.
- 17- NEGROPONTE, Nicholas, *Being Digital*, Alfred Knopf, New York, 1995.
- 18- O'BRIEN, James A. Y Marakas George M., *Management Information Systems*, McGraw-Hill, 2006
- 19- OZ, Efy, *Administración de sistemas de información*, Segunda edición, Thomson Learning, México, 2001.
- 20- PETERS, Tom, *Thriving on Chaos: Handbook for a Management Revolution*, Alfred A. Knopf Inc, New York, 1987.
- 21- PORTER, Michael E., *What is Strategy?*. HBR, November - December 1996.
- 22- PRIMOZIC, Kenneth y PRIMOZIC, Edward, *Strategic Choices: Supremacy, Survival, or Sayonara*, McGraw-Hill, New York, 1991.
- 23- SÁNCHEZ ORTIZ, Aurora, *Evaluación de desempeño de Sistemas de Información desde una Perspectiva Estratégica*, Tesis para Optar al Grado de Magister en Ciencias de la Ingeniería Mención Ingeniería Industrial, Departamento de Industrias, Universidad de Chile, 1997.
- 24- SIMONS, Robert, *Performance Measurement & Control Systems for Implementing Strategy*, Prentice Hall, New Jersey, EEUU, 2000.
- 25- ZACHMAN & ZOWA, Extending and formalizing the framework for information system architecture, *IBM System Journal*, Vol. 31, N° 3, 1992.

Contexto

El siguiente proyecto es resultado de la integración de dos subsistemas creados por el instituto INTIA división INCA. Se investigan dichos subsistemas para lograr una integración en un macro sistema que permita el control de flotas de vehículos.

Resumen

Se presenta aquí el proyecto de desarrollo de un sistema tal que permita controlar una flota de vehículos. El mismo hace referencia a la telemetría y georreferenciación brindando información de una unidad móvil en tiempo real, principalmente su estado y ubicación. Se utilizan tecnologías tales como GPS y la red de telefonía móvil GSM.

Palabras claves

GPS, GSM, telemetría, georreferenciación, tiempo real, rastreo satelital.

1. Introducción

Dada la alta competitividad actual del mercado, toda empresa se enfrenta día a día con la necesidad de reducir sus costos y aumentar la calidad del servicio prestado. Los recursos utilizados deben ser administrados de la manera más eficiente para reducir gastos y aumentar las ganancias. La optimización de los recorridos a partir del conocimiento de la ubicación de unidades representa un pequeño aporte. Adicionalmente, el mantenimiento preventivo de unidades implica una disminución de fallos mecánicos. La prevención de los mismos permite mantener una unidad en actividad la mayor parte de su vida útil.

1.1 Motivación

Las empresas que cuentan con unidades móviles como herramienta de trabajo mantienen cierta infraestructura. La misma está compuesta por una planta de

trabajadores, un espacio físico, una flota de unidades, entre otros componentes que hacen a la operatividad de la empresa. Para llevar adelante la actividad consumen una amplia variedad de recursos. La optimización de estos recursos es de gran importancia ya que representan un elevado gasto económico. De manera adicional, existen muchos otros factores que impactan directamente en la economía de una empresa de este tipo. Entre ellos podemos mencionar la inseguridad relacionada al robo o hurto de unidades y/o mercancías, siniestros por falla mecánica o humana, mala administración de recursos, etc. Es muy importante tener información del desempeño de las unidades tanto on-line como off-line. En el primer caso para poder llevar un control de la flota en tiempo real. En el segundo para procesar y calcular estadísticas relacionadas a la utilización de recursos, para optimizar recorridos, para realizar mantenimiento preventivo de unidades, y más.

1.2 Objetivo

Este trabajo tiene como objetivo general el desarrollo de un sistema de control de flotas. Un sistema compuesto tanto por hardware como por software que asista en el control de una flota de unidades móviles. El propósito principal es el de brindar una forma de optimización de recursos por medio de la informatización de sus actividades. Dicha informatización debe permitir al personal idóneo de la empresa tomar decisiones. Como objetivo de menor prioridad se propone la reutilización de componentes de software y hardware existentes para reducir los impactos tanto económicos como en los tiempos de desarrollo.

Por consiguiente, el objetivo del presente proyecto se enfoca esencialmente en comunicar o integrar dos subsistemas existentes (subsistema de coordenadas

geográficas y sensores y subsistema de visualización georreferenciada) para lograr una visualización on line de la ubicación y estado de una unidad móvil.

El mismo se desarrolla dentro del grupo de investigación INTIA división INCA de la Universidad del Centro de la Provincia de Buenos Aires.

2. Líneas de investigación y desarrollo

Los ejes principales de investigación del presente proyecto son la georreferenciación, como proceso de localizar un objeto en un sistema geográfico de coordenadas [1, 2, 10], y la telemetría que permite la medición remota de magnitudes físicas en lugares distantes o de difícil acceso y el posterior envío de la información hacia el centro de control correspondiente [3, 4].

3. Resultados obtenidos/esperados

Al concluir con la etapa de desarrollo se logran los objetivos planteados para el presente proyecto. Se obtiene un sistema de control de flotas de vehículos de tiempo real re utilizando subsistemas y módulos que estaban en funcionamiento antes de comenzar el mismo.

3.1 Arquitectura general

La arquitectura del sistema se basa en un modelo cliente-servidor, conectados a través de la red de telefonía GSM [5]. Por medio del envío de mensajes de texto mediante ésta red GSM los clientes le comunican al visualizador las coordenadas geográficas emitidas por el GPS [6] y los valores recolectados de los sensores.

3.1.1 Cliente

En el contexto del sistema un cliente es un vehículo con un GPS emitiendo a intervalos fijos de tiempo información acerca de su posición global y estado de sensores. El cliente está representado por el subsistema de coordenadas geográficas y sensores.

3.1.2 Servidor

El servidor es el subsistema de visualización georreferenciada ejecutándose sobre una

computadora. El mismo recibe la información de posición y estado de una unidad y actualiza las propiedades del objeto visual que representa al cliente que envió el mensaje de texto. Esta información entrante de la unidad es procesada por el subsistema encargado de tratar los mensajes de texto de la red GSM.

3.2 Subsistema de coordenadas geográficas y sensores

Se cuenta con un dispositivo hardware que integra la tecnología GPS con GSM. El mismo obtiene una ubicación geográfica válida y la envía a través del servicio de mensajería proveído por la red de comunicaciones móviles (GSM). El dispositivo cuenta con un microcontrolador [7, 8, 11] que es el encargado de recolectar datos y transmitirlos.

El componente hardware aquí mencionado posee un software de control el cual está diseñado en torno a un scheduler que atiende a cada una de las tareas dependiendo de su prioridad. Cabe destacar que existen diferentes prioridades en las tareas involucradas. Se asegura mediante mensajes que todas las tareas son ejecutadas en algún momento no importa cual sea su prioridad. Por ejemplo, en el caso ideal las tareas como el cálculo de posición y rumbo son realizadas en un tiempo finito.

3.2.1 Colaboración en la funcionalidad del sistema global

El subsistema de coordenadas geográficas y sensores es el encargado de generar las coordenadas correspondientes y obtener los valores de los sensores apostados en la unidad. De esta manera cumple con la funcionalidad inicial de todo el sistema global. Es el primer subsistema que actúa y genera información necesaria para la entrada de los restantes subsistemas en el hilo de ejecución. Posee su funcionalidad bien definida y funciona de manera independiente a los demás. El formato de los mensajes de texto producidos por el mismo es:

[ESTADO] en [LAT] [LONG] ads:
[X1][X2][X3][X4].

Donde:

[ESTADO] puede ser NORMAL o PANICO. El mismo se activa mediante un pulsador cuando sucede una situación de emergencia y su estado pasa a ser PANICO de otra forma el estado es siempre NORMAL.

[LAT] es el valor de la latitud perteneciente a la coordenada geográfica obtenida en un determinado instante. El formato es #####.##### de donde se toman los dos primeros dígitos de adelante para calcular el valor en grados, los dos siguientes antes del punto para los minutos y los restantes cuatro luego del punto para los segundos.

[LONG] idem al anterior pero es el valor de la longitud.

[X1][X2][X3][X4] son los valores obtenidos de los sensores en un instante dado. Cada uno representa la medición de un factor diferente como puede ser la temperatura, presión de aceite, revoluciones de motor, etc. La función que cumple cada sensor se puede cambiar a conveniencia del usuario.

Como ejemplo se muestra el mensaje de texto producido por una unidad situada en el campus universitario de Tandil, Buenos Aires, Argentina

NORMAL en 3719.2818 5904.9411 ads: 89 30 256 0

3.3 Subsistema de tratado de mensajes de texto

Luego de investigar diferentes tecnologías se concluye que el subsistema encargado de tratar los mensajes de texto se desarrolla en Java usando la biblioteca SMSLib [9]. SMSLib es una biblioteca Java para el envío y recepción de mensajes SMS vía módem GSM o teléfono móvil. SMSLib está distribuido bajo los términos de la licencia Apache v2. En la actualidad se encuentra en su tercera versión. Se implementa esta biblioteca dado que brinda una gran funcionalidad y abstrae al programador de las características físicas de bajo nivel de los diferentes módems GSM.

3.3.1 Colaboración en la funcionalidad del sistema global

El subsistema encargado de recibir los datos (mensajes de texto) provenientes de la red de

telefonía celular GSM es un sistema basado en eventos que espera la notificación de la entrada de un nuevo mensaje. El subsistema accede al módem GSM a través de un puerto virtual que está ligado directamente mediante tecnología bluetooth al dispositivo móvil. El subsistema espera la notificación de la entrada de un nuevo mensaje y ante este evento captura el contenido e información de procedencia. Analiza si el origen del mensaje es correcto y luego publica en un socket el contenido del mismo. Hasta este punto colabora el subsistema de tratado de mensajes de texto.

3.4 Subsistema de visualización georreferenciada

El subsistema provee la visualización en capas de información geográfica. Cada una de éstas capas esta conformada por imágenes vectorizadas que contienen por ejemplo: espejos de agua, rutas y caminos, actividades humanas, puntos geográficos, curvas de nivel, etc. Con el objeto de brindar una simple interacción al usuario con dicha información, el subsistema contiene herramientas para visualizar una selección de dichas capas y navegación georreferenciada sobre ellas (Ej.: alejar, acercar, trasladar). Además de navegación georreferenciada sobre capas de información geográfica en tiempo real, el sistema permite el almacenamiento de los recorridos y su posterior reproducción.

Cabe destacar que cada simulación en tiempo real puede ser almacenada y posteriormente reproducida [12], ya sea para análisis y optimización de recorridos, corrección de errores, entre otras aplicaciones.

3.4.1 Colaboración en la funcionalidad del sistema global

El visualizador recibe mediante un determinado socket el mensaje de texto tratado por el subsistema dedicado a tal fin y lo parsea. De esta manera obtiene la información necesaria para determinar la ubicación y estado de la unidad y por consiguiente refresca la pantalla con estos nuevos datos.

Tenemos un subsistema capaz de mostrar de manera georreferenciada un móvil emitiendo a intervalos constantes su ubicación y estado. Cabe aclarar que este subsistema puede estar ejecutándose en otra ubicación física que el de tratado de mensajes de texto gracias a la utilización de sockets como medio de comunicación. Esta tecnología condice con el modelo productor-consumidor donde el productor es el subsistema encargado de tratar los mensajes de texto entrantes del cliente y el consumidor por su parte es el presente subsistema. El socket es un canal de comunicación establecido entre dos partes. De un lado se encuentra el cliente y del otro el servidor del socket.

Además, esto permite que los subsistemas sean independientes de la plataforma y del lenguaje en que están desarrollados. En este caso se tiene el subsistema de visualización desarrollado en Delphi y el encargado de tratar los mensajes de texto entrantes en Java.

3.5 Integración de los subsistemas presentados

Luego de haber descrito los tres subsistemas se explica como estos colaboran entre sí.

El microcontrolador embebido en el subsistema de coordenadas geográficas y sensores realiza cada cierto tiempo un ciclo de obtención de datos. Obtiene del GPS una coordenada geográfica, lee la entrada de los sensores, almacena toda esta información y por último envía un SMS al destino con este contenido. Dicho SMS viaja por la red de telefonía celular al número de destino ingresado. El receptor del SMS es un teléfono celular Sony Ericsson W300i, que está conectado a una computadora portátil HP Pavilion dv6835nr mediante tecnología inalámbrica Bluetooth. El puerto virtualizado Bluetooth esta ligado a un puerto de comunicaciones (COM11). El subsistema encargado de tratar los mensajes de texto una vez activo se queda a la espera de la entrada de un mensaje proveniente del módem GSM. Una vez recibido el mensaje lo analiza verificando en primera instancia si el origen del mismo coincide con el esperado. De esta manera se introduce una verificación de

autenticidad del mensaje ya que el mismo puede ser enviado por cualquier dispositivo que tenga acceso a la red GSM y conozca el número de destino. Igualmente el formato del mensaje esta predeterminado por lo cual si cumple con la condición anterior pero no con el formato interno del mismo tampoco será tratado para su envío al socket. Una vez establecido el origen y la estructura del mensaje, el mismo es publicado en un socket determinado. Por último el servidor instanciado en este caso por el subsistema de visualización georreferenciada ejecutándose en la computadora portátil antes mencionada es el encargado de crear el socket servidor y quedarse a la espera. Cuando se produce una entrada en el socket, éste notifica al servidor para que consuma ese ítem. Por lo tanto, el visualizador lo consume y analiza para poder obtener los datos necesarios para la representación visual.

4. Formación de recursos humanos

Durante el desarrollo de los diferentes subsistemas pertenecientes al instituto INTIA división INCA, participaron cuatro estudiantes avanzados de la carrera por el período de dieciocho meses. De los mismos proyectos se realizaron tesinas de grado, además se incorporaron tres estudiantes como pasantes en empresas del polo informático también por un lapso de dieciocho meses.

De los alumnos avanzados y pasantes hoy día tenemos tres becarios, dos de la Agencia y uno de la CIC.

5. Bibliografía

[1] “Información geográfica y sistemas de información geográfica (SIGs)” Juan A. Cebrian, Santander servicio de publicaciones, Universidad de Cantabria 1992, ISBN: 84-87412-81-5. Pág. 10.

[2] “Tecnología de los sistemas de información geográfica” F. Javier Moldes, Editorial RA-MA, Madrid 1995, ISBN: 84-7897-164-5. Págs. 1-6.

[3] Telemetría
<http://es.wikipedia.org/wiki/Telemetría>

[4] “Elementos de teledetección” Carlos Pinilla, Universidad de Jaén, RA-MA editorial, Madrid Noviembre de 1995, ISBN: 84-7897-202-1. Pág. 1.

[5] GSM (Sistema Global de Comunicaciones Móviles)
<http://es.wikipedia.org/wiki/GSM>

[6] GPS definición
<http://www.alegsa.com.ar/Dic/gps.php>

[7] “Microcontroladores PIC: diseño práctico de aplicaciones” José María Angulo Usategui e Ignacio Angulo Martínez, 2º Edición McGraw-Hill Interamericana de España, Madrid 2000, ISBN: 84-4812496-0. Pág. 1.

[8] “Microcontrollers and Microcomputers – Principles of software and hardware engineering” Fredrick M. Cady, Oxford University Press 1997, ISBN: 0-19-511008-0. Pág. 2.

[9] Biblioteca Java SMSLib
<http://smslib.org/>

[10] “Fundamentals of geographic information systems” Michael N. DeMers, John Willey & Sons, Inc., U.S.A. 1997, ISBN: 0-471-14284-0. Pág. 5-7.

[11] “Programming Embedded Systems in C and C++” Michael Barr, O'Reilly & Associates, Inc., Primera Edición Enero de 1999 U.S.A., ISBN: 1-56592-354-5. Págs. 1-2.

[12] ACOSTA N. y MOSCA N. “Proyecto de desarrollo de herramienta para seguimiento de ejercicios tácticos y estrategias de guerra”. En VII Workshop de Investigadores en Ciencias de la Computación - WICC 2005, págs. 429–433. Univ. de Río Cuarto, Córdoba, Argentina. (May 2005).

Estructuras de Datos Métricas para la Recuperación de Información Multimedia en la Web. *

Roberto Uribe Paredes, Eduardo Peña Jaramillo

Departamento de Ingeniería en Computación

Universidad de Magallanes

Punta Arenas, Chile

(roberto.uribe@umag.cl, eduardo.pena@umag.cl)

and

Osiris Sofia

Universidad Nacional de la Patagonia Austral

Río Gallegos, Argentina

osofia@unpa.edu.ar

Resumen

La *búsqueda por similitud* consiste en recuperar todos aquellos objetos dentro de una base de datos que sean parecidos o relevantes a una determinada consulta. Este concepto tiene una amplia gama de aplicaciones en áreas como bases de datos multimediales, reconocimiento de patrones, minería de datos, recuperación de información, etc.

En este contexto, tres grupos de investigación aunan esfuerzos en una misma dirección con el objetivo de posibilitar un mayor avance en torno al diseño, desarrollo e implementación de nuevas y eficientes estructuras métricas, así como también en la construcción de aplicaciones que permitan acercar a la realidad este tipo de investigaciones.

El presente artículo describe algunos de los avances realizados en el último año en torno a esta línea de investigación realizados por grupos conformados por la Universidad de Magallanes, Chile y por las unidades académicas de Río Turbio y Río Gallegos de Universidad Nacional de la Patagonia Austral, Argentina.

Palabras claves: bases de datos, estructuras de datos, algoritmos, espacios métricos, consultas por similitud, paralelismo, modelo BSP, CBIR.

* Este trabajo es parcialmente financiado por los proyectos y programas de investigación: PR-F1-02IC-08, UMAG, Chile; 29/C035 UART-UNPA y 29/A216 UARG-UNPA, Argentina.

1. Introducción

1.1. Antecedentes

Uno de los problemas de gran interés en ciencias de la computación es el de “búsqueda por similitud”, es decir, encontrar los elementos de un conjunto más similares a una muestra. Esta búsqueda es necesaria en múltiples aplicaciones, como ser en reconocimiento de voz e imagen, compresión de video, genética, minería de datos, recuperación de información, etc. En casi todas las aplicaciones la evaluación de la similaridad entre dos elementos es cara, por lo que usualmente se trata como medida del costo de la búsqueda la cantidad de similaridades que se evalúan.

Interesa el caso donde la similaridad describe un espacio métrico, es decir, está modelada por una función de distancia que respeta la desigualdad triangular. En este caso, el problema más común y difícil es en aquellos espacios de “alta dimensión” donde el histograma de distancias es concentrado, es decir, todos los objetos están más o menos a la misma distancia unos de otros.

El aumento de tamaño de las bases de datos y la aparición de nuevos tipos de datos sobre los cuales no interesa realizar búsquedas exactas, crean la necesidad de plantear nuevas estructuras para búsqueda por similaridad o búsqueda aproximada. Asimismo, se necesita que dichas estructuras sean dinámicas, es decir, que permitan agregar o eliminar elementos sin necesidad de crearlas nuevamente, así como también que sean óptimas en la administración de memoria secundaria. La necesidad de procesar grandes volúmenes de datos obligan a aumentar la capacidad de procesamien-

to y con ello la paralelización de los algoritmos y la distribución de las bases de datos.

Las distintas problemáticas mencionadas en el párrafo anterior son abarcadas por diferentes equipos, donde hay que incluir, además, el desarrollo y evaluación de prototipos de prueba.

1.2. Marco teórico

La similitud se modeliza en muchos casos interesantes a través de un espacio métrico, y la búsqueda de objetos más similares a través de una búsqueda por rango o de vecinos más cercanos.

Definición 1 (*Espacios Métricos*): Un espacio métrico es un conjunto X con una función de distancia $d : X^2 \rightarrow R$, tal que $\forall x, y, z \in X$,

1. $d(x, y) \geq 0$ and $d(x, y) = 0$ ssi $x = y$. (*positividad*)
2. $d(x, y) = d(y, x)$. (*Simetría*)
3. $d(x, y) + d(y, z) \geq d(x, z)$. (*Desigualdad Triangular*)

Definición 2 (*Consulta por Rango*): Sea un espacio métrico (X, d) , un conjunto de datos finito $Y \subseteq X$, una consulta $x \in X$, y un rango $r \in R$. La consulta de rango alrededor de x con rango r es el conjunto de puntos $y \in Y$, tal que $d(x, y) \leq r$.

Definición 3 (*Los k Vecinos más Cercanos*): Sea un espacio métrico (X, d) , un conjunto de datos finito $Y \subseteq X$, una consulta $x \in X$ y un entero k . Los k vecinos más cercanos a x son un subconjunto A de objetos de Y , donde la $|A| = k$ y no existe un objeto $y \in A$ tal que $d(y, x)$ sea menor a la distancia de algún objeto de A a x .

El objetivo de los algoritmos de búsqueda es minimizar la cantidad de evaluaciones de distancia realizadas para resolver la consulta. Los métodos para buscar en espacios métricos se basan principalmente en dividir el espacio empleando la distancia a uno o más objetos seleccionados. El no trabajar con las características particulares de cada aplicación tiene la ventaja de ser más general, pues los algoritmos funcionan con cualquier tipo de objeto [6].

Existen distintas estructuras para buscar en espacios métricos, las cuales pueden ocupar funciones discretas o continuas de distancia. Algunos son BKTtree [4], MetricTree [14], GNAT [2], Vp-Tree [19], FQTree [1], MTree [7], SAT [9], Slim-Tree [13], Spaghettis [5], SSS-Tree [3], EGNAT [15].

Algunas de las estructuras anteriores basan la búsqueda en pivotes y otras en clustering. En el primer caso se seleccionan pivotes del conjunto de datos y se precálculan las distancias entre los elementos y los pivotes. Cuando se realiza una consulta, se calcula la distancia de la consulta a los pivotes y se usa la desigualdad triangular para descartar candidatos.

Los algoritmos basados en clustering dividen el espacio en áreas, donde cada área tiene un *centro*. Se almacena alguna información sobre el área que permita descartar toda el área mediante sólo comparar la consulta con su centro. Los algoritmos de clustering son los mejores para espacios de alta dimensión, que es el problema más difícil en la práctica.

Existen dos criterios para delimitar las áreas en las estructuras basadas en clustering, *hiperplanos* y *radio cobertor* (*covering radius*). El primero divide el espacio en particiones de *Voronoi* y determina el hiperplano al cual pertenece la consulta según a qué centro corresponde. El criterio de radio cobertor divide el espacio en esferas que pueden intersectarse y una consulta puede pertenecer a más de una esfera.

1.3. Modelo de computación paralela BSP

El modelo BSP de computación paralela fue propuesto en 1990 con el objetivo de permitir que el desarrollo de software sea portable y tenga desempeño eficiente y escalable [18, 11]. BSP propone alcanzar este objetivo mediante la estructuración de la computación en una secuencia de pasos llamados *supersteps* y el empleo de técnicas aleatorias para el ruteo de mensajes entre procesadores. El computador paralelo, independiente de su arquitectura, es visto como un conjunto de pares procesadores-memoria, los cuales son conectados mediante una red de comunicación cuya topología es transparente al programador. Los *supersteps* son delimitados mediante la sincronización de procesadores. Los procesadores proceden al siguiente *superstep* una vez que todos ellos han alcanzado el final del *superstep*, los cuales son agrupados en bloques para optimizar la eficiencia de la comunicación. Durante un *superstep*, los procesadores trabajan asincrónicamente con datos almacenados en sus memorias locales. Cualquier mensaje enviado por un procesador está disponible para procesamiento en el procesador destino sólo al comienzo del siguiente *superstep*. Dada la estructura particular del modelo de computación, el costo de los programas BSP puede ser obtenido utilizando técnicas similares a las empleadas en el análisis de algoritmos secuenciales. En BSP, el costo de cada *superstep* esta dado por la suma del costo en computación

(el máximo entre los procesadores), el costo de sincronización entre procesadores, y el costo de comunicación entre procesadores (el máximo enviado/recibido entre procesadores).

2. Resultados Preliminares

Durante la primera etapa del trabajo conjunto entre los distintos grupos, básicamente se abordaron tres líneas de trabajo, la primera corresponde al rediseño de estructuras, su implementación y prueba. La segunda línea tiene relación con la paralelización de algoritmos y esquemas de distribución de estructuras sobre un cluster de PC's. Finalmente, la tercera corresponde a la implementación de aplicaciones.

El rediseño de estructuras ha sido orientada a aumentar la eficiencia en las búsquedas, es el caso de los trabajos [16, 17] como del diseño de nuevas estructuras [8]. En la actualidad se está trabajando sobre la optimización en memoria secundaria para el *SSS-Tree* y una nueva versión para la estructura *Lista de Cluster*, para ambos casos se espera presentar los resultados en congresos latinoamericanos durante el presente año.

La paralelización se a abordado sobre la estructura métrica *Spaghettis*, sobre ésta se han experimentado esquemas de distribución de datos sobre un cluster como la paralelización de sus algoritmos. También sobre esta estructura se han hecho modificaciones para darle características dinámicas, específicamente eliminación y inserción, considerando en estos procesos el balance de la estructura en el cluster.

La tercera línea de trabajo está abocada a la implementación de aplicaciones, en este sentido, se implementó, en una etapa temprana, un *Digesto Digital Paralelo para Búsqueda por Similitud sobre Documentos* ([12]) y una segunda aplicación, ahora en etapa de inicial de prueba, se presenta en la siguiente subsección.

2.1. Sistema Recuperador de Imágenes Basado en Contenidos sobre Estructuras Métricas

Recuperar información desde una imagen basada en contenido (*CBIR: Content Based Image Retrieval*) corresponde a una metodología de recuperación con respecto al dominio de aplicación del proceso de recuperación en sí. Usa un análisis y procesamiento digital para generar descriptores a partir de los datos. Los méritos principales de sistemas basados en el contenido son: soporta el procesamiento de consultas visuales, la consulta es intuitiva y amistosa al usuario, la generación de los descriptores es automática, siendo objetiva y consistente.

El prototipo de prueba es una continuación del trabajo desarrollado en [10]. El prototipo está soportado por la estructura *EGNAT*, sin embargo, se realizaron experimentos del mismo *CBIR* sobre la estructura *GNAT* y *Spaghettis*. En la figura 1 se puede observar resultados preliminares para un conjunto de consultas (primera columna) y los 5 primeros objetos recuperados. Este experimento fue realizado sobre una base de datos de 1,000 imágenes y un conjunto de 5 consultas.

3. Conclusiones

En este artículo se ha presentado una descripción breve del contexto y de algunos de los avances logrados por los equipos conformados por investigadores de las Universidades de Magallanes, Chile y Nacional de la Patagonia Austral, Argentina.

Los avances corresponden a trabajos realizados en conjunto en torno a la búsqueda por similitud en espacios métricos. Durante la primera mitad del presente año se espera enviar a evaluar estos resultados a congresos latinoamericanos.

Se espera continuar el trabajo con énfasis en el aumento de la eficiencia en las estructuras diseñadas y la continuación de pruebas sobre los prototipos indicados. Se espera contar, al finalizar el proyecto de investigación conjunto, con el desarrollo de parte de una máquina de búsqueda por similitud, soportada sobre un cluster de PCs, que pueda ser utilizada como prototipo en aplicaciones de tipo real.

Referencias

- [1] R. Baeza-Yates, W. Cunto, U. Manber, and S. Wu. Proximity matching using fixedquery trees. In *5th Combinatorial Pattern Matching (CPM'94)*, LNCS 807, pages 198–212, 1994.
- [2] Sergei Brin. Near neighbor search in large metric spaces. In *the 21st VLDB Conference*, pages 574–584. Morgan Kaufmann Publishers, 1995.
- [3] Nieves R. Brisaboa, Oscar Pedreira, Diego Seco, Roberto Solar, and Roberto Uribe. Clustering-based similarity search in metric spaces with sparse spatial centers. In *SOFSEM 2008: 34rd Conference on Current Trends in Theory and Practice of Computer Science*, volume 4910 of *Lecture Notes in Computer Science*, pages 186–197, Novy Smokovec, High Tatras, Slovakia, January, 19-25 2008. Springer.

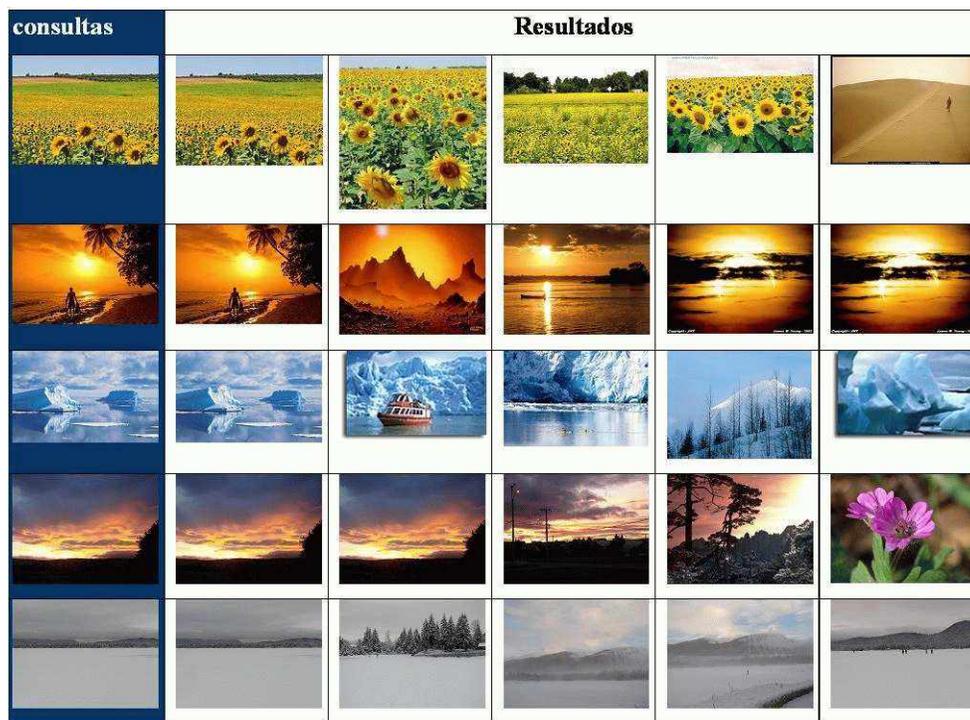


Figura 1: Primeras imágenes recuperadas por el *CBIR* para un conjunto de consultas (primera columna).

- [4] W. Burkhard and R. Keller. Some approaches to best-match file searching. *Communication of ACM*, 16(4):230–236, 1973.
- [5] E. Chavéz, J. Marroquín, and R. Baeza-Yates. Spaghettis: An array based algorithm for similarity queries in metric spaces. In *6th International Symposium on String Processing and Information Retrieval (SPIRE'99)*, pages 38–46. IEEE CS Press, 1999.
- [6] Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates, and José L. Marroquín. Searching in metric spaces. In *ACM Computing Surveys*, pages 33(3):273–321, September 2001.
- [7] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. In *the 23rd International Conference on VLDB*, pages 426–435, 1997.
- [8] Mauricio Marín, Veronica Gil-Costa, and Roberto Uribe. Hybrid index for metric space databases. In *Proc. of International Conference on Computational Science 2008 (ICCS 2008)*, volume 5101 of *Lecture Notes in Computer Science*, pages 327–336, Krakow, Poland, Jun 2008. Springer.
- [9] Gonzalo Navarro. Searching in metric spaces by spatial approximation. *The Very Large Databases Journal (VLDBJ)*, 11(1):28–46, 2002.
- [10] Eduardo Peña-Jaramillo. Estructuras métricas paralelas en la recuperación de imágenes. Master's thesis, Escuela de Ingeniería, Departamento de Ciencias de la Computación, Pontificia Católica de Chile, Santiago, Chile, Nov. 2006.
- [11] D.B. Skillicorn, J.M.D. Hill, and W.F. McColl. Questions and answers about BSP. Technical Report PRG-TR-15-96, Computing Laboratory, Oxford University, 1996. Also in *Journal of Scientific Programming*, V.6 N.3, 1997.
- [12] Roberto Solar, Roberto Uribe-Paredes, Esteban Gesto, and Osiris Sofia. Implementación de un digesto digital paralelo para búsquedas por similitud sobre documentos. In *Congreso Argentino de Ciencias de la Computación*, La Rioja, Argentina, Octubre 2008. CACIC 2008.
- [13] Caetano Traina, Agma Traina, Bernhard Seeger, and Christos Faloutsos. Slim-trees: High performance metric trees minimizing overlap between nodes. In *VII International Conference on Extending Database Technology*, pages 51–61, 2000.
- [14] J. Uhlmann. Satisfying general proximity/similarity queries with metric trees.

In *Information Processing Letters*, pages 40:175–179, 1991.

- [15] Roberto Uribe-Paredes. Manipulación de estructuras métricas en memoria secundaria. Master's thesis, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile, Abril 2005.
- [16] Roberto Uribe-Paredes, Claudio Márquez, and Roberto Solar. Estrategias de construcción sobre estructuras métricas para búsquedas por similitud. In *Conferencia Latinoamericana de Estudios en Informática (CLEI2008)*, Santa Fé, Argentina, 2008. CLEI 2008.
- [17] Roberto Uribe-Paredes, Claudio Márquez, and Roberto Solar. Sstree v2.0: Búsqueda por similitud en espacios métricos con solapamiento de planos. In *Congreso Argentino de Ciencias de la Computación*, La Rioja, Argentina, Octubre 2008. CACIC 2008.
- [18] L.G. Valiant. A bridging model for parallel computation. *Comm. ACM*, 33:103–111, Aug. 1990.
- [19] P. Yianilos. Data structures and algorithms for nearest neighbor search in general metric spaces. In *4th ACM-SIAM Symposium on Discrete Algorithms (SODA'93)*, pages 311–321, 1993.

METODOLOGÍA PARA ESPECIFICACION DE REQUISITOS EN PROYECTOS DE EXPLOTACION DE INFORMACION

F. Pollo-Cattaneo, P. Britos, P. Pesado, R. García-Martínez

Programa de Doctorado en Ciencias Informáticas. Facultad de Informática. UNLP
Centro de Ingeniería de Software Ingeniería del Conocimiento. Escuela de Postgrado. ITBA
Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP - CIC
Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA

{fcattane,pbritos}@itba.edu.ar, ppesado@lidi.info.unlp.edu.ar, rgarciamar@fi.uba.ar

CONTEXTO

El Proyecto articula líneas de trabajo del Proyecto “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaria de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires.

RESUMEN

Este proyecto tiene como objetivo general definir una metodología de educación de requisitos para proyectos de explotación de información que permita ser integral al ciclo de vida de este tipo de proyectos. Se considera que el proyecto implica investigación básica en la definición del proceso de educación de requisitos de este tipo de proyectos.

Palabras clave: explotación de información, proyectos, ingeniería de requisitos, especificación.

1. INTRODUCCION

La Ingeniería de Requerimientos es una fase importante en las metodologías de Ingeniería del Software (IEEE, 1993; Winter & Strauch 2004; Maiden et al. 2004, 2007; Solheim et al. 2005; Jiang & Eberlein 2007) la que permite especificar las necesidades de los clientes. Las Metodologías de Explotación de Información buscan organizar los procesos de descubrimientos de patrones en el datawarehouse de la organización. Estas metodologías consideran la especificación de requerimientos como una fase temprana en las actividades de este tipo de proyectos. (Chapman et al. 2000; Pyle 2003).

En (Winter & Strauch 2002; Silva & Freire 2003; Yang and Wu, 2006) se destaca la necesidad que tienen las metodologías de Explotación de Información en focalizarse en la definición de objetivos y sus tareas especialmente en la fase de exploración de los datos; proponen el uso de herramientas conceptuales para la documentación de estos procesos, la construcción de modelos y la

búsqueda de patrones. La comunidad dedicada a la Explotación de Información no presta mucha atención a los aspectos a tener en cuenta en la especificación de requerimientos, no identifica técnicas de elicitación ni sugiere plantillas para una documentación sistemática.

En el dominio de la Explotación de Información, durante los procesos de elicitación de requerimiento son identificados los conceptos relativos a la extracción, transformación, agregación y descubrimiento de patrones negocios dentro de la organización.

Una suposición de los Ingenieros en Requerimientos involucrados en proyectos de Explotación de Información es que los recursos humanos involucrados en el proyecto conocen lo suficiente acerca de los requerimientos. Es conocido que en situaciones típicas, clientes y aun usuarios “hablen en otro lenguaje”, lo mismo que el equipo de desarrolladores (Maiden et al. 2007). La tarea de traducir a clientes y usuarios es realizados por los Ingenieros en Requerimientos y los Analistas del Negocio utilizando diversas notaciones (Jiang & Eberlein 2007).

En este contexto los stakeholders y los ingenieros de requerimientos trabajan en forma conjunta para identificar “que” y “donde” buscar dentro de los datos de las organizaciones, para proveer las bases del descubrimiento de patrones de negocios. El proceso de elicitación de requerimiento es direccionado comúnmente por el uso de metodologías de explotación de datos (Chapman et al, 2000; Pyle, 2003, SAS, 2008), estas mencionan la necesidad de entender en negocio como punto de partida para el desarrollo de este tipo de proyectos.

La metodología CRISP-DM (Chapman et al, 2000) consisten en 4 niveles de abstracción jerárquicos organizados desde las áreas generales a los casos específicos. El proceso esta divide en 6 fases, cada una de las cuales tienen subfases. Las tareas generes son proyectadas como una, donde las acciones deben ser desarrolladas para situaciones descriptas específicamente. Como consecuencia, nos encontramos con tareas muy generales, por ejemplo “limpieza de datos”; para lo cual se cuenta con un

tercer nivel en el cual se desarrollan tareas específicas para esos casos, como por ejemplo “limpieza de datos numéricos”, o “limpieza de datos categóricos”. Un cuarto nivel recolecta las acciones del grupo, decisiones y resultados específicos del proyecto de explotación de información. La metodología CRISP-DM presenta dos documentos diferentes como herramientas de ayuda durante el desarrollo del proyecto: el modelo de referencia y la guía de usuario. El modelo de referencia describe en términos generales las fases, tareas generales y salidas previstas en cada una de ellas. La guía de usuarios brinda en detalle la documentación acerca de la aplicación del modelo de referencia en proyectos de explotación de datos; también sugiere lista de validación acerca de cada una de las fases.

La metodología P³TQ (Product, Place, Price, Time, Quantity) consta de dos partes (Pyle, 2003): [a] Modelado (PI): provee una guía paso a paso para desarrollar y construir un modelo de negocio, problema u oportunidad. El modelado depende de las circunstancias de negocios que sean señaladas, en primer lugar identifica 5 escenarios para esta fase. Principalmente provee una lista de acciones a ser completadas dependiendo del escenario planteado; y [b] Data Mining (PII): provee una guía paso a paso de cómo llevar adelante el proceso de explotación de información para el modelo indicado en la fase anterior (PI). Esta fase consiste en una serie de acciones que deben ser completados en orden. Para los diversos modelos se deben realizar un conjunto de tareas al mismo tiempo, el proceso de explotación pasa de una actividad a otra. Cada una de las partes está basada en 4 tipos de “cajas de actividades”; *cajas de acciones*: indican una o más actividades requeridas en los próximos pasos a realizar; *cajas de descubrimiento*: que proveen acciones de exploración necesarias para realizar la acción y decidir que realizar en el próximo paso; siempre contienen una “acción de descubrimiento” la cual tiene resultados asociados, interpretaciones y posibles problemas; *cajas técnicas*: proveen información suplementaria acerca de las recomendación de los pasos descritos en las cajas de acciones o de descubrimiento; y *cajas de ejemplo*: proveen de una descripción detallada de cómo usar una técnica específica.

SEMMA es una metodología orientada a seleccionar, explotar y modelar un gran conjunto de datos; destinado al descubrimiento de patrones de negocio (SAS, 2008). El proceso comienza con la extracción de una muestra de los datos para los cuales el análisis es aplicado. Con la muestra seleccionada, la metodología propone explotar los datos en orden a simplificar el modelo. Una tercera fase está orientada a seleccionar el algoritmo de explotación de datos más adecuado. La cuarta fase está orientada a ejecutar el algoritmo seleccionado con la muestra. La última fase consiste en la

evaluación de los resultados a través del contraste con modelos estadísticos o nuevas muestras.

En este contexto, las metodologías no cubren adecuadamente la fase de elicitación de requerimiento, los conceptos necesarios ni su correspondiente documentación.

2. LINEAS DE INVESTIGACION y DESARROLLO

Para construir el conocimiento asociado al presente proyecto de investigación, se seguirá un enfoque de investigación clásico [Kumar, 1996; Creswell, 2003; Marczyk, *et al*, 2005] en el que se han identificado métodos y materiales necesarios para desarrollar el proyecto. Se dispone de los materiales: [a] metodologías CRISP-DM (Chapman, *et al*, 2000), P³TQ (Pyle, 2003) y SEMMA (SAS, 2008) para identificar los conceptos necesarios para la educación de conocimiento en proyectos de explotación de información, [b] norma IEEE 830-1993 (IEEE, 1993) sobre educación de requisitos para la ingeniería del software, [c] manual SWEBOOK (IEEE, 2004), sobre educación de requisitos en ingeniería del software, [d] técnicas de educación de conocimientos para sistemas inteligentes (García-Martínez, *et al*, 2004). Y se prevé utilizar el método basado en las siguientes tareas: [a] se buscará identificar los conceptos necesarios a educir en proyectos de explotación de información a través de revisión bibliográfica, estudio de casos y consulta a expertos, [b] identificación de la relación existente entre los diversos conceptos a educir, [c] identificación de técnicas de elicitación de requisitos para poder obtener los conceptos de este tipo de proyecto a través de revisión bibliográfica y consulta con expertos, [d] se propondrán: plantillas para la educación de conceptos, técnicas de elicitación para los conceptos a educir, cuestionarios modelos para facilitar la educación de conceptos.

En este proyecto se realizarán investigaciones sobre:

- [a] Las distintas metodologías para proyectos de explotación de información.
- [b] Los conceptos necesarios a ser educidos para estas metodologías.
- [c] Las técnicas de educación de conocimiento aplicables a la identificación de requisitos de proyectos de software.
- [d] La fiabilidad de las técnicas identificadas.
- [e] La relación entre los conceptos a ser educidos con la metodología propuesta.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se ha avanzado en la identificación de conceptos que deben ser educidos en el dominio del proyecto de explotación de información e identificado la necesidad de definir procesos que definan como llevar esa educación y herramientas que den soporte a la documentación de la misma [Britos *et al*, 2008].

4. FORMACION DE RECURSOS HUMANOS

En el marco de este proyecto se esta desarrollando una tesis doctoral y tres tesis de ingeniería informática (una de ellas defendida y dos en curso).

5. BIBLIOGRAFIA

- Britos, P., Dieste, O., García-Martínez, R. (2008). *Requirements Elicitation in Data Mining for Business Intelligence Projects*. IFIP Series, 274: 139–150.
- Chapman P, Clinton J, Keber R, Khabaza T, Reinartz T, Shearer C, Wirth R (2000) *CRISP-DM 1.0 Step by step BIguide Edited by SPSS*. <http://www.crisp-dm.org/CRISPWP-0800.pdf>
Acceso Marzo 2008.
- García Martínez, R. y Britos, P. (2004). *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4
- IEEE (1993) *Standard IEEE 830-1993: Recommended Practice for Software Requirements Specifications*. Institute of Electronic and Electrical Engineers Press.
- IEEE (2004) *Guide to the Software Engineering Body of Knowledge*. IEEE Comp. Society Press
- Jiang L, Eberlein A (2007) *Selecting Requirements Engineering Techniques based on Project Attributes - A Case Study*. 14th Annual IEEE ECBS: 269-278
- Maiden N, Ncube C, Robertson S (2007) *Can Requirements Be Creative? Experiences with an Enhanced Air Space Management System* Proceedings 29th ICSE: 632-641
- Maiden N, Robertson S, Gizikis A (2004) *Provoking Creativity: Imagine What Your Requirements Could be Like*. IEEE Software 21(5): 68-75
- Pyle D (2003) *Business Modeling and Business intelligence*. Morgan Kaufmann
- SAS (2008) *SAS Enterprise Miner: SEMMA* <http://www.sas.com/technologies/analytics/data-mining/miner/semma.html>. Acceso Marzo 2009
- Silva F, Freire J (2003) *DWARF: An Approach for Requirements Definition and Management of Data Warehouse Systems*. RE'03: 75-84
- Solheim H, Lillehagen F, Petersen S, Jorgensen H, Anastasiou M (2005) *Model-driven visual requirements engineering* Proceedings RE'05: 421-428
- Winter R, Strauch B (2002) *A Method for Demand-driven Information Requirements Analysis in Data Warehousing Projects*. HICSS-36:231-239
- Yang Q, Wu X (2006) *10 Challenging Problems in Data Mining Research*. Int. J. Inf. Tech. & Decis. Mak. 5(4):597–604

MÉTODOS DE EDUCACIÓN EN INGENIERÍA DE SISTEMAS INTELIGENTES

Enrique Calot, Ramón García-Martínez

Programa de Doctorado en Ingeniería, Facultad de Ingeniería, Universidad de Buenos Aires.
Laboratorio de Sistemas Inteligentes, Facultad de Ingeniería, Universidad de Buenos Aires.

ecalot@fi.uba.ar, rgarciamar@fi.uba.ar

CONTEXTO

El Proyecto articula líneas de trabajo del Proyecto “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaría de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires.

RESUMEN

Enmarcado dentro del procesamiento de la información y la minería de datos se encuentra el problema del reconocimiento de patrones. Un patrón es un conjunto de características comunes que pueden ser localizadas en una región, admitiéndose un cierto error o tolerancia. Estas regiones pueden ser cadenas de texto, de ADN o imágenes entre otras cosas.

El desarrollo de algoritmos computacionales que permiten buscar sobre estas regiones representa un área en crecimiento. Pese a esto, la incorporación de sistemas inteligentes a la misma todavía requiere mucho desarrollo.

Esta línea de investigación pretende incorporar estas técnicas a algoritmos de este tipo; definir y enmarcar los problemas existentes y buscar soluciones a los mismos dentro de este campo.

Palabras clave: Reconocimiento de patrones, sistemas inteligentes, búsqueda difusa

1. INTRODUCCION

Muchas veces se tiende a modelizar los problemas mediante sistemas matemáticos rígidos y sin error. En algunas áreas, incluso, puede estar mal vista la consideración de un modelo que admita error y luego la búsqueda de su minimización.

Es importante reconocer que todo lo que es procesado por la inteligencia natural admite internamente errores y que la inteligencia artificial, en parte, como su emulación, también lo hace.

La utilización de algoritmos difusos de búsqueda sobre cadenas de texto da de una utilidad inmediata en campos crecientes como lo son la educación del

conocimiento y la biología computacional [Navarro. 2001].

Nuestro interés radica en la creación y luego aplicación de algoritmos inteligentes y/o difusos basados en actuales algoritmos convencionales.

La utilidad de algoritmos de este tipo se ve ejemplificada en los algoritmos de búsqueda sobre cadenas de caracteres. Éstos suelen ser una comparación booleana “si/no”, mientras que se utiliza muy poco las comparaciones por distancia entre el patrón y la posición en el texto a comparar.

Otro ejemplo es la búsqueda sobre imágenes, donde lo que se busca es un patrón común, como puede ser una figura o propiedad, sobre un arreglo de píxeles de 2 y hasta 3 dimensiones.

Existe un enorme número de algoritmos destinados a resolver este tipo de problemas, muchos utilizan distintas nociones de distancias entre el patrón y la región de búsqueda admitiendo errores y basándose su minimización.

La aplicación de estos algoritmos tiene mucha importancia en el área de minería de datos ya que los bancos de datos pueden contener errores humanos – denominados ruido – en la carga y no siempre se los puede asumir correctos. Este ruido no solo puede estar en las relaciones de identificadores y claves, sino también en los campos de texto y/o imágenes. En el proceso de minería de datos muchas veces es necesaria la comparación basada en este tipo de campos y eso conlleva a la creación de algoritmos que sean capaces de calcular distancias entre datos o bien encontrar patrones comunes

2. LINEAS DE INVESTIGACION Y DESARROLLO

Dentro de lo que a la minería de datos representa, el tema a investigar es el reconocimiento de patrones difusos o en entornos ruidosos, es decir, donde la clasificación admite un porcentaje de error.

Dentro de este tema se propone la línea de investigación que combina las técnicas utilizadas en sistemas inteligentes con la búsqueda y reconocimiento de patrones en arreglos tanto lineales (cadenas de textos) como multidimensionales (imágenes o volúmenes).

Sobre esta línea se ha publicado un trabajo introductorio previo sobre educación de información

[Calot, *et al.* 2008] mediante *repertory grids* y otro sobre reconocimiento de patrones con redes neuronales *back propagation* aplicados a sistemas médicos –en este caso reconocimiento de lesiones mamarias– [Calot *et al.*, 2009].

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto se encuentra en su fase inicial desde marzo de 2009 y prevén las siguientes tareas para ser realizadas en sus primeros dos años:

- Refinamiento e investigación documental orientada a la identificación de trabajos previos vinculados a la utilización de sistemas inteligentes y/o difusos en la detección de patrones.
- Identificación de las posibles futuras líneas de trabajo de los mismos que puedan ser realizadas mediante sistemas inteligentes.
- Identificación de las últimas técnicas utilizadas en el área de la detección de patrones.
- Identificación de casos de estudio aceptados por la comunidad científica en general para ser utilizados en pruebas de concepto y validación del proyecto.
- Identificación y delimitación de problemas vinculados a la búsqueda de patrones mediante estas técnicas
- Exploración y propuesta de nuevos algoritmos que resuelvan estos problemas.
- Validación de pruebas de concepto. Cálculo del orden y estimación de la velocidad de estos algoritmos y contraste de los mismos frente a otros.
- Creación de ejemplos de aplicación que prueben su utilidad concreta.

4. FORMACION DE RECURSOS HUMANOS

En el marco de este proyecto se esta desarrollando una tesis doctoral y una tres tesis de ingeniería informática.

5. BIBLIOGRAFIA

- BAEZA-YATES, R., NAVARRO, G. 1996. *A faster algorithm for approximate string matching*. In Dan Hirschberg and Gene Myers, editors, *Combinatorial Pattern Matching (CPM'96)*, LNCS 1075, pages 1-23, Irvine, CA.
- BISHOP, C. M. 2005. *Neural Networks for Pattern Recognition*. Oxford University Press. ISBN 0198538642, 9780198538646.
- CALOT, E., BRITOS, P., GARCÍA-MARTÍNEZ, R. 2008. *Optimizing Relationships Information in Repertory Grids*. En *Artificial Intelligence and Practice II*, Max Bramer Ed. (Boston: Springer), IFIP Series, 276: 163-172. ISBN 978-0387-09694-0.
- CALOT, E., MERLINO, H., RANCAN, C., GARCÍA-MARTÍNEZ, R. 2009. *Tumor Classification on Mammographies Based on BPNN and Sobel Filter*. Serie: *Studies in Computational Intelligence*, Springer-Verlag (en prensa).

LINHARTA, C., SHAMIR, R. 2009. *Faster pattern matching with character classes using prime number encoding*. *Journal of Computer and System Sciences*. Volume 75, Issue 3, May 2009, Pages 155-162.

MYERS, G. 1999. *A fast bit-vector algorithm for approximate string matching based on dynamic programming*. *J. ACM* 46, 3, 395-415. Earlier version in *Proceedings of CPM'98 (LNCS, vol. 1448)*.

NAVARRO, G. 2001. *A guided tour to approximate string matching*. *ACM Computing Surveys (CSUR) archive* 33(1), pp 31-88.

UKKONEN, E., WOOD, D. 1993. *Approximate string matching with suffix automata*. *Algorithmica* 10, 353-364. Preliminary version in Rep. A-1990-4, Dept. of Computer Science, Univ. of Helsinki, Apr. 1990.

ZOBEL, J., DART, P. 1996. *Phonetic string matching: lessons from information retrieval*. In *Proceedings of the 19th ACM International Conference on Information Retrieval (SIGIR '96)*, 166-172.

E-CITIZEN E E-LEARNING. UTILIZANDO LAS TICS PARA REDUCIR LA BRECHA DIGITAL Y MEJORAR LOS SERVICIOS AL CIUDADANO.

Beatriz Depetris⁽¹⁾, Guillermo Feierherd⁽¹⁾, Armando De Giusti⁽²⁾, Patricia Pesado^(2,3), Cecilia Sanz⁽²⁾, Marcelo Naiouf⁽²⁾

⁽¹⁾*Grupo de Investigación en Tecnología Informática Aplicada (GITIA)
Facultad de Ingeniería. UNPSJB (Sede Ushuaia)*

Darwin y Canga – Ushuaia (TF) – CP 9410

⁽²⁾*Instituto de Investigación en Informática LIDI (III LIDI)*

Facultad de Informática. UNLP

50 y 120 – 2do Piso - La Plata (BA) – CP 1900

⁽³⁾*Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)*

{depetrisb,feierherdge}@speedy.com.ar, {degiusti,ppesado,csanz,mnaiouf}@lidi.info.unlp.edu.ar

CONTEXTO

Esta línea de investigación corresponde a los Proyectos “E-Citizen e E-Learning. Utilizando las TICS para reducir la brecha digital y mejorar los servicios al ciudadano” de la UNPSJB Sede Ushuaia y “Sistemas de Software Distribuidos. Aplicaciones en procesos industriales, E-government y E-learning” del Instituto de Investigación en Informática LIDI de la Facultad de Informática acreditado por la UNLP.

RESUMEN

Actualmente se observa una utilización creciente de las nuevas tecnologías de la información y la comunicación (TICS) para mejorar la prestación de servicios al ciudadano: E-Government (más adecuadamente E-democracy), E-Learning, E-Health, son algunos de los casos de denominaciones que atienden a la presencia de las TICS para mejorar la prestación de distintos servicios. Obviamente, para que los resultados obtenidos estén a la altura de las expectativas, es preciso que exista un E-Citizen, es decir, un ciudadano en condiciones de utilizar de modo significativo las tecnologías que se utilizan para mediar la prestación de los distintos servicios. Este proyecto busca trabajar diferentes líneas de la tecnología que se requieren para ejercer la E-ciudadanía, aprovechando la experiencia previa en E-Learning y la posibilidad de transferir resultados a la comunidad regional. De este modo se plantean dos ejes temáticos: Gobierno (Democracia) Electrónico y Educación a Distancia y se los relaciona en las temáticas de

reducción de la brecha digital e integración de servicios para el ciudadano.

Palabras clave: *E-Learning, E-Democracy, E-Government, E-Citizen*

1. INTRODUCCION

Desde hace unos años los temas vinculados a gobierno electrónico han merecido especial interés, tanto por parte del público en general como de los profesionales de distintas disciplinas.

En particular este grupo de investigación comparte la idea de que el gobierno electrónico (aplicación de las TICS para mejorar la prestación de los servicios gubernamentales a los ciudadanos) debe considerarse inserto en un campo más amplio que es el de la democracia electrónica. Esta convicción se fundamenta en el hecho de que los receptores de los servicios del Estado (aquellos que son objeto de lo que estudia el gobierno electrónico) no pueden ser considerados exclusivamente *clientes* que deben recibir una atención satisfactoria, sino que se trata de *ciudadanos*, lo que los convierte a su vez en propietarios de las organizaciones que les brindan los servicios.

A su vez el concepto de democracia electrónica, más abarcativo, consiste en la aplicación de TICS en las distintas etapas del proceso democrático con el fin de favorecer la participación de los ciudadanos electrónicos (E-Citizens).

Es evidente que para lograr la participación de la mayor parte de los ciudadanos es condición necesaria (aunque no suficiente) que los mismos estén en condiciones de participar, lo que implica que, entre otros requisitos, deben estar en condiciones de utilizar de modo significativo las

tecnologías que se emplean para mediar en las distintas etapas del proceso democrático. El fenómeno denominado brecha digital (para algunos fractura digital) se refiere a la existencia de importantes diferencias entre los que pueden y no pueden acceder a las TICs y los que pueden y no pueden hacer un uso significativo de ellas. En la explicación de estas diferencias se recurre a distintos factores económicos y culturales. La intención de este grupo es trabajar sobre herramientas que permitan disminuir la brecha digital causada por algunos factores culturales, auxiliándose para ello con las mismas tecnologías que se deberán utilizar en el proceso democrático (E-Learning) combinando su uso con instancias presenciales cuando sea necesario (Blended Learning).

Por otra parte, la aplicación de tecnologías para facilitar la participación de los ciudadanos en el proceso democrático (incluyendo en este proceso el acceso a los servicios que directa o indirectamente presta el Estado), resulta de interés para las disciplinas de las ciencias sociales y la informática, ocupadas ambas en buscar soluciones (sociales y tecnológicas) a los problemas de la brecha digital.

Para el caso de las ciencias informáticas el tema presenta diversas cuestiones de interés. Además de los tradicionalmente vinculados al E-Learning (que plantearé a continuación), los vinculados a la E-Democracy. Estos incluyen el uso pasivo y activo de las tecnologías para distribuir información, para recoger opiniones y para permitir la participación, así como los que se relacionan con el procesamiento de las participaciones de los ciudadanos, abarcando en este caso investigaciones relacionadas al procesamiento del lenguaje natural (NLP) (que permitan procesar las participaciones obtenidas de un modo no estructurado), hasta los mecanismos que hagan factible todas las etapas del voto electrónico en sus distintas variantes (reservado para aquellos casos en los que hay una fuerte estructuración de las consultas).

Por el lado del E-Learning, el diseño de metodologías y el desarrollo de herramientas que permitan incorporar las TICs en el proceso de enseñanza y de aprendizaje constituyen temas de investigación y aplicación permanente dentro de las Ciencias de la Educación y de la Computación. La importancia de estos temas para ambas disciplinas es consecuencia del crecimiento de la Educación a Distancia (hoy casi un sinónimo de educación mediada por las TICs), producto a su vez de una necesidad de actualización continua de los conocimientos requeridos tanto para un correcto desempeño profesional como para el ejercicio de una ciudadanía responsable, actualización que no puede ser satisfecha por los mecanismos clásicos del aula presencial.

Corresponde señalar aquí que tanto la provincia de Tierra del Fuego como las instituciones universitarias involucradas en este proyecto

(Universidades Nacionales de La Plata y de la Patagonia San Juan Bosco) presentan un conjunto de condiciones y antecedentes que permiten suponer el interés en utilizar las conclusiones de un emprendimiento de esta naturaleza.

Para mencionar algunos, los distintos mecanismos de participación ciudadana incorporados a la moderna Constitución de la Provincia de Tierra del Fuego (consulta popular, iniciativa popular y revocatoria de mandatos), así como a las aún más modernas Cartas Orgánicas Municipales de las ciudades de Ushuaia y Río Grande (consulta e iniciativa populares, presupuesto participativo), requieren para su implementación eficiente la aplicación de TICs y la existencia de ciudadanos en condiciones de utilizarlas significativamente. Por otra parte, cabe señalar que en la Municipalidad de Ushuaia se produjo una primera experiencia de voto electrónico en las elecciones municipales de 2003 que, aunque debe ser perfeccionada y no ha tenido continuidad, permite suponer una cierta predisposición a la incorporación de TICs en el proceso democrático.

En el caso de la Universidad Nacional de La Plata ya se han aplicado mecanismos de voto electrónico y de voto a través de Internet para resolver las dificultades que plantean las elecciones del claustro alumnos en la Facultad de Informática (claustro único con alumnos en Sedes alejadas). La extrapolación de dicha experiencia al ámbito de la Universidad Nacional de la Patagonia San Juan Bosco (caracterizada por ser una Universidad regional con Sedes distantes en algunos casos miles de kilómetros), es inmediata y sólo requiere contar con el interés y el aval de las autoridades universitarias y de los demás actores involucrados.

En el campo de la aplicación directa de los desarrollos propuestos en el Proyecto, se trata básicamente de extender la formación adquirida por varios de los miembros del proyecto para la aplicación de las TICs a cursos universitarios, a los problemas vinculados a la disminución de la brecha digital y a la formación de E-Ciudadanos.

La Sede Ushuaia de la Facultad de Ingeniería de la UNPSJB ha celebrado convenios marcos con la Municipalidad y con el Concejo Deliberante de la ciudad de Ushuaia, con el fin de facilitar la realización de acciones conjuntas cuyos detalles deben plasmarse en Actas Acuerdo específicas. Distintas cuestiones vinculadas a los objetivos del proyecto forman parte de posibles Actas Acuerdo.

Asimismo, la UNPSJB ha firmado (en el marco de un convenio entre el CIN y el PAMI), un acuerdo para la capacitación de adultos mayores en el uso de TICs el cual ya ha comenzado a implementarse en la sede Ushuaia.

2. LINEAS DE INVESTIGACION Y DESARROLLO

- Las TICs y sus aplicaciones en educación.
- Entornos de aprendizaje virtuales.
- Enseñanza y aprendizaje centrados en la Web.
- Evaluación de calidad en procesos educativos semipresenciales y a distancia.
- Realidad virtual y sus aplicaciones en educación.
- E-Citizen. Modelo de Aplicaciones.
- Tecnología GRID para el E-Citizen.
- Voto electrónico

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto ha comenzado este año, por lo que a la fecha no se dispone de resultados. Los objetivos que nos hemos fijado dentro del proyecto son:

1. Investigar en soluciones tecnológicas para temas de Democracia y Gobierno Electrónico, en particular en cuatro grandes áreas:
 - Voto Electrónico.
 - Uso de TICs para favorecer la participación ciudadana.
 - Servicios WEB con identificación segura.
 - Integración de datos heterogéneos (ej. Fichas de hospitales)
2. Investigar en la utilización de entornos virtuales para la capacitación del ciudadano en temas específicos, en particular en cuatro grandes áreas:
 - Formación continua por áreas del conocimiento (ejemplo: actualización de médicos/enfermeros).
 - Reducción de la brecha digital en sectores socialmente postergados.
 - Voto electrónico y sus variantes (ej. presupuesto participativo, consultas no vinculantes).
 - Servicios WEB e interacción ciudadano-estado.
3. Formación de recursos humanos con capacidad de desarrollar e integrar tecnologías de Informática y Comunicaciones para aplicarlas en los campos del proyecto.

4. FORMACION DE RECURSOS HUMANOS

Uno de los beneficios más valiosos del proyecto será la consolidación de parte del grupo de I/D que se formara inicialmente con los proyectos anteriores (*“Evaluación y desarrollo de herramientas multimediales para análisis de competencias y aplicación de una metodología didáctica para mejorar el aprendizaje inicial en Informática”*, *“Metodologías y herramientas para la educación no presencial utilizando tecnología multimedial”* y

“Experiencias de intervenciones docentes en espacios virtuales”) oportunamente apoyados por la UNPSJB y en los que han participado recursos humanos de ambas instituciones.

5. BIBLIOGRAFIA

Abbey B.; Instructional and Cognitive Impacts of Web-Based Education; Idea Group Publishing, London, 2000

Alakeson V., Aldrich T., Goodman J., Jorgensen B., Miller P; 2003; Social Responsibility in the Information Society. Brussels: DEESD - Digital Europe: e-business and sustainable development. March, Report Deliverable 7 (D7) DEESD IST-2000-28606

Biblioteca Digital de la ACM (Magazines: eLearn; Proceedings of Special Interest Groups: on Computer and Society (SIGCAS), on Computer Science Education (SIGCSE) on Computer Uses in Education, etc.)

Biblioteca Digital de la Computer Society de IEEE (Magazines: Distributed Systems, Internet Computing, Multimedia, Software, Security & Privacy; Transactions: on Learning Technologies, on Parallel & Distributed Systems, etc.)

Caballero Sybil L.; Prácticas emergentes: la ciberdemocracia, las telecomunidades de conocimiento y los telecentros como alternativas para el desarrollo; CDC, jan. 2005, vol. 22, no. 58, p.97-114. ISSN 1012-2508.

Cabero J.; 2001; Las TICs: una conciencia global en la educación. En VV.A.A., Ticomur. Jornadas Nacionales TIC y Educación. Murcia:CEP de Lorca, XIX-XXXXVI.

Cantijoch Cunill, M.; 2005; El voto electrónico ¿Un temor justificado?, Revista TEXTOS de la CiberSociedad, 7. Temática Variada. Disponible en <http://www.cibersociedad.net>.

Center for democracy and technology; 2002; E-Government Handbook; <http://www.cdt.org/egov/handbook/>

Clift S. L.; 2004; e-Government and Democracy: Representation and Citizen Engagement in the Information Age (February)

De Giusti A., Sanz C., Gonzalez A., Zangara A., Ibáñez E., Iglesias L.; Entorno de Aprendizaje WebINFO y sus posibilidades para el trabajo colaborativo; Publicado en anales y CD Global Congress on Engineering and Technology Education - GCETE05. San Pablo, Brasil. Marzo 2005 Pág: 1767-1771

Diaz Barriga F.; 2005; Principios de diseño instruccional de entornos de aprendizaje apoyados con TIC: un marco de referencia sociocultural y situado. Tecnología y Comunicación Educativas. Nº 41. 4-16.

European Commission; 2003; The Role of eGovernment for Europe's Future. Brussels: European Commission. September 26, Report COM(2003) 567 Final, 26 p

European Commission – Eurydice. Focus on the Structure of Higher Education in Europe 2003/04; National Trends in the Bologna Process; Eurydice

European Unit. September 2003.
<http://www.eurydice.org>

Feierherd G., De Giusti A., Pesado P., Depetris B.; Una aproximación a los requerimientos del software de voto electrónico de Argentina; CACIC 2004

GOV3; 2006; Citizen Centric Government: Global Best Practice in Delivering Agile Public Services to Citizens and Businesses. London: Gov3 Ltd, 22 p

Lave J., Wenger E.; 1998; Communities of Practice: Learning, Meaning, and Identity; Cambridge University Press. También consultado el 10 de julio de 2008 en <http://www.learning-theories.com/communities-of-practice-lave-and-wenger.html>

Malbrán M., Villar C.; Agosto, 2000; VirtualMente (Technology test). Un espacio diferente para el desarrollo de procesos cognitivos. PIP. CONICET. Publicado en actas de las 4tas. Jornadas de Educación a Distancia del MERCOSUR, Buenos Aires (junio 2000) y de la 3ra. Reunión Regional de América Latina y el Caribe del ICDE, San Pablo, Brasil.

Mena M. y otros; El diseño de proyectos de educación a distancia. Páginas en Construcción; La Crujía, Buenos Aires, 2005

Muilenburg L.Y., Berge Z.L.; Barriers to distance education: A factor-analytic study; The American Journal of Distance Education. 15(2): 7-22. 2001.

O'Donnell S., McQuillan H., Malina A.; 2003; eInclusion: expanding the Information Society in Ireland. Dublin: Government of Ireland, Information Society Commission.

Office of Management and Budget; 2002; E-Government Strategy: Simplified Delivery of Services to Citizens. Washington DC: Office of Management and Budget. February 27.

Pasini A, Estrebou C.; 2005; Voto Electronico; XIII Jornadas de Jóvenes Investigadores AUGM. Setiembre, 2005

Pesado P, Feierherg G, Pasini A.; Especificación de requerimientos para sistemas de voto electrónico; Congreso Argentino de Ciencias de la Computación, Octubre 2005.

Pesado P., De Giusti A., Pasini A., Estrebou C.; Voto Informatizado en la Facultad de Informática UNLP; Reporte Técnico III-LIDI. Facultad de Informática UNLP. Diciembre, 2007

Rosenberg M.; E-LEARNING: Extrategias para transmitir conocimientos en la era digital; Colombia, MC Hill, 2001

Rumble G.; The costs and costing of networked learning; Publicado en Journal of Asynchronous Learning Networks [serial online], Septiembre de 2001, Volume 5, Issue 2

Tula, Maria (comp.); Voto Electrónico; Ariel Ciencias Políticas, 2005.

Wang L., Bretschneider S., Gant J.; 2005; Evaluating Web-based e-government services with a citizen-centric approach Proceedings of the 38th Hawaii

International Conference on System Sciences - 2005, [cited June 4 2006]

Gestión de la degradación de las comunicaciones en sistemas de vídeo bajo demanda a gran escala en Internet

Eduardo Grosclaude, Claudio Zanellato, Javier Balladini
Departamento de Ciencias de la Computación
Universidad Nacional del Comahue, Argentina
{oso, czanella, jballadi}@uncoma.edu.ar

Remo Suppi
Departamento de Arquitectura de
Computadores y Sistemas Operativos
Universidad Autónoma de Barcelona, España
remo.suppi@uab.es

Contexto

Dentro del marco del proyecto “Software para Aprendizaje y Trabajo Colaborativo” perteneciente a la Universidad Nacional del Comahue, se ha definido una línea de investigación cuyo objetivo es diseñar aplicaciones educativas de Vídeo bajo Demanda (VoD, *Video on Demand*) para dar soporte a los procesos de enseñanza y aprendizaje colaborativo.

Resumen

Los sistemas de Vídeo bajo Demanda (VoD) ofrecen servicios interactivos y personalizados de visualización de contenidos multimedia. Al localizarse estos servicios sobre redes con concepto *best-effort*, como Internet, se presentan problemas de servicio asociados a la congestión y pérdida de comunicación. En este trabajo se presenta una línea de investigación en sistemas de VoD a gran escala, con énfasis en las propiedades de Calidad de Servicio y Tolerancia a Fallos. Se describen los módulos de Transmisión de *Streams* y de Planificación de Canales Lógicos, ya desarrollados; y el módulo de Garantía de Calidad de Servicio, actualmente en diseño.

1. Introducción

Los sistemas de vídeo bajo demanda a gran escala (LVoD, *Large-Scale Video-on-Demand*) brindan un servicio de visualización de vídeos a una gran cantidad de usuarios dispersos geográficamente, permitiendo numerosas opciones, dentro de las que se destaca la interactividad y personalizado de la programación ofrecida a los usuarios (servicio conocido como vídeo bajo demanda verdadero —T-VoD, *True Video-on-Demand*—). Esto significa que es posible elegir, a cualquier hora, la película que queremos ver como si fuera reproducida en nuestro dispositivo de vídeo particular, interactuando con el emisor enviándole nuestras preferencias de visualización (pausar, reanudar, avanzar rápidamente, etc).

La implementación de sistemas de VoD sobre Internet es una tendencia evidente; sin embargo, hasta la actualidad la mayoría de los sistemas de VoD han sido diseñados para trabajar en redes locales y dedicadas, o que permiten hacer reserva de recursos. Cuando el entorno de red pasa de las redes locales a una red como Internet, aumenta la probabilidad de fallos, disminuye el ancho de banda, y se pierde la calidad y clasificación de servicios la que es reemplazada por un modelo de servicio de “mejor esfuerzo”. El tráfico presenta fluctuaciones originadas por estados de congestión.

La congestión puede producirse ya sea por un aumento del tráfico o por una falla física como la caída de enlaces o encaminadores. Cuando un enlace se cae, los algoritmos de encaminamiento derivan el tráfico hacia otros enlaces, pudiendo producir congestión en los enlaces que han recibido la derivación del tráfico. Un interesante trabajo [1], muestra el impacto de la falla de los enlaces en sistemas de voz sobre IP (VoIP). En él se encontró que, a pesar de la protección del encaminamiento IP, las caídas de enlaces y encaminadores fueron seguidas por largos períodos de inestabilidad en el encaminamiento de paquetes, produciendo múltiples descartes de los mismos a causa de reenvíos por caminos incorrectos. Por lo tanto, no sólo hay fluctuaciones en el ancho de banda de las comunicaciones, sino que también es posible que por problemas de encaminamiento, un destino se vuelva inaccesible.

La presente investigación propone una nueva arquitectura de un sistema de LVoD distribuido, que permita ofrecer un servicio de T-VoD, con comunicaciones unitransmisión sobre una red sin calidad de servicio como Internet. Este sistema, denominado VoD-NFR (*Video-on-Demand with Network Fault Recovery*), tiene como fin garantizar, ante fallos de la red y caídas de servidores, la entrega del contenido multimedia a los clientes por una red como Internet, sin degradar su calidad y sin sufrir interrupciones durante su visualización.

Dadas las deficiencias que presenta la red de comunicaciones analizada, un sistema de LVoD con QoS (*Quality of Service*) debe ser capaz de: 1. Detectar estados de congestión y ajustar la tasa de transmisión para darles solución. 2. Determinar el ancho de banda de comunicación con los clientes. 3. Planificar la entrega del contenido multimedia de acuerdo al ancho de banda disponible. 4. Detectar inconvenientes en la comunicación entre servidores y clientes, y con la debida antelación, tomar medidas que permitan continuar prestando un servicio de calidad. Estas cuatro responsabilidades que deben cubrir los sistemas de LVoD, pueden ser enmarcadas en tres componentes. Las dos primeras responsabilidades son propias de un *módulo transmisor de streams* (STM, *Stream Transmission Module*), mientras que la tercera y cuarta responsabilidad son tareas de un *planificador de canales lógicos* (LCS, *Logical Channel Scheduler*) y un *módulo de garantía de la calidad del servicio* (QoSA, *Quality of Service Assurance*), respectivamente.

Los componentes STM y LCS ya han sido diseñados, implementados, y parcialmente publicados, mientras que el restante, el módulo QoSA, actualmente se encuentra en proceso de desarrollo. En este trabajo se plantea la necesidad del módulo QoSA y sus desafíos, se describen las generalidades de los módulos STM y LCS, y se explica la interacción entre estos tres componentes claves para proveer gestión de la degradación de las comunicaciones al sistema VoD-NFR. Una vez construido el módulo QoSA, nuestro sistema intentaría representar un modelo de las relaciones generales de componentes de un sistema de LVoD adaptable a la variabilidad de las condiciones de la red, posiblemente implantado en otras arquitecturas y modelos de servicio, desde esquemas centralizados, hasta distribuidos, P2P y P4P.

1.1. Trabajos relacionados

Las soluciones actuales de sistemas de LVoD se centran fundamentalmente en el alto nivel de gestión, es decir, en los modelos de servicio, y prácticamente sin considerar que el modelo de servicio no funcionará de manera adecuada en una red con alta probabilidad de fallos. Pocos son los trabajos orientados a la tolerancia a fallos de la red, no obstante, hay mucha investigación con respecto a la tolerancia a fallos de servidores, entre los que se puede mencionar a [2].

Varios trabajos (entre ellos [3]) dan soluciones adecuadas para *streaming* pero inadecuadas

para VoD y, además, la mayoría no describe el mecanismo de detección de fallos de red. En [4] se presenta un mecanismo de detección de fallos de red que dice ser para VoD aunque en realidad está orientado a *streaming* en vivo. Este esquema no funciona para T-VoD, debido a que los contenidos no necesariamente son transmitidos a la velocidad de reproducción, por lo tanto el cliente no dispone de suficiente información para determinar cuándo hay un fallo de red o cuándo el servidor ha decidido suspender momentáneamente su servicio mientras atiende a otros clientes de mayor urgencia. Algo similar ocurre con [5], donde se utiliza un mecanismo de *heartbeat* para proveer tolerancia a fallos de la red en un sistema de VoD. Mediante este esquema, se puede detectar la pérdida total de comunicación y el incremento del retraso de los paquetes pero, sin embargo, no es posible obtener suficiente información de la red para tomar medidas apropiadas.

Pueden encontrarse trabajos relacionados con los módulos STM y LCS en publicaciones que nuestro grupo ha realizado en [6] y [7], respectivamente.

2. Líneas de Investigación y Desarrollo

El sistema VoD-NFR consiste de tres componentes que conforman la arquitectura base del sistema: el STM (*Stream Transmission Module*), el LCS (*Logical Channel Scheduler*) y el módulo QoS (Quality of Service Assurance). Estos tres componentes son descriptos a continuación.

2.1. Módulo transmisor de *streams*

El STM es uno de los módulos del sistema VoD-NFR más complejo de diseñar e implementar, y es el componente base que permite al resto de los componentes hacer su trabajo correctamente. Su objetivo es proporcionar un transporte de datos de los vídeos del servidor a los clientes, con soporte de servicio T-VoD, y también de proveer información del estado de las comunicaciones.

El STM incluye un Planificador del Tráfico de Red (NTS, *Network Traffic Scheduler*) para cumplir con parte de sus responsabilidades. El NTS se encarga de las transmisiones de datos por la red y de generar información sobre el estado de las comunicaciones con cada uno de los clientes. Para esto, el NTS utiliza un protocolo de control de congestión TCP-Friendly, denominado ERAP, especialmente diseñado para VoD-NFR y presentado en [6]. ERAP extrema el ahorro de recursos para soportar la elevada carga de trabajo típica de servidores de VoD. El STM aumenta la funcionalidad del NTS, mediante un protocolo de capa superior denominado STP (*Stream Transmission Protocol*) que se ocupa de mantener el origen y la sección de contenidos enviados, para que el cliente pueda clasificar los datos recibidos.

2.2. Planificador de canales lógicos

El LCS es el encargado de decidir qué canal lógico es servido en cada momento para garantizar la QoS, y de adaptarse dinámicamente al ancho de banda disponible de las comunicaciones. Es decir, es el responsable de distribuir el ancho de banda del servidor entre los distintos canales, respetando las necesidades de contenido multimedia de los clientes asociados a esos canales. El LCS desarrollado para este sistema soporta vídeos VBR (*Variable Bit Rate*) y es del tipo *anticipación*. Esta última propiedad permite, al sistema, adaptarse a las fluctuaciones del ancho de banda de las comunicaciones, evitando degradar la calidad de los vídeos cuando el ancho de banda es menor al

requerido. Las soluciones de *anticipación* son aplicables a vídeos prealmacenados (en el servidor) o que el cliente recibe con cierto retraso desde el momento de la generación de la señal de vídeo. El LCS define un mecanismo que prioriza la atención de ciertos clientes sobre otros. En [7] hemos publicado un esquema de priorización basado en las necesidades de contenido multimedia de los clientes. Teniendo en cuenta también los estados de reproducción de cada cliente, con posterioridad hemos mejorado el esquema, el cual se encuentra en proceso de publicación.

2.3. Módulo de garantía de QoS

El módulo QoS tiene tres responsabilidades fundamentales: aceptar o rechazar solicitudes de vídeos de acuerdo a la disponibilidad de recursos del servidor, balancear la carga del sistema, y detectar los fallos de comunicación con el cliente para luego tomar las medidas apropiadas que permitan garantizar la QoS de la entrega de los vídeos.

Las dos primeras responsabilidades son normalmente encontradas en cualquier sistema de VoD. La aceptación de una petición de un nuevo cliente, significa que ha sido aceptado por el balanceador de carga y también que el servidor tiene recursos disponibles para prestarle servicio. Este componente debe garantizar que la aceptación de la petición no compromete el buen funcionamiento del servidor, lo que significa que la QoS del contenido multimedia entregado a los clientes que ya habían sido aceptados para ser servidos, no será comprometida.

La tercer responsabilidad es el problema al que apunta nuestra línea de investigación. VoD-NFR ha sido diseñado desde un principio para soportar la detección de los fallos de red. El módulo QoS dispone de información actualizada del estado de las comunicaciones que obtiene del STM, así como también del estado de reproducción del cliente, provisto por el LCS. Con esta información, el componente QoS, estaría en condiciones de decidir si una determinada comunicación con un cliente es suficiente o no para continuar con la prestación del servicio. Además de determinar cuál es la mejor forma de aprovechar esta información para detectar los fallos de red, es necesario diseñar un esquema de recuperación del fallo. Una vez determinado el proceso de detección del fallo de red y el de recuperación, el módulo QoS sería capaz de cumplir con la tercera responsabilidad asignada.

3. Resultados Obtenidos/Esperados

Con el objetivo de evaluar el comportamiento del nuevo sistema VoD-NFR, se ha diseñado y desarrollado un prototipo que funciona en entorno real y de simulación (N2 [8]).

El estudio experimental se ha dividido en tres fases incrementales, comenzando por un único componente hasta valorar el sistema completo. En la primera fase, se evaluó el módulo STM mediante pruebas reales y de simulación, donde se midieron parámetros de rendimiento y consumo de recursos bajo una alta carga de trabajo, típica de servidores de VoD. En la segunda fase, se analizó el LCS y su integración con el STM que demuestran la adaptabilidad del sistema a las fluctuaciones del ancho de banda de la red. En la última fase, tenemos previsto desarrollar pruebas del sistema completo para verificar el funcionamiento del mecanismo de detección y recuperación de fallos del módulo QoS a construir.

Los resultados obtenidos con el STM y el LCS han sido plenamente satisfactorios, y pueden observarse en [6] y [7], respectivamente. Actualmente, nos hemos planteado estudiar el compor-

tamiento del flujo de tráfico en la red de comunicaciones de la Universidad Nacional del Comahue y, de acuerdo a esta experimentación, diseñar un esquema de detección y recuperación de fallos a implantar en el módulo QoS, que permita continuar prestando servicio a la mayor cantidad posible de clientes afectados por problemas temporales de la red.

4. Formación de Recursos Humanos

Parte de los estudios del sistema VoD-NFR han sido realizados en la Universidad Autónoma de Barcelona, España, dentro del proyecto “Procesamiento de Altas Prestaciones: Arquitectura, Entornos de Desarrollo y Aplicaciones” (TIN-2004-03388), que han dado lugar a la tesis doctoral de Javier Balladini, quien actualmente es miembro del grupo de investigación que trabaja en la línea presentada en este trabajo. Nuestro grupo de investigación está orientado a continuar desarrollando y adaptar el sistema de acuerdo a las necesidades planteadas por el proyecto “Software para Aprendizaje y Trabajo Colaborativo” (E04/065) de la Universidad Nacional del Comahue, Argentina.

Referencias

- [1] Catherine Boutremans, Gianluca Iannaccone, and Christophe Diot. Impact of link failures on voip performance. In *NOSSDAV '02: Proceedings of the 12th international workshop on Network and operating systems support for digital audio and video*, pages 63–71, New York, NY, USA, 2002. ACM.
- [2] Steven Berson, Leana Golubchik, and Richard R. Muntz. Fault tolerant design of multimedia servers. *SIGMOD Rec.*, 24(2):364–375, 1995.
- [3] Tai T. Do, Kien A. Hua, and Mounir A. Tantaoui. Robust video-on-demand streaming in peer-to-peer environments. *Comput. Commun.*, 31(3):506–519, 2008.
- [4] T. Anker, D. Dolev, and I. Keidar. Fault tolerant video on demand services. *Proceedings of 19th IEEE International Conference on Distributed Computing Systems*, pages 244–252, 1999.
- [5] A. Maharana and G.N. Rathna. Fault-tolerant video on demand in rserpool architecture. *International Conference on Advanced Computing and Communications (ADCOM)*, pages 534–539, 20–23 Dec. 2006.
- [6] Javier Balladini, Leandro Souza, and Remo Suppi. A network scheduler for an adaptive VoD server. In *E-Business and Telecommunication Networks, Communications in Computer and Information Science*, volume 9, pages 237–251. Springer-Verlag Berlin Heidelberg, 2008.
- [7] Javier Balladini, Leandro Souza, and Remo Suppi. Un planificador de canales lógicos para un servidor de VoD en internet. *XII Congreso Argentino de Ciencias de la Computación (CACIC 2006)*, 2006.
- [8] Steven McCanne and Sally Floyd. NS2 - Network Simulator. <http://www.isi.edu/nsnam/ns/>, 2009.

Programación Paralela en Sistemas Híbridos

Verónica Gil-Costa, Fernando Saez, Cristian Tissera y Marcela Printista
LIDIC, Departamento de Informática
Facultad de Ciencias Físico, Matemáticas y Naturales
Universidad Nacional de San Luis

CONTEXTO

Esta línea de investigación pertenece al Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de la UNSL, el cual tiene por objetivo actual el diseño e implementación de nuevas herramientas de computación inteligente para desarrollar Sistemas Inteligentes. También se incluyen entre las áreas de investigación del LIDIC los Sistemas Multiagente, los Sistemas Híbridos y las Técnicas de Distribución y Paralelismo para acelerar los procesos de búsqueda y mejorar la calidad de los resultados.

RESUMEN

Con la aparición de las CPU multi-cores (o Chip-level-Multi-Processor -CMP-), es importante el desarrollo de las técnicas que exploten las ventajas de las CMP para acelerar las aplicaciones paralelas que poseen una gran demanda de cómputo paralelo. En particular, las aplicaciones que requieren de un gran poder computacional de los recursos disponibles, es esencial poder desarrollar estrategias y algoritmos que aprovechen el uso adecuado del hardware. Esto es especialmente crítico cuando se consideran sistemas o aplicaciones en las que los requerimientos ingresan en intervalos variables. En este trabajo se propone el desarrollo de técnicas híbridas basadas en el uso de MPI para la comunicación entre procesadores y OpenMP para la comunicación entre cores de un mismo procesador. OpenMP ha sido desarrollado para tomar ventaja de las facilidades multithreading de los nodos CMP.

Palabras clave: *sistemas multi-core, OpenMP, MPI, simulación paralela, algoritmos paralelos, clusters*

1. INTRODUCCION

Modificar un programa secuencial para distribuirlo o para que se ejecute en paralelo tiene dos atractivas razones: acelerar la velocidad de ejecución y/o mejorar la cantidad de memoria disponible [Hol07]. Actualmente existen dos alternativas para paralelizar aplicaciones secuenciales: (a) utilizar clusters de computadoras, y (b) utilizar sistemas multi-cores. Los sistemas multi-core difieren en varios aspectos respecto de un cluster de computadoras. Un sistema multi-core ofrece a todos los cores un acceso rápido a una única memoria compartida, evitando la transferencia de datos entre las máquinas a través de la red del cluster, pero aún así la cantidad de

memoria disponible es limitada. En cambio los clusters de computadoras permiten incrementar el espacio de almacenamiento (principal y secundario) aunque deben pagar el precio de la latencia de la red. Una tercera alternativa (c) consiste en un sistema híbrido que permite combinar las características de ambos sistemas, e incrementar aun más la capacidad y el poder de los sistemas computacionales. Esta última alternativa es el eje central de la línea de investigación propuesta en este trabajo. Sin embargo el diseño de aplicaciones para este tipo de sistema trae aparejado una serie de complejidades que se describen en esta sección.

La tendencia actual de la tecnología destaca la aparición de procesadores CMP [Str69, Hamm97, Hamm00]. En realidad, la mayoría de los sistemas que incorporan estos chips descartan la vieja idea de un sistema multiprocesador como varios nodos mono-procesador. Las tendencias de la tecnología indican que el número de cores (núcleos) en un chip seguirá creciendo a medida que lo indiquen los planes de trabajo de los fabricantes más importantes. Hoy en día, AMD trabaja con chips de cuatro núcleos (Quad tecnología nativa) e Intel ha comenzado a incorporar el procesador Intel Core™ de cuatro y ocho núcleos de procesador en sus sistemas.

Un procesador multi-core combina dos o más núcleos (normalmente una CPU) en un único paquete compuesto de un sólo circuito integrado (CI) denominado *die*. Un procesador dual-core contiene dos núcleos, y un procesador quad-core posee cuatro núcleos. Un microprocesador multi-core implementa multiprocesamiento en un solo paquete físico. Un procesador con todos los núcleos en una sola *die* se denomina procesador monolítico. Los cores pueden compartir una única caché coherente del nivel más alto del dispositivo de memoria caché (por ejemplo, L2 para el Intel Core 2) o puede tener cachés separadas (por ejemplo, los procesadores AMD dual-core). Los procesadores también comparten la misma interconexión con el resto del sistema. Cada core implementa independientemente optimizaciones como ejecución superescalar, pipelining, y multithreading. Un sistema con n cores es efectivo cuando se presenta con n o más hilos simultáneamente. La ganancia obtenida por el uso de sistemas multi-cores, depende de las características del problema que se quiere

resolver y del algoritmo utilizado. La Figura 1 muestra el diseño de un chip dual core.

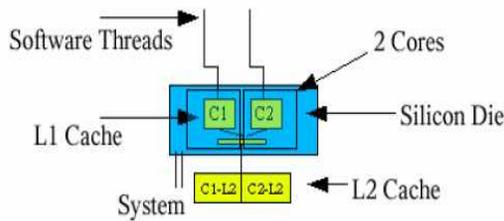


Figura 1: Procesador con CMP.

Las mejoras que se obtienen al utilizar sistemas multi-cores se reflejan a través del paralelismo desarrollado para las aplicaciones. Si se ejecuta un programa secuencial sobre un sistema multi-core no se obtiene ninguna ganancia. Por lo tanto, la ventaja de estos sistemas consiste en dividir eficientemente la ejecución de una aplicación, para que cada parte sea ejecutada por un thread independiente. La cantidad de threads disponibles en estos sistemas también es un tema importante a analizar, debido a que la creación y administración de threads requiere del uso de los recursos tales como la memoria; además los threads deben ser cuidadosamente planificados e incorporados en el stack de ejecución.

Por lo tanto, el uso de chips multi-core en los sistemas de computación de uso común conllevará, sin lugar a dudas, grandes cambios en la forma de desarrollar aplicaciones y software. En esta nueva etapa, será necesario reescribir los programas existentes, paralelizando el código para sacar provecho de los sistemas con múltiples núcleos. Una de las cuestiones básicas cuando se diseñan programas en paralelo es coordinar los accesos a los datos compartidos de manera que los resultados de la ejecución sean correctos. El uso de mecanismos de coordinación entre procesadores basados en el acceso exclusivo a datos (uso de exclusiones mutuas de grano fino) no se percibe como posibles soluciones por la complejidad que supone su uso para la mayoría de los programadores. Las zonas críticas no sólo son complejas de administrar, sino que son posibles fuentes de cuellos de botella cuando varios threads deben acceder a ellas para realizar una operación.

Los trabajos realizados hasta el momento en sistemas multi-core, se focalizan en la eficiente paralelización de un conjunto de instrucciones de un programa secuencial [Fri06, Ben05, Cho07, Goo07], para ellos proponen y evalúan dos métodos principales denominados parallel depth-first (pdf) y work-stealing (ws). También existen algunas variantes que optimizan estas estrategias [Gu]. Sin embargo, estos trabajos no aprovechan las características inherentes en las aplicaciones que se desean paralelizar. Es decir que ninguno de estos

trabajos toma ventaja de las características particulares de las aplicaciones que se desean paralelizar. Es importante estudiar la forma en particular que trabajan las aplicaciones, debido que para ciertas operaciones es conveniente utilizar uno u otro tipo de planificación paralela.

Todavía no hay estudios con esta tecnología que evalúen los paradigmas de programación paralela que trabajan en el contexto de las aplicaciones que resuelven requerimientos de los usuarios en un corto plazo y que requieren de un gran poder computacional, y si son las más apropiadas.

Otro aspecto que no ha sido estudiado hasta el momento en sistemas multi-core, es el efecto que tienen los cambios de tráfico de los requerimientos ingresados al sistema. Si el procesador posee varios cores que pueden trabajar independientemente pero compartiendo los datos de memoria, cuando los requerimientos de los usuarios ingresan en intervalos cortos o con una gran afluencia en el sistema, es conveniente utilizar threads que trabajen en forma independiente sobre cada requerimiento. Sin embargo, cuando los requerimientos ingresan entre intervalos de tiempo lo suficientemente grandes, es conveniente que los requerimientos se dividan en tareas atómicas y que los threads que se encuentran ociosos “roben” tareas de los threads que tienen una carga de trabajo elevada. De esta manera es posible optimizar el uso de los recursos disponibles en el sistema.

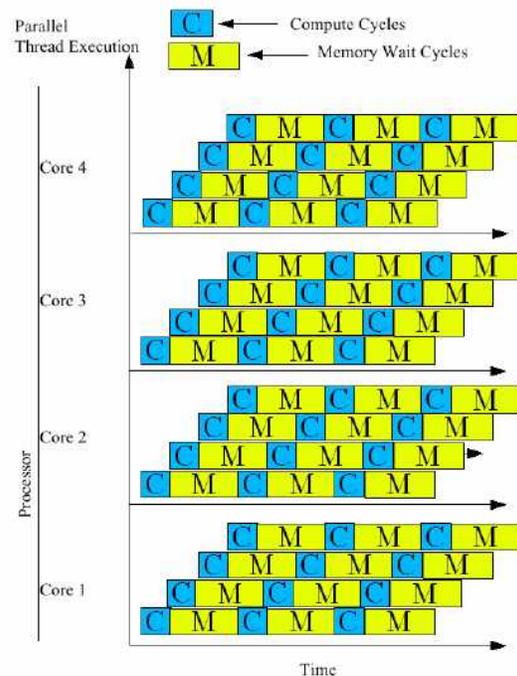


Figura 2: Con CMP se incrementa la velocidad de procesamiento sin modificar la velocidad de acceso a la memoria.

Este mismo sistema híbrido de comunicación puede ser utilizado a nivel de procesadores del cluster de la siguiente manera. Cuando el tráfico de los requerimientos es alto, conviene administrar y procesar estos requerimientos en masa (*bulk*). De esta manera se logra reducir los costos (latencia) de enviar mensajes por la red de intercomunicación y por lo tanto mejorar los tiempos de respuestas. Por otro lado, cuando los tiempos entre arribos de los requerimientos son lo suficientemente grande (es decir que llegan pocos requerimientos al sistema), es conveniente utilizar un modo de comunicación completamente asíncrono, donde cada procesador resuelve sus tareas independientemente [Mar08a, Mar08b].

Las técnicas de planificación tanto a nivel micro (dentro de cada procesador) y macro (en un cluster de computadoras) deben ser capaces de obtener un buen balance de carga. Si los todos los cores/procesadores logran realizar aproximadamente la misma cantidad de operaciones, esto permite mejorar el tiempo de respuesta a los requerimientos, así como incrementar el throughput. Pero para ello es esencial un uso equitativo de los recursos por parte de todos los requerimientos. Para lograr una planificación adecuada de las operaciones a realizar, es posible utilizar un *scheduler* que sea capaz de predecir el comportamiento de los cores/recursos dependiendo del tipo de requerimiento que se esté procesando. Si el sistema opera en modo síncrono la tarea del scheduler es sencilla y puede predecir la cantidad de tareas que cada core/procesador debe realizar en cada etapa realizando mediciones y operando cada N_q consultas terminadas. Esto se puede llevar a cabo utilizando un modelo de colas $G/G/\infty$, donde el cociente de la tasa de llegada de eventos y la tasa de servicios (λ/μ) determina el número de servidores activos. La tasa de llegada se define como $\lambda = n/\Delta$, durante un período de tiempo Δ en el que se procesaron n requerimientos. La tasa de servicio se define como $\mu = n/S$, donde S es la suma de las diferencias de los tiempos de salida menos los tiempos de arribo $\delta q [Tpo_Salida - Tpo_arribo]$.

Sin embargo, cuando el sistema opera en modo asíncrono, la planificación de los nuevos requerimientos tiene una complejidad adicional, debido a que en un período de tiempo determinado los cores/procesadores pueden tener un número diferente de requerimientos en espera o en servicio. Además, no existe ningún tipo de control que permite visualizar el estado del sistema en un instante dado. En este caso es posible utilizar un sistema de planificación síncrono ficticio que permita predecir el comportamiento de del sistema real.

Teniendo en cuenta esta tendencia, el objetivo principal de este trabajo es estudiar, cuál es la forma

más eficiente para explotar esta nueva tecnología, aplicándola sobre un cluster de computadoras. De esta forma se logra obtener un sistema híbrido que combina las propiedades de “divide y vencerás” en un nivel macro (a nivel de cluster con memoria distribuida), y en un nivel micro (dentro de cada procesador con memoria compartida).

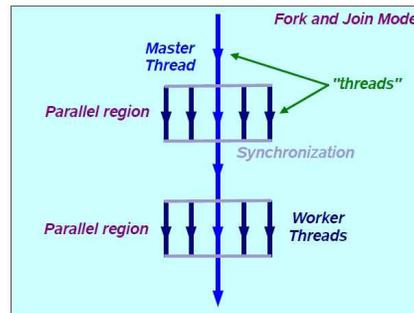


Figura 3: Programación con OpenMP.

2. LINEAS DE INVESTIGACION y DESARROLLO

La línea de investigación involucra una serie de desarrollos individuales que en su conjunto logran obtener el objetivo planteado. Para ello es necesario estudiar formas eficientes de programación paralela utilizando el lenguaje OpenMP sobre arquitecturas multi-cores. Como se mencionó anteriormente, es de suma importancia reducir las regiones críticas o zonas de memoria compartidas para evitar los cuellos de botellas. También es necesario estudiar las ventajas y desventajas provistas por los sistemas de clusters de computadora que se comunican mediante librerías paralelas como MPI (asíncrona) y/o BSP (síncrona).

Para optimizar la eficiencia de los algoritmos paralelos, es necesario determinar las características particulares de las aplicaciones. Además, es necesario utilizar estructuras de datos que sean capaces de manejar eficientemente grandes cantidades de datos (miles de Terabytes).

2.1. Línea 1: Algoritmos de Búsquedas Avanzados

Una aplicación que requiere de un gran poder computacional y que puede beneficiarse de las características provistas por los sistemas multicores son las aplicaciones que soportan búsquedas de objetos multimediales. Estas deben ser capaces de resolver requerimientos en un tiempo razonable y al mismo tiempo dependen de operaciones computacionales complejas generalmente de alta dimensionalidad. Un ejemplo de estos sistemas es el sitio Flickr (<http://www.flickr.com/>) que utiliza estructuras de datos poco convencionales basadas en

árboles y mapas. Este tipo de estructuras es poco conveniente al momento de utilizar un sistema multi-core, debido a los accesos simultáneos a celdas de memoria. Por lo tanto, es necesario diseñar cuidadosamente las estructuras de almacenamiento y de indexación [GilC08a, GilC08b].

2.2. Línea 2: Programación Paralela Estructurada

Esta línea trata con nuevas formas de programación paralela que tienen por objetivo reducir el tiempo de diseño, testing y codificación de aplicaciones paralelas. En esta dirección estamos desarrollando un sistema de programación esquelético que simplifique la tarea de desarrollar software paralelo. Con el advenimiento de las arquitecturas actuales (cluster con nodos multicore) es necesario la adaptación de estos sistemas para obtener las ventajas del modelo híbrido (Pasaje Mensajes y de Memoria compartida). Además, esto implica abordar nuevos problemas que surgen en este entorno, como la sincronización, planificación, mapeo de procesos sobre procesadores, balance de carga, etc [Sae08a, Sae08b].

2.3 Línea 3: Simulación en Ciencia Computacional

En esta línea se aborda la simulación de alta performance. El propósito es realizar una aplicación paralela utilizando un sistema híbrido basado en OpenMP y MPI, para desarrollar un modelo de simulación basado en autómatas celulares (AC) y agentes inteligentes, destinado al estudio de las dinámicas pedestres en situaciones de emergencia, especialmente aquellas que involucran la evacuación forzada debido a la amenaza del fuego dentro de un área definida cerrada con un número específico de salidas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados obtenidos hasta el momento son:

- Diseño e implementación de algoritmos eficientes que permiten resolver consultas ingresadas por los usuarios en un sistema de búsqueda. El sistema ha sido implementado para soportar consultas de texto y multimediales sobre un cluster de computadoras [GilC08c].
- Los sistemas de búsquedas desarrollados operan en forma síncrona y asíncrona [Mar08a].
- Se han implementado estrategias que permiten resolver eficientemente los requerimientos de los usuarios en sistemas multi-core y que realizan una planificación de las tareas para optimizar el uso de los recursos.
- El desarrollo de esqueletos paralelos que resuelven algoritmos del tipo “Divide y Vencerás” y “Autómata Celular” para el modelo de pasaje de mensajes [Sae07].

- Un modelo teórico para predecir el tiempo de ejecución de una aplicación que se instancia en el esqueleto “Divide y Vencerás” [Sae08b].
- Un modelo secuencial de la simulación basada en autómatas celulares para el estudio de procesos de evacuación [Tis06a, Tis06b].

Los resultados esperados son:

- Combinar diferentes modos de comunicación (síncrono/asíncrono) sobre el sistema que opera sobre un cluster de computadoras.
- Obtener un sistema híbrido que permite utilizar los beneficios provistos por los sistemas multi-core y los clusters de computadoras. Para ello es posible utilizar programas que utilicen tanto la librería OpenMp como MPI/BSP.
- Aplicar algoritmos de scheduling inteligentes que se adapten a los cambios del ambiente (tráfico de requerimientos variable) para balancear la carga de trabajo a nivel macro y micro.
- Aprovechar las características de las estructuras de datos utilizadas como índices, para explotar la localidad de los datos y reducir los costos de comunicación (latencia), y los fallos a memoria caché.
- Adaptar otros esqueletos algorítmicos a los sistemas de alta performance actuales (híbridos).

4. FORMACION DE RECURSOS HUMANOS

Los investigadores de esta línea de trabajo se desarrollan en el Laboratorio de Investigación y desarrollo en Inteligencia Computacional (LIDIC). Actualmente, el mismo cuenta con un subsidio de Ciencia y Técnica de la UNSL y un Proyecto PICT del FONCYT.

En la temática de la línea se están desarrollando tres doctorados en la UNSL:

- 1- El doctorando es becario de CONICET y el trabajo se ha realizado en cooperación con investigadores del Centro de Investigación de la Web-Chile (Próximo a terminar).
- 2- El doctorando es becario de CONICET. (En desarrollo, plan de tesis aprobado).
- 3- El doctorando es docente exclusivo de la UNSL y el trabajo se desarrolla en cooperación con la Universidad Autónoma de Barcelona. (En desarrollo, se está confeccionando su plan de tesis).

Subsidios externos:

Durante el año 2007 y 2008 estos investigadores han sido parcialmente subsidiados por el proyecto CYTED GRID (finalizado) para realizar viajes de

estudios tendientes a completar su formación de posgrado.

Durante el 2008, el Centro Yahoo! Research, becó en 6 oportunidades a una investigadora para realizar investigaciones conjuntas.

Durante el 2008, una investigadora fue becada por el centro de investigación High Performance Computing Center Stuttgart (HLRS) de la Universidad de Stuttgart, Alemania. Financiado por el programa HPC-Europa Transnational Access para realizar una estadía de perfeccionamiento de 2 meses.

5. BIBLIOGRAFIA

[Ben05] Concurrent cache-oblivious B-trees. M. A. Bender, J. T. Fineman, S. Gilbert, and B. C. Kuszmaul. In ACM SPAA, 2005.

[Cho07] The cache-oblivious gaussian elimination paradigm: Theoretical framework, parallelization and experimental evaluation. R. Chowdhury and V. Ramachandran. In ACM SPAA, 2007.

[Fri06] The cache complexity of multithreaded cache oblivious algorithms. M. Frigo and V. Strumpfen. In ACM SPAA, 2006.

[GilC08a] Parallel query processing on distributed clustering indexes. V. Gil-Costa, M. Marin and N. Reyes Journal of Discrete Algorithms (7) 03-17, 2009 (Elsevier).

[GilC08b] An Empirical Evaluation of a Distributed Clustering-Based Index for Metric Space Databases, G.V. Costa, M. Marin and N. Reyes. In International Workshop on Similarity Search and Applications (SISAP 2008), Cancun, Mexico, April 11-12, pp. 386-393, IEEE-CS Press, 2008

[GilC08c] Distributed Sparse Spatial Selection Indexes, G.V. Costa and M. Marin. In 16th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP 2008), Toulouse, France, Feb. 13-15, pp. 440-444, IEEE-CS Press, 2008

[Goo07] Sorting in parallel external-memory multicores. M. T. Goodrich, M. Nelson, and N. Sitchinava Technical report, U.C. Irvine, 2007.

[Gu] Provably Good Multicore Cache Performance for Divide-and-Conquer Algorithms. Guy E., Blleloch, Rezaul A. Chowdhury, Phillip B. Gibbons, Vijaya Ramachandran, Shimin Chen, and Michael Kozuch.

[Ham00] L. Hammond, B. A. Hubbert, M. Siu, M. K. Prabhu, M. Chen, and K. Olukotun. The Stanford

Hydra CMP. IEEE Micro, 20(2), 2000.

[Hamm97] L. Hammond, B. Nayfeh, and K. Olukotun. A single-chip multiprocessor. IEEE Computer, 30(9), 1997.

[Hol07] The Design of a Multi-Core Extension of the SPIN Model Checker. Gerard J. Holzmann and Dragan Bošnački. PDMC 2007, J. Electronic Notes in Theoretical Computer Science, pp. 33-46, 2007.

[Mar08a] M. Marin, G.V. Costa, (SynclAsync)+ MPI Search Engines, In Euro PVM/MPI: Recent Advances in Parallel Virtual Machine and Message Passing Interface, Paris, France, Sep. 30-Oct. 3, Lecture Notes in Computer Science 4757, pp. 117-124, Springer, 2007

[Mar08b] High-Performance Distributed Inverted Files, M. Marin, G.V. Costa. In ACM 16th Conference on Information and Knowledge Management (CIKM 2007), Lisbon, Portugal, Nov 6-9, pp. 935-938, ACM, 2007.

[Sae07] Conceptos Fundamentales de Diseño en Sistemas de programación Esqueletal. XIII Congreso Argentino de Ciencias de la Computación (CACIC 2007). Universidad Nacional del Nordeste, Corrientes y Resistencia, Argentina, Octubre 2007.

[Sae08a] Predicting the Performance of Hypercube Divide and Conquer Skeleton. F. Saez, M. Printista. Jornadas Chilenas de computación. Punta Arenas, Chile. Noviembre de 2008. ISBN 978-956-319-507-1

[Sae08b] A Performance Model for Divide and Conquer Skeletons on Homogeneous Cluster". F. Saez, M. Printista. 7th International Information and Telecommunication Technologies Symposium (I2TS'08). Brasil. Págs. 139-147.

[Str69] V. Strassen. Gaussian elimination is not optimal. Numerische Mathematik, 13(4), 1969. Springer.

[Tis06a] Tesis de Licenciatura en Cs. de la Computación: "Simulación basada en autómatas celulares para el estudio de procesos de evacuación". Director: Marcela Printista, Co-director: Dr. Marcelo Errecalde. Resolución 490/06.

[Tis06b] Tissera C., Errecalde M. Printista A.M (2006). Simulación De Evacuaciones basada en Autómatas Celulares. (XII CACIC 2006). Págs. 1837 a 1848. ISBN 950-609-050-5. Argentina.

Diseño y clasificación de los Sistemas de Descubrimiento de recursos Grid

Martín Chuburu*

Javier Echaiz

Jorge Ardenghi

Laboratorio de Investigación de Sistemas Distribuidos (LISiDi)

Departamento de Ciencias e Ingeniería de la Computación

Teléfono: +54 291 4595135, Fax: +54 291 4595136

Universidad Nacional del Sur, Bahía Blanca (8000), Argentina

{mic,je,jrap}@cs.uns.edu.ar

Resumen

En la actualidad, contamos con grandes redes de recursos computacionales interconectadas a través de Internet. La *computación grid* provee una plataforma a través de la cuál se puede acceder a la mayoría de estos recursos haciendo abstracción de su localización. Para lograr esto, es necesaria la implementación de *sistemas de descubrimiento* (Discovery) que permitan localizar los recursos necesarios para llevar a cabo una tarea dentro del entorno grid.

En este artículo se exploran los aspectos a tener en cuenta a la hora de diseñar un sistema de descubrimiento de recursos, así como también se observan como estos aspectos fueron considerados en el diseño de los sistemas existentes.

Palabras Clave: Computación Grid, Descubrimiento de recursos, aspectos de diseño.

1. Introducción

La disponibilidad de computadoras poderosas conectadas a Internet, así como también redes de alta velocidad, está cambiando la forma de hacer computación y de utilizar las computadoras en el presente [1]. También está creciendo el interés en agrupar recursos computacionales que se encuentran geográficamente distribuidos para resolver problemas de gran escala. Esta filosofía es en la que se basa el concepto de *Computación Grid*. En este contexto, una amplia variedad de recursos de procesamiento, dispositivos de visualización, instrumentos especiales de tipo científico, sistemas de almacenamiento y bases de datos son agrupados lógicamente y presentados como un único recurso integrado al usuario.

Sin embargo, a medida que el sistema crece, la variedad de recursos se hace más extensa dificultando la tarea de localizar los recursos necesarios para la ejecución de aplicaciones. Por ello es de vital importancia que el sistema cuente con mecanismos para encontrar esos recursos. El *descubrimiento de recursos* constituye la habilidad de localizar los recursos que cumplen

*Becario de la *Comisión de Investigaciones Científicas* (CIC) de la Provincia de Buenos Aires, Argentina

con un conjunto de requerimientos establecidos en una consulta. Las entidades involucradas en esta capacidad son tres: el *proveedor del recurso*, el *usuario del recurso* y el *servicio de descubrimiento*.

El *proveedor del recurso* es aquella entidad que permite el uso del recurso en cuestión. En computación grid, la administración de recursos se suele llevar a cabo a través de consorcios de individuos y/o organizaciones conocidos como *Organizaciones Virtuales* (VOs por su sigla en inglés) [2, 3]. El *usuario del recurso* es la entidad que se encuentra conectada a la red y que hace uso de los recursos compartidos. Se debe tener en cuenta que esta entidad podría ser un usuario humano así como también una aplicación que requiere de determinados recursos para llevarse a cabo. Por último, el *servicio de descubrimiento* es el servicio que devuelve la localización de los recursos que concuerdan con los requerimientos especificados en la consulta realizada por el usuario. En cuanto a este servicio, podría constituir un módulo como parte del usuario o el proveedor del recurso o ser una entidad independiente de estos.

A la hora de implementar un servicio de descubrimiento de recursos es necesario considerar ciertos aspectos relativos al entorno en el que funcionará, como por ejemplo su arquitectura subyacente, la forma en que se manejan los recursos, entre otros. Las decisiones tomadas en cada uno de estos aspectos determinan cómo debe ser diseñado el sistema de descubrimiento y constituirán las características que harán distintivo al servicio implementado con respecto al resto de los servicios existentes.

En la sección 2 se exponen estos aspectos de diseño mencionados, así como también las distintas alternativas que se pueden tomar en cada uno de los aspectos. Luego, en la sección 3 se muestra como estos aspectos son considerados en los sistemas de descubrimiento actuales. Por último, en la sección 4 se exponen las conclusiones y las posibles líneas de investigación relacionadas con este tópico.

2. Diseño del sistema de descubrimiento

Como se mencionaba en la sección 1, los aspectos de diseño establecen una división espacial de los distintos tipos de servicios de descubrimiento dependiendo de las decisiones que se hayan tomado para cada uno de estos aspectos. Estas decisiones están influenciadas por el tipo de plataforma o qué tan distribuido es el sistema, por lo que no siempre la elección de una alternativa constituye una solución que pueda ser catalogada como la mejor alternativa ya que depende del entorno y del tipo de aplicaciones que solicitarán el servicio.

En las siguientes subsecciones se mencionan algunos de los aspectos de diseño que fueron contemplados en [8] y [7], así como también las diversas alternativas para cada uno de estos aspectos.

2.1. Localización del proveedor del servicio

Una de las primeras decisiones de diseño es establecer quién va a proveer el servicio de descubrimiento. Como se mencionaba en la sección 1, un caso bastante común en los sistemas es que el servicio quede a cargo de una entidad independiente del proveedor del recurso o de la entidad que va a hacer uso del mismo. Esta entidad generalmente constituye un servidor o conjunto de servidores que se encarga de recolectar la información sobre los recursos disponibles con el fin de responder las consultas de los usuarios.

La alternativa a esto sería contar con un servicio que sea genuinamente distribuido, esto es, que el servicio de descubrimiento de recursos se encuentre distribuido entre todos los proveedores

y usuarios de recursos, sin ningún componente central que esté coordinando.

2.2. Configuración de la red solapada

Los sistemas distribuidos construyen redes solapadas a la red física sobre la que se encuentran. Esta red solapada es un grafo dirigido: los vértices representan los nodos de la red y los arcos representan el conocimiento de un nodo acerca de otro nodo. Existen dos alternativas para la construcción de esa red solapada: puede ser configurada manualmente o que se organice de manera automática.

- **Redes configuradas de forma manual:** Un administrador es responsable de proveer a cada servidor de la información sobre el resto de los componentes, y los usuarios y proveedores de recursos deben saber las direcciones físicas de estos servidores. Esto es, el administrador debe construir el grafo dirigido. Mientras que esta alternativa provee de mayor control, es más complicado para escalarlo.
- **Redes configuradas de forma automática:** Constituyen una alternativa interesante cuando las redes empiezan a hacerse grandes y las configuraciones manuales se hacen caras o no hay una autoridad central para coordinar las configuraciones. La desventaja es el incremento en el tráfico de la red con datos extras que únicamente sirven para mantener configurada la red solapada.
- **Híbridas:** También es posible utilizar las ventajas de ambos enfoques en una alternativa híbrida. Por ejemplo, algunas redes utilizan *multicast* o *broadcast* para organizarse a nivel de LAN y tienen una configuración manual para interconectar estos subgrafos a nivel de WAN.

2.3. Identificación de los recursos

Como se decía anteriormente, existe una gran variedad de recursos dentro de un grid. Por lo tanto, es necesario poder identificarlos para poder realizar consultas sobre estos y que el sistema de descubrimiento pueda encontrarlos. A la hora de especificar el nombre de un recurso, existen tres alternativas:

- **Identificadores *hash*:** Se aplica una función de dispersión (*hash*) sobre alguna propiedad (o conjunto de propiedades) del recurso predeterminada para todos los recursos del grid. Esto da por un resultado un identificador único para el recurso.
- **Texto plano:** Se identifican los recursos por una cadena de caracteres significativa. Esta alternativa permite realizar consultas más complejas sobre el nombre que no podrían ser realizadas sobre los *hashes* ya que no guardan relación sobre los mismos (salvo al momento de calcularlos).
- **Atributos:** Los recursos son descriptos por medio de un número de atributos predeterminados que toman un valor. Este mecanismo es muy utilizado en los sistemas de descubrimiento ya que facilita el mapeo entre los requerimientos de usuarios y las propiedades de los recursos disponibles.

2.4. Registración de los recursos

Antes de que se puedan realizar solicitudes sobre los recursos, es necesario generar y almacenar una referencia a los mismos en algún lugar donde estas referencias puedan ser accedidas por los usuarios. Este proceso se denomina *registración* y existen diferentes técnicas que se pueden usar para realizarla:

- **Registración local:** Solamente el proveedor del recurso mantiene contacto directo con los recursos a compartir, y el resto puede localizarlos consultando al proveedor.
- **Referencias *Hash*:** Las referencias a los recursos son almacenadas en los nodos cuyos identificadores *hash* estén más próximos al identificador *hash* del recurso en cuestión. Con este método se evita la centralización y la ubicación de las referencias se obtiene de manera intuitiva.
- **Registración en un servidor local:** Los proveedores de recursos anuncian sus recursos al servidor local independiente de los mismos.

3. Ejemplos de sistemas concretos

A continuación, se verán algunos sistemas de descubrimiento de recursos y se explorarán sus características teniendo en cuenta los aspectos de diseño analizados en la sección anterior.

En primer lugar, uno de los sistemas que se pueden destacar es el MDS (*Monitoring and Discovery System*) que constituye el módulo de descubrimiento de *Globus Toolkit* [4]. Si bien su implementación ha variado desde su primera versión, mantiene la mayoría de sus aspectos de diseño. La provisión del servicio de descubrimiento está a cargo de un módulo denominado *index* formando una entidad separada del proveedor y del usuario del recurso como se mencionaba en la sección 2.1. Este mismo módulo constituye el servidor local para la registración que era una de las alternativas mostradas en la sección 2.4. Los mensajes de registración por parte del proveedor de recurso se envían de forma periódica hacia la dirección del servidor local que haya sido especificada de forma manual (sección 2.2) por el administrador del sistema. La identificación de los recursos se hace por medio de los atributos (sección 2.3) de los mismos, lo cual facilita a la hora de localizarlos.

Otros sistemas importantes para mencionar son aquellos que hacen uso de técnicas *peer-to-peer* (P2P) para soportar el servicio de descubrimiento de recursos.

En [6] se presenta SDRD (*Super node and Dht approach to Resource Discovery*). En este modelo, el grid es dividido en VOs dentro de cada cual se utilizan DHTs (*Distributed Hash Tables*) para mejorar la eficiencia en el ruteo local. Además cada VO cuenta con uno o más super nodos que se encargan de transferir los pedidos a otros super nodos en otras VOs. Este caso, sería similar a una construcción híbrida como la planteada en la sección 2.2, ya que a nivel de VO (una red acotada, como sucede en una LAN) se utilizan tablas hash mejoradas como forma de configuración automática en la que se hace intercambio de mensajes con el fin de actualizar las tablas. Luego, a nivel de comunicación entre las VOs, se utilizan los super nodos, donde cada uno de ellos estarán configurados de forma manual para poder hacer transferencia de los pedidos al super nodo que corresponda.

Otro sistema P2P es el presentado en [5], el cuál a diferencia del anterior, posee una configuración automática en la cuál tanto los requerimientos de recursos como los nodos que poseen los mismos son identificados por referencias *hash* (como se explicaba en la sección 2.4).

4. Conclusiones y trabajos futuros

En el presente artículo se analizaron algunos aspectos a tener en cuenta a la hora de diseñar un servicio de descubrimiento de recursos y se mostró cómo estos aspectos fueron considerados en el diseño de sistemas que se utilizan en la actualidad. Si bien detalles como la forma de identificar los recursos, cómo llevar a cabo la registración de los mismos y las posteriores consultas son cuestiones de gran importancia, existen otros aspectos que deberían ser explorados en el futuro que en conjunto con los antes mencionados conformarán los ejes de un criterio no solamente destinado a diseñar nuevos sistemas de descubrimiento sino también para clasificar los sistemas existentes, como se menciona brevemente en la sección 3.

Se plantea como línea de investigación, la exploración de estas características adicionales y la posible implementación de un esquema de clasificación de los sistemas de descubrimiento de recursos.

Referencias

- [1] Rajkumar Buyya, Steve J. Chapin, and David C. DiNucci. Architectural models for resource management in the grid. In Rajkumar Buyya and Mark Baker, editors, *GRID*, volume 1971 of *Lecture Notes in Computer Science*, pages 18–35. Springer, 2000.
- [2] Karl Czajkowski, Carl Kesselman, Steven Fitzgerald, and Ian Foster. Grid information services for distributed resource sharing. *10th IEEE International Symposium on High Performance Distributed Computing*, 00:0181, 2001.
- [3] Ian Foster, Carl Kesselman, and Steven Tuecke. The anatomy of the Grid: Enabling scalable virtual organization. *The International Journal of High Performance Computing Applications*, 15(3):200–222, Fall 2001.
- [4] Ian T. Foster. Globus toolkit version 4: Software for service-oriented systems. *J. Comput. Sci. Technol.*, 21(4):513–520, 2006.
- [5] Abhishek Gupta, Divyakant Agrawal, and Amr El Abbadi. Distributed resource discovery in large scale computing systems. *saint*, 00:320–326, 2005.
- [6] Yiduo Mei, Xiaoshe Dong, Weiguo Wu, Shangyuan Guan, and Junyang Li. SDRD: A novel approach to resource discovery in grid environments. In Ming Xu, Yinwei Zhan, Jiannong Cao, and Yijun Liu, editors, *APPT*, volume 4847 of *Lecture Notes in Computer Science*, pages 301–312. Springer, 2007.
- [7] Ioan Toma, Kashif Iqbal, Dumitru Roman, Thomas Strang, Dieter Fensel, Brahmananda Sapkota, Matthew Moran, and Juan Miguel Gómez. Discovery in grid and web services environments: A survey and evaluation. *Multiagent and Grid Systems*, 3(3):341–352, 2007.
- [8] Koen Vanthournout, Geert Deconinck, and Ronnie Belmans. A taxonomy for resource discovery. *Lecture Notes in Computer Science*, 9:81–89, March 2005.

APLICACIONES DEL CÓMPUTO DE ALTAS PRESTACIONES

Nilda M. Pérez Otero – Sandra A. Mendez – Cristina V. Ayusa – Marino I. Aucapiña – Victor J. Lopez

Facultad de Ingeniería – Universidad Nacional de Jujuy (UNJu)
{nilperez, smendez, cristina_ayusa, maucapiña, vlopez}@fi.unju.edu.ar

CONTEXTO

El presente proyecto se encuadra en el campo del Procesamiento paralelo y Distribuido particularmente en el Cómputo de Altas Prestaciones y se desarrolla en la Facultad de Ingeniería de la UNJu en el marco del Programa de Investigación Cód. 08/D058.

RESUMEN

Debido al avance de las arquitecturas multiprocesador y la importancia que tiene actualmente el Cómputo de Altas Prestaciones (HPC) se vió la necesidad de proporcionar a los alumnos, docentes investigadores de la provincia de Jujuy y a la región noroeste un lugar de referencia donde se pueda aprovechar los servicios para aplicaciones que requieran una gran capacidad de procesamiento. Es así que se inicia en el año 2008 el proyecto de Investigación “APLICACIONES DEL CÓMPUTO DE ALTAS PRESTACIONES” formado por jóvenes docentes investigadores de la Facultad de Ingeniería de la UNJu con el objetivo de profundizar en el área del cómputo de altas prestaciones, incorporar HPC en la currícula de las carreras informáticas y colaborar con otros grupos de investigación que requieran HPC.

Palabras clave: *Cómputo de Altas Prestaciones, HPC, Procesamiento paralelo*

1. INTRODUCCION

Los organismos gubernamentales, militares y grandes centros de investigación que cuentan con aplicaciones científicas para la simulación de procesos naturales (previsión del tiempo, análisis de cambios climáticos, entre otros procesos), modelaje molecular, simulaciones físicas (túneles de viento, criptoanálisis), etc. emplean costosas supercomputadoras que permiten realizar los cálculos complejos requeridos por tales aplicaciones. Una supercomputadora o superordenador es una computadora con altísima capacidad de procesamiento y gran capacidad de memoria, diseñada para procesar enormes cantidades de información en poco tiempo y dedicada a una tarea específica en investigaciones científicas y militares. Las 4 tecnologías más importantes en el diseño de supercomputadoras son:

- La tecnología de registros vectoriales que permite la ejecución de innumerables operaciones aritméticas en paralelo.
- El sistema conocido como M.P.P. por las siglas de Massively Parallel Processors o Procesadores

Masivamente Paralelos, que consiste en la utilización de cientos y a veces miles de microprocesadores estrechamente coordinados.

- La tecnología de computación distribuida: los clusters de computadoras de uso general y relativo bajo costo, interconectados por redes locales de baja latencia y el gran ancho de banda.
- Cuasi-Supercómputo: con la popularización de la Internet, han surgido proyectos de computación distribuida en los que software especiales aprovechan el tiempo ocioso de miles de ordenadores personales para realizar grandes tareas por un bajo costo. A diferencia de las tres últimas categorías, el software que corre en estas plataformas debe ser capaz de dividir las tareas en bloques de cálculo independientes que no se ensamblarán ni comunicarán por varias horas. En esta categoría destacan BOINC y Folding@home.

El costo de las supercomputadoras, millones de dólares, hace que las universidades e instituciones de investigación cuyos presupuestos no permiten inversiones de tal envergadura opten por sistemas de cómputo de altas prestaciones basadas en clúster de PCs que funcionan como un supercomputador. La implementación de este tipo sistemas es posible debido a las características de las actuales computadoras de escritorio [1]:

- Arquitecturas especializadas VLIW, que permiten ejecutar varias instrucciones simultáneamente, especialmente en el área de procesamiento de gráficos.
- Paralelismo asimétrico a nivel de sistema, en cuanto a que los sistemas suelen tener procesadores de propósito general junto con procesadores especializados (los más potentes son los procesadores gráficos, pero también hay procesadores de física, por ejemplo).
- Arquitecturas multithread de altas prestaciones, que usan sistemas tales como el Hyperthreading para ejecutar varias tareas de forma concurrente.
- Paralelismo simétrico a nivel de sistema, con arquitecturas que incluyen dos (o potencia de dos) núcleos por microprocesador (en la práctica, varios microprocesadores). En el año 2006, por ejemplo, las arquitecturas multi núcleo ya eran habituales tanto en ordenadores de sobremesa como en portátiles. En poco tiempo, este tipo de arquitecturas se extienden a dispositivos pequeños de tipo PDA, e incluso móviles.

La conectividad es un hecho en todo tipo de ordenadores: desde la conectividad inalámbrica (WiFi, habitualmente) o por red, hasta Bluetooth en los móviles, o la conectividad intrínseca a los mismos.

Como ya se indicó, la computación de altas prestaciones hace uso de computadores paralelos o sistemas distribuidos. Éstos disponen de múltiples procesadores que son capaces de trabajar conjuntamente en una o varias tareas a la vez para resolver un problema computacional. Las arquitecturas que dan soporte a estos sistemas se pueden clasificar de acuerdo a [2] [3] [4] [5] [6] en:

- SIMD (Simple Instruction Multiple Data): un solo flujo de instrucciones se aplica concurrentemente a múltiples conjuntos de datos. En una arquitectura SIMD procesos homogéneos (con el mismo código) ejecutan sincrónicamente la misma instrucción sobre sus datos, o bien la misma operación se aplica sobre unos vectores de tamaño fijo o variable.
- MISD (Multiple Instruction Simple Data): los procesadores tratan de forma diferente el mismo conjunto de datos. Son útiles en casos donde se deban realizar muchas operaciones diferentes sobre el mismo conjunto de datos.
- MIMD (Multiple Instruction Multiple Data): paralelismo funcional y/o de datos. No sólo se distribuyen los datos entre los diferentes procesadores, sino también las tareas a realizar. Varios flujos (posiblemente diferentes) de ejecución se aplican a diferentes conjuntos de datos.
- SPMD (Simple Program Multiple Data): en paralelismo de datos [7], se utiliza el mismo código con distribución de datos. Se crean varias instancias de las mismas tareas, cada una ejecutando el código de forma independiente.

Otra forma de clasificar las arquitecturas es en base a los mecanismos de control y la organización del espacio de direcciones [8] [9] [10], por ejemplo, para las arquitecturas MIMD:

- Sistemas MIMD con memoria compartida (Shared Memory – MIMD), o también denominados Multiprocesadores. En este modelo, los procesadores comparten el acceso a una memoria común por lo que existe un único espacio de direcciones de memoria para todo el sistema, accesible para todos los procesadores. Éstos se comunican a través de la memoria compartida. Los desarrolladores no tienen que preocuparse sobre la posición actual de almacenamiento de los datos.
- Sistemas MIMD con memoria distribuida (Distributed Memory – MIMD), o también llamados Multicomputadores. En este modelo, cada procesador ejecuta un conjunto separado de instrucciones sobre sus propios datos locales. La memoria, no centralizada, está distribuida entre los procesadores del sistema, cada uno con su

propia memoria local donde posee su propio programa y datos asociados. El espacio de direcciones de memoria no está compartida entre los procesadores. Una red de interconexión conecta los procesadores (y sus memorias locales), mediante enlaces (links) de comunicación, usados para el intercambio de mensajes entre los procesadores. Los procesadores intercambian datos entre sus memorias cuando se pide el valor de variables remotas.

Debido a la fuerte dependencia de las arquitecturas de las máquinas que se usen y de los paradigmas de programación usados en su implementación, aún no existe una metodología claramente definida para la creación de aplicaciones paralelas [11]. Una metodología simple de creación de aplicaciones paralelas [9], es la que estructura el proceso de diseño en cuatro etapas: Partición, Comunicación, Aglomeración y Mapping (PCAM). Las dos primeras se enfocan en la concurrencia y la escalabilidad, y su objetivo es desarrollar algoritmos que prioricen estas características. En las dos últimas etapas, la atención se desplaza hacia la localidad y las prestaciones ofrecidas. Como resultado de esta metodología y dependiendo de los paradigmas y arquitecturas físicas, se puede obtener:

- una aplicación paralela, con un mapping estático entre tareas y procesadores a la hora de iniciar una aplicación o
- una aplicación que crea tareas (y las destruye) en forma dinámica, por medio de técnicas de balanceo de carga.

Por esto, una de las decisiones importantes en las aplicaciones paralelas es, precisamente, la elección del paradigma de programación.

Los paradigmas de programación paralelos son una clase de algoritmos que solucionan diferentes problemas bajo las mismas estructuras de control [12]. Bajo la idea de paradigma se encapsula una determinada forma de realizar el control y las comunicaciones de la aplicación. Hay diferentes clasificaciones de paradigmas de programación, pero un subconjunto habitual [13] en todas ellas es:

- Master-Worker: también conocido como Task-farming, consiste en dos tipos de entidades, un master y múltiples esclavos (workers). El master es responsable de descomponer el problema a computar en diferentes trabajos más simples (o en subconjuntos de datos), los cuales distribuye sobre los workers y finalmente recolecta los resultados parciales para componer el resultado final de la computación. Los workers ejecutan un ciclo muy simple: reciben un mensaje con el trabajo, lo procesan y envían el resultado parcial de vuelta al master. Usualmente sólo hay comunicación entre el master y los workers. Uno de los parámetros importantes de este modelo es el número de workers utilizados.

- SPMD (Single Program Multiple Data): se crea un número de tareas fijas idénticas al inicio de la aplicación, y no se permite la creación o destrucción de tareas durante la ejecución. Cada tarea ejecuta el mismo código, pero con diferentes datos. Normalmente esto significa partir los datos de la aplicación entre los nodos (o máquinas) disponibles en el sistema que participan en la computación. Este tipo de paralelismo también se conoce como Paralelismo Geométrico, Descomposición por Dominios o Paralelismo de Datos.
- Pipelining: este paradigma está basado en una aproximación por descomposición funcional, donde cada tarea se tiene que completar antes de comenzar la siguiente, en sucesión. Cada proceso corresponde a una etapa del pipeline, y es responsable de una tarea particular. Todos los procesos son ejecutados concurrentemente y el flujo de datos se realiza a través de las diferentes fases del pipeline, de manera que la salida de una fase es la entrada de la siguiente. Uno de los parámetros importantes aquí es el número de fases.
- Divide and Conquer: el problema a computar se divide en una serie de subproblemas. Cada uno de estos subproblemas se soluciona independientemente y sus resultados se combinan para producir el resultado final. Si el subproblema consiste en instancias más pequeñas del problema original, se puede entonces, realizar una descomposición recursiva hasta que no puedan subdividirse más. Normalmente en este paradigma son necesarias operaciones de partición (split o fork), computación, y unión (join). Normalmente se trabaja con dos parámetros, la anchura del árbol, el grado de cada nodo, y la altura del árbol, el grado de recursividad.
- Paralelismo especulativo: se utiliza como alternativa si no puede obtenerse paralelismo con alguno de los paradigmas anteriores. Consiste en realizar algunos intentos de paralelización de cómputo, en base a elegir diferentes algoritmos para resolver el mismo problema. El primero en obtener la solución final es el que se escoge entre las diferentes opciones probadas. Otra posibilidad es, si el problema tiene dependencias de datos complejas, puede intentarse ejecutar de forma optimista, con una ejecución especulativa de la aplicación, con posibles vueltas atrás, si aparecen problemas de incoherencias o dependencias no tenidas en cuenta o previstas, que obliguen a deshacer cómputo realizado previamente si algo falla. En caso de acertar, o si no aparecen problemas, se puede incrementar el rendimiento de forma significativa.

Cada paradigma de programación se refiere a una clase de algoritmos que tienen la misma estructura de control [12] [14] y que se pueden implementar usando un modelo genérico de programación paralela. Los modelos de programación solucionan la distribución de código y la interconexión entre las unidades de ejecución y las tareas. Cualquiera de estos modelos puede implementar los diferentes paradigmas de programación paralela. Sin embargo, las prestaciones de cada combinación resultante (paradigma en un determinado modelo) dependerán del modelo de ejecución subyacente (la combinación de hardware paralelo, red de interconexión y software de sistema disponibles).

Se puede clasificar los modelos de programación, según [8] [9], en:

- Memoria compartida: en este modelo [15] [16] los programadores ven sus programas como una colección de procesos que acceden a variables locales y un conjunto de variables compartidas. Cada proceso accede a los datos compartidos mediante una lectura o escritura asíncrona. Por tanto, como más de un proceso puede realizar las operaciones de acceso a los mismos datos compartidos en el mismo tiempo, es necesario implementar mecanismos (semáforos o bloqueos) para resolver los problemas de exclusiones mutuas que se puedan plantear. En este modelo el programador ve la aplicación como una colección de tareas, que normalmente se asignan a threads de ejecución en forma asíncrona. OpenMP [17] es una de las implementaciones más utilizadas para programar bajo este modelo.
- Paralelismo de Datos: el paralelismo de datos, es un paradigma en el que [3] se realizan operaciones semejantes (o iguales) sobre varios elementos de datos simultáneamente, por medio de la ejecución simultánea en múltiples procesadores. Este modelo [18] es aconsejable para las aplicaciones que realizan la misma operación sobre diferentes elementos de datos. La idea principal es explotar la concurrencia que deriva de realizar las mismas operaciones sobre múltiples elementos de las estructuras de datos. Un programa paralelo de datos consiste en una lista de las mismas operaciones a realizar en una estructura de datos. Entonces, cada operación en cada elemento de datos puede realizarse en forma de una tarea independiente. En cuanto a los lenguajes que soportan este modelo, se puede citar a Fortran 90, y su extensión High Performance Fortran [19] [9].
- Paso de Mensajes: este es el modelo más ampliamente usado [8] [20]. En él los programadores organizan sus programas como una colección de tareas con variables locales privadas y la habilidad de enviar, y recibir datos entre tareas por medio del intercambio de mensajes. Definiéndose así por sus dos atributos

básicos: un espacio de direcciones distribuido y soporte únicamente al paralelismo explícito. Los mensajes pueden ser enviados vía red o usando memoria compartida si está disponible, dependiendo del modelo arquitectural del sistema (o de la virtualización software empleada). Las comunicaciones entre dos tareas ocurren a dos bandas, donde los dos participantes tienen que invocar una operación. Se puede denominar a estas comunicaciones como operaciones cooperativas ya que deben ser realizadas por cada proceso, el que tiene los datos y el que quiere acceder a ellos. También, en algunas implementaciones [21] [22], pueden existir comunicaciones de tipo (one-sided), si es sólo un proceso el que invoca la operación, colocando todos los parámetros necesarios, y la sincronización se hace de forma implícita. Como ventajas, el paso de mensajes permite un enlace con el hardware existente, ya que se corresponde bien con arquitecturas que tengan una serie de procesadores conectados por una red de comunicaciones (ya sea interconexión interna, o red cableada). En cuanto a la funcionalidad, incluye una mayor expresión disponible para los algoritmos paralelos, proporcionando control no habitual en el paralelismo de datos, o en modelos basados en paralelismo implícito por compilador paralelizante. PVM [23] y MPI [24] [25] constituyen ejemplos de librerías que soportan el modelo de paso de mensajes.

2. LINEAS DE INVESTIGACION y DESARROLLO

Aunque a nivel mundial el tema de altas prestaciones está en auge, a nivel regional, particularmente en el NOA, los profesionales de informática no han incursionado en el área del Cómputo de Altas Prestaciones y los investigadores de las diferentes disciplinas científicas que requieren utilizar aplicaciones que requieren gran capacidad de cómputo no tienen acceso a estos sistemas.

A continuación se enumeran los ejes del tema de investigación:

- Evaluación de las aplicaciones existentes para la instalación y configuración de un cluster de computadoras.
- Instalación y Configuración de un cluster de PCs con dos sistemas operativos.
- Evaluación de librerías MPI para el desarrollo de programas paralelos.
- Evaluación de Gestores de Recursos para Cluster de PC (PBS, SGE).
- Selección y Prueba de aplicaciones clásicas que requieren procesamiento paralelo (TSP, 9 Reinas).
- Análisis del rendimiento de la aplicaciones de prueba en el cluster.

3. RESULTADOS OBTENIDOS/ESPERADOS

Dada la reciente formación del grupo de investigación aun se trabaja en la evaluación del clúster y el análisis del rendimiento de las aplicaciones que se ejecutan en este.

Los resultados obtenidos hasta ahora son:

- Instalación y configuración de un clúster formado por 12 PCs.
- Instalación y Configuración de la librería de paso de mensajes LAM-MPI.
- Estudio y selección de los gestores de recursos TORQUE, SGE y LFS. De este estudio se optó por elegir PBS (versión gratuita de TORQUE). Aunque no se lo utiliza en la actualidad por un problema logístico.

Actualmente se trabaja en:

- La paralelización de aplicaciones científicas y su evaluación en el cluster.
- La implementación del gestor de recursos.

Además, se espera seguir avanzando en la optimización del cluster para poder proporcionar su servicio a quienes lo requieran.

4. FORMACION DE RECURSOS HUMANOS

La mayoría de los miembros del grupo de investigación realizan estudios de posgrado en HPC en la Universidad Nacional de La Plata y se cuenta con la asesoría constante de los Dres. Emilio Luque y Dolores Rexachs, profesores de la Universidad Autónoma de Barcelona y de la UNLP y del personal de la UAB que trabaja en el área del cómputo de Altas Prestaciones. Además, algunos integrantes de este proyecto integran un grupo de investigación (formado en el 2009) dedicado a la Tolerancia a Fallos en Cómputo de Altas Prestaciones, donde tienen como tarea, entre otras, la formación en cuestiones de HPC de alumnos avanzados de la carrera Ingeniería Informática.

Este año se dictarán cursos para los jóvenes de la facultad en temas de HPC, dándoles la oportunidad de realizar las pruebas en el cluster real de reciente creación.

También se tiene previsto ofrecer opciones de trabajos finales de carrera en HPC a los alumnos de Ingeniería en Informática.

5. BIBLIOGRAFIA

- [1] J. Merelo, J. L. Jiménez Laredo. "Introducción a las Arquitecturas de Altas Prestaciones". www.geneura.ugr.es/~jmerelo/asignaturas/AAP/AAP-Tema-1.mhtml.
- [2] S.G. Akl. The Design and Analysis of Parallel Algorithms. Prentice Hall, Englewood Cliffs, NJ, 1989.
- [3] George S. Almasi and Alan Gottlieb. Highly Parallel Computing, 2nd ed. Benjamin/Cummings division of Addison Wesley Inc., Redwood City, CA, 1st edition, 1994.
- [4] Oliver A. McBryan. An overview of message passing environments. Parallel Computing, 1994.

- [5] Aad J.vander Steen and Jack J. Dongarra. Overview of recent supercomputers. Technical Report UT-CS-96-325, Department of Computer Science, University of Tennessee, April 1996.
- [6] Aad van der Steen and Jack Dongarra. Overview of high performance computers. In Handbook of Massive Data Sets. Kluwer Academic Publishers Group, Norwel, MA, USA, and Dordrecht, The Netherlands, 2002.
- [7] W. Daniel Hilis and Guy L. Steele. Data paralel algorithms. Communications of the ACM, 29(12):1170–1183, December 1986.
- [8] Vipin Kumar, Ananth Grama, Anshul Gupta, and George Karypis. Introduction to Parallel Computing: Design and Analysis of Algorithms. Benjamin/Cummings, Redwood City, CA, 1994.
- [9] I Foster. Designing and Building Parallel Programs. Addison Wesley, 1995.
- [10] David E. Culler, Jaswinder Paul Singh and Anoop Gupta. Parallel Computer Architecture: A Hardware/Software Approach. Morgan Kauffman Publishers, inc., San Francisco, CA, 1999.
- [11] Jorba Esteve, Josep. “Análisis automática de prestaciones de aplicaciones basadas en paso de mensajes”. Tesis doctoral – Universidad Autónoma de Barcelona. Barcelona. España. 2006.
www.tesisenxarxa.net/TESIS_UAB/AVAILABL E/TDX-1013106-132034//jje1de1.pdf
- [12] Per Brinch Hansen. Model programs for computational science: A programming methodology for multicomputers. Concurrency: Practice & Experience, 5(5):407–423, August 1993.
- [13] R. Buyya, editor. High Performance Cluster Computing: Architectures and Systems. Prentice Hal, June 1999.
- [14] D.I. Moldovan. Parallel Processing -From Applications to Systems. Morgan Kaufmann, San Mateo, 1993.
- [15] Moti Ben-Ari. Principles of Concurrent Programming. Prentice-Hal International, Inc., Englewood Clifs, 1990.
- [16] Satish Chandra, James R. Larus, and Anne Rogers. Where is time spent in message-passing and shared-memory programs? Technical Report TR-463-94, Princeton University, Computer Science Department, July 1994.
- [17] F. Wolf and B. Mohr. Automatic performance analysis of hybrid mpi/openmp applications. In Parallel, Distributed and Network Based Processing, 2003. Proceedings. Eleventh Euromicro Conference on, pages 13–22, 2003.
- [18] Philip J. Hatcher and Michael J. Quinn. Data-Parallel Programming on MIMD Computers. The MIT Pres, Cambridge, MA, 1991.
- [19] B. Chapman, P. Mehrotra, and H. Zima. Extending HPF for advanced data-parallel applications. IEEE Parallel and Distributed Technology, 2(3):15–27, 1994.
- [20] Bary Wilkinson and Michael Allen. Parallel Programming. Prentice Hall, 1999.
- [21] Message Passing Interface Forum. Mpi-2: Extensions to the message-passing interface. <http://www.mpi-forum.org/docs/mpi-20-html/mpi2-report.html>, 1997.
- [22] Weihang Jiang, Jiuxing Liu, Hyun-Wook Jin, D.K. Panda, W.Gropp, and R. Thakur. High performance mpi-2 one-sided communication over infiniband. In Cluster Computing and the Grid, 2004. CC Grid 2004. IEEE International Symposium on, pages 531–538, 2004.
- [23] Al Geist, Adam Beguelin, Jack Dongarra, Weicheng Jiang, Robert Manchek, and Vaidyalingam S. Sunderam. PVM: Parallel Virtual Machine A Users’ Guide and Tutorial for Network Parallel Computing. Scientific and Engineering Computation Series. MIT Pres, Cambridge, MA, 1994.
- [24] Wiliam Gropp, Ewing Lusk Anthony Skjelum, and Nathan Doss. A high-performance, portable implementation of the MPI message passing interface standard, December 26 1996.
- [25] J. Dongarra, S. Otto, M. Snir, and D. Walker. A message passing standard for mpp and workstations. Communications of the ACM, 39(7):84–92, 1996.

Aplicaciones de Búsquedas por Similitud en un Portal GRID Orientado a Aspectos*

Osiris Sofia, Sandra Casas

Universidad Nacional de la Patagonia Austral
Lisandro de la Torre 1070
CP 9400. Río Gallegos, Santa Cruz, Argentina
Tel/Fax: +54-2966-442313/17
{osofia;lis}@unpa.edu.ar

y

Roberto Uribe Paredes

Departamento de Ingeniería en Computación
Universidad de Magallanes
Punta Arenas, Chile
ruribe@ona.fi.umag.cl

Resumen

En el marco del Proyecto de Investigación *Paralelización de Estructuras de Datos y Algoritmos para la Recuperación de Información*, de la Universidad Nacional de la Patagonia Austral se desarrolla la línea de investigación que da continuidad al trabajo previo de dos grupos de investigación de esa Universidad en las áreas de paralelismo y programación orientada a aspectos, integrados en esta oportunidad en la búsqueda de alternativas para la implementación del paradigma de orientación a aspectos en entornos paralelos, particularmente en entornos GRID, a través del desarrollo de un *Portal GRID*, utilizando técnicas y herramientas de la programación orientada a aspectos, y del área de búsquedas por similitud investigadas por un grupo de la Universidad de Magallanes, Chile, con el fin de desarrollar aplicaciones arqueológicas para el Portal GRID.

Palabras claves: Computación GRID, Programación Orientada a Aspectos, Portales GRID, Paralelismo, Búsquedas por Similitud.

1. CONTEXTO

En los últimos años, de manera separada han surgido y crecido dos líneas de investigación en

*Este trabajo fue financiado por la Universidad Nacional de la Patagonia Austral, Santa Cruz, Argentina, proyecto 'Paralelización de Estructuras de Datos y Algoritmos para la Recuperación de Información'

la Unidad Académica Río Gallegos de la UNPA. A partir del año 1999 se ha trabajado sobre problemas de la distribución y paralelización de bases de datos y desde el año 2005 un grupo diferente de investigadores ha estudiado el paradigma orientado a aspectos en la búsqueda de estrategias de resolución de conflictos. Por otra parte desde el comienzo de la investigación en paralelismo se ha trabajado en colaboración con la Universidad de Magallanes, Chile, sede Punta Arenas. Actualmente ese grupo desarrolla sus investigaciones en temas relacionados con las búsquedas por similitud. Los tres grupos se encuentran en un estado de consolidación suficiente para confluir sus esfuerzos en un proyecto que unifique sus líneas de investigación.

2. INTRODUCCIÓN

La *Computación Grid* [10] se ha establecido como un nuevo paradigma para la computación científica de gran escala (o redes de investigación): la aplicación de recursos computacionales coordinados, interconectados vía una red pública de alta velocidad, para solucionar problemas en áreas científicas tales como: astrofísica, química, física, geofísica, meteorología y climatología, neurobiología, biología molecular, etc. Las aplicaciones Grid son sistemas computacionales distribuidos que proveen mecanismos para compartir de manera controlada recursos computacionales. La Computación Grid requiere componentes middleware genéricos, que oculten a las aplicaciones específicas los detalles de acceso

y usar configuraciones de recursos heterogéneos, procesadores, almacenamiento y conexiones de redes. Estos garantizan la interoperabilidad de los recursos a través del uso de protocolos estándares. El término *Tecnología Grid* [14] usualmente se refiere a éste tipo de middleware.

Uno de los enfoques más utilizados para proporcionar acceso a la Computación GRID son los Portales GRID. Los Portales GRID [7] son herramientas muy eficaces que proporcionan a los usuarios de Computación GRID interfaces simples e intuitivas para el acceso a la información y recursos GRID [10]. La construcción de un Portal GRID debe cumplir con todos los requerimientos de servicios y recursos, para lo cual se han desarrollado APIs, Toolkits, Frameworks específicos, tales como GridPort [12], GPKD [11] y P-GRADE [3].

La Programación Orientada a Aspectos [8] (POA) es un relativamente nuevo paradigma para el desarrollo de software que proporciona abstracciones para la implementación de los crosscutting concerns, de manera separada y aislada a los componentes de funcionalidad básica. La POA además proporciona mecanismos para la composición de las diferentes unidades. En otras palabras, la Orientación a Aspectos, es una técnica que permite aplicar el principio de Separación de Concerns [6] y de esta forma, obtener los beneficios enunciados por dicho principio. La unidad de implementación que representa a la funcionalidad transversal se denomina aspecto, dando origen al nombre del paradigma. De esta forma, se suele referir casi sin distinción para indicar el mismo concepto a los términos aspecto, funcionalidad transversal y/o “crosscutting concern”.

Ciertas funcionalidades y requerimientos han sido identificados como “clásicos” crosscutting concerns. Entre estos suelen identificarse: coordinación, distribución, sincronización, concurrencia, balance de carga, seguridad, logging y autenticación. En el desarrollo e implementación de portales GRID, todos o algunos de estos concerns estarán presentes, por lo que se puede suponer a priori, que al desarrollar una aplicación GRID pueden ser implementados bajo el enfoque POA.

2.1. Portales GRID

Un portal GRID es en esencia una aplicación WEB, por lo cual tiene requerimientos (o características) similares a los portales orientados al consumo o usuario (Yahoo, CNN, IBM intranet). Estos servicios suelen incluir soporte para el contexto (login, customización, personali-

zación, etc.); soporte para interfaces de usuario basadas en navegadores; páginas dinámicas disponibles a usuarios anónimos o autenticados. En particular los portales de e-Science deben además soportar cuestiones relacionadas con la integración de aplicaciones de dominio específicas basadas en GRID. Aquí surge la principal diferencia, un portal GRID debe manejar computación que se ejecute por días o semanas sobre cientos de nodos para procesar terabytes de datos científicos. Específicamente estos portales son requeridos para manejar credenciales, lanzar trabajos, manejar ficheros y ocultar la complejidad de la GRID como los trabajos batch distribuidos. En la Figura 1 se comparan ambos tipos de portales [17].

	Grid	Webs	Comentarios
Principales Usos / Usuarios	eScience, eEngineering	Comunicación científica (inicialmente), eCommerce, eContent (multimedia)	Existen algunos solapamientos y habrán más en el futuro.
Principales funciones	Computación de alta performance, compartición de recursos computacionales	Información, comunicación, transacciones	
Aplicaciones	Problemas que requieren computo intensivo en ciencias e ingeniería	Servicios I&C, educación & entrenamiento, eBusiness, eCommerce (B2B, B2C, B2A, etc.), etc.	Las Webs son interfaces principales para acceso a las aplicaciones
Volúmenes de Datos	XXL (y mayores)	S - XL	Grid futuras pueden también trabajar sobre volúmenes más pequeños.
Recursos	Almacenamiento, ancho de banda, tiempo de procesador, ficheros, etc.	Contenido digital y relacionado a servicios.	Contenedores, transportadores & Procesadores vs. Contenido y aplicaciones
Usuarios	Grupos de usuarios especiales (científicos, ingenieros)	Público en general, negocios, administradores, etc.	Estos son solo algunos de los grupos objetivo.
Standares	Faltan estándares para middleware	Existen algunos estándares y recomendaciones	Las comunidades Grid y Web están todavía lejanamente separadas.

Figura 1: *Comparación de Portales Grids y Portales Webs Orientados al consumo/usuario.*

A continuación se detallan los servicios básicos que un Portal GRID debe ofrecer y el modelo arquitectónico de base empleado en el diseño e implementación de los mismos.

2.2. Servicios soportados por un Portal GRID

Los servicios que un Portal GRID típicamente incluye son:

- *Seguridad*: los usuarios se loguean en un portal usando un navegador WEB y se autentican mediante un user-id y password. El portal GRID

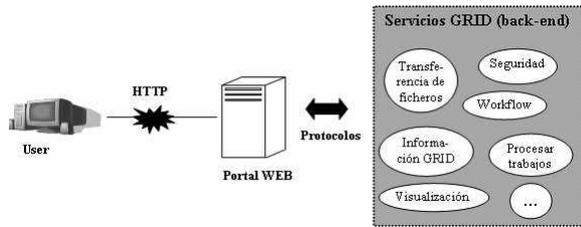


Figura 2: *Arquitectura Base de Portales Grid.*

mapea el user-id a credenciales GRID.

- *Gestión de Datos*: provee acceso a ficheros, colecciones y metadatos locales y remotos, soporta transferencia de ficheros;
- *Job submission*: se refiere a la habilidad de que los procesadores conectados a la GRID ejecuten un trabajo (secuencial o paralelo) y puedan monitorear su estado. Este es un servicio clásico soportado por el portal;
- *Servicios de Información*: el acceso a directorios y estado de herramientas es un rol esencial del Portal;
- *Interfaces de Aplicación*: permite ocultar convenientemente los detalles GRID detrás de una interfaz de aplicación;
- *Colaboración*: los portales sirven como entradas a organizaciones virtuales para compartir recursos;
- *Workflow*: presenta los usuarios con sus tareas y asume la responsabilidad de integrar estas tareas en secuencias;
- *Visualización*: provee herramientas que ofrecen a los usuarios acceso a los datos, renderización y visualización de recursos. Puede proveer algún nivel de vista de los datos o puede ser usada para ofrecer herramientas más avanzadas.

2.3. Arquitectura base de portales GRID

Un Portal GRID se puede ver como una interface WEB a un sistema distribuido. La arquitectura básica de un portal responde a un esquema de arquitectura de tres capas: (1) la capa cliente, que se ejecuta mediante un navegador WEB; (2) la capa servidor que cumple la función de representar la lógica del negocio y (3) la capa de recursos y servicios GRID. Los clientes y el servidor típicamente se comunican vía HTTP permitiendo que cualquier navegador WEB sea usado. La capa servidor simplemente accede a ficheros locales para servir páginas pero también puede dinámicamente generar páginas web mediante la ejecución de scripts CGI y/o mediante interacción directa o indirecta con los recursos back-end. La interacción con la tercera capa puede lograrse en algún protocolo o de manera apropiada. Usando esta arquitectura general, los portales pueden ser

construidos para que soporten aplicaciones de una amplia variedad. Para hacerlo efectivamente, sin embargo, se requiere un conjunto de herramientas de construcción de portales que puedan ser personalizados para cada área de aplicación. En la Figura 2 se presenta gráficamente una arquitectura de tres capas aplicada al acceso y utilización de recursos y servicios GRID.

2.4. Búsquedas por similitud en espacios métricos

Uno de los problemas de gran interés en ciencias de la computación es el de “búsqueda por similitud”, es decir, encontrar los elementos de un conjunto más similares a una muestra. Esta búsqueda es necesaria en múltiples aplicaciones, como ser en reconocimiento de voz e imagen, compresión de video, genética, minería de datos, recuperación de información, etc. En casi todas las aplicaciones la evaluación de la similaridad entre dos elementos es cara, por lo que usualmente se trata como medida del costo de la búsqueda la cantidad de similaridades que se evalúan. Interesa el caso donde la similaridad describe un espacio métrico, es decir, está modelada por una función de distancia que respeta la desigualdad triangular. En este caso, el problema más común y difícil es en aquellos espacios de “alta dimensión” donde el histograma de distancias es concentrado, es decir, todos los objetos están más o menos a la misma distancia unos de otros. El aumento de tamaño de las bases de datos y la aparición de nuevos tipos de datos sobre los cuales no interesa realizar búsquedas exactas, crean la necesidad de plantear nuevas estructuras para búsqueda por similaridad o búsqueda aproximada. Asimismo, se necesita que dichas estructuras sean dinámicas, es decir, que permitan agregar o eliminar elementos sin necesidad de crearlas nuevamente, así como también que sean óptimas en la administración de memoria secundaria. La necesidad de procesar grandes volúmenes de datos obligan a aumentar la capacidad de procesamiento y con ello la paralelización de los algoritmos y la distribución de las bases de datos. El objetivo de los algoritmos de búsqueda es minimizar la cantidad de evaluaciones de distancia realizadas para resolver la consulta. Los métodos para buscar en espacios métricos se basan principalmente en dividir el espacio empleando la distancia a uno o más objetos seleccionados. El no trabajar con las características particulares de cada aplicación tiene la ventaja de ser más general, pues los algoritmos funcionan con cualquier tipo de objeto [4]. Existen distintas estructuras para buscar en espacios métricos, las cuales pueden ocupar funciones discretas o continuas de distancia. Algunos son BKTree [2], GNAT [1], VpTree [23], MTree [5],

SAT [13], EGNAT [22]. Algunas de las estructuras anteriores basan la búsqueda en pivotes y otras en clustering. En el primer caso se seleccionan pivotes del conjunto de datos y se precálculan las distancias entre los elementos y los pivotes. Cuando se realiza una consulta, se calcula la distancia de la consulta a los pivotes y se usa la desigualdad triangular para descartar candidatos.

3. LÍNEA DE INVESTIGACIÓN Y DESARROLLO

El objetivo del trabajo es demostrar empíricamente que la orientación a aspectos es una técnica de desarrollo de software más conveniente para la construcción de Portales GRID que las técnicas convencionales de desarrollo de software (orientadas a objetos y/o componentes).

La hipótesis principal consiste en que el diseño e implementación de un Portal GRID empleando la orientación a aspectos genera una aplicación más reutilizable, mantenible, escalable, evolucionable y traceable.

El método de investigación es eminentemente empírico, para lo cual se seleccionará un caso de estudio particular y real. Para este caso concreto se desarrollará una aplicación consistente en un simulador arqueológico que implementará además búsquedas por similitud.

Hasta el momento se han realizado las etapas de *Estudio del Estado del Arte*, *Identificación de requerimientos* y *Diseño Arquitectónico*. Las etapas pendientes se describen a continuación.

Implementación: En esta etapa se debe en principio seleccionar las herramientas de programación más adecuadas, garantizando la compatibilidad entre las mismas. El lenguaje de componentes para la funcionalidad base y un lenguaje orientado a aspectos, además del servidor web y herramientas para servicios específicos del Portal GRID.

Pruebas: En principio se establecen pruebas funcionales que garanticen el correcto funcionamiento del Portal GRID, pero además se establecerán pruebas de performance y rendimiento.

Comparación: En esta etapa se pretende realizar diversos estudios comparativos con portales GRID desarrollados bajo enfoques diferentes.

4. RESULTADOS OBTENIDOS / ESPERADOS

Durante el primer año del proyecto se han realizado esfuerzos tendientes al cabal entendimiento de las distintas tecnologías que se pretende integrar. De este manera distintos subgrupos han abordado la investigación y publicación de temas tales como GRID [18, 19, 21], Portales GRID [15], simuladores grid, desarrollo de aplicaciones

paralelas utilizando la programación orientada a aspectos [9], y paralelización de estructuras de búsqueda por similitud [16].

Durante esta segunda etapa se espera integrar las distintas tecnologías, de manera de implementar aplicaciones reales que realicen búsquedas por similitud en el entorno de un Portal GRID desarrollado bajo el modelo de programación orientada a aspectos.

5. FORMACION DE RECURSOS HUMANOS

La integrante del grupo de investigación de la Universidad Nacional de la Patagonia Austral Natalia Bibiana Trejo se encuentra desarrollando a través de una beca obtenida de la Fundación Carolina y la propia Universidad, el doctorado en Informática otorgado por la Universidad Complutense de Madrid, España. Para la obtención de dicho título es requisito obtener previamente el de Master en Investigación en Informática, lo que ha realizado en el transcurso del año 2008 con la tesis denominada 'Aplicación de Multicast IPv6 Seguro a Servicios de Información en Entornos Grid', bajo la dirección del Dr. Juan Carlos Fabero Jiménez [20]. Actualmente continúa sus estudios para obtener el título de doctorado.

Durante el año 2008 se ha postulado un alumno como becario del proyecto, con el objetivo de desarrollar una aplicación que valide empíricamente el modelo propuesto para el portal GRID. El alumno Franco Herrera desarrollará una aplicación consistente en un simulador arqueológico. Se ha promovido la participación de los auxiliares de docencia a través de la presentación de los trabajos aprobados en los distintos congresos donde se realizaron aportes, de manera de fomentar su interés no sólo en la etapa de investigación sino también de la transferencia de los resultados y la comunicación con otros grupos del país y del exterior.

Actualmente dos alumnos de la carrera de Licenciatura en Sistemas participan como integrantes en el proyecto de investigación.

Por último durante el 2008 se han realizado pasantías entre integrantes de los grupos de la Universidad Nacional de la Patagonia Austral y la Universidad de Magallanes para intercambiar habilidades en la aplicación de las áreas de especialización de cada grupo.

Referencias

- [1] Sergei Brin. Near neighbor search in large metric spaces. In *the 21st VLDB Conference*, pages 574–584. Morgan Kaufmann Publishers, 1995.

- [2] W. Burkhard and R. Keller. Some approaches to best-match file searching. *Communication of ACM*, 16(4):230–236, 1973.
- [3] Nemeth C., Dozsa G., Lovas R., and Kacsuk P. The p-grade grid portal. *LNCS*, v. 2044, pages 10–19, 2004.
- [4] Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates, and José L. Marroquín. Searching in metric spaces. In *ACM Computing Surveys*, pages 33(3):273–321, September 2001.
- [5] P. Ciaccia, M. Patella, and P. Zezula. M-tree : An efficient access method for similarity search in metric spaces. In *the 23rd International Conference on VLDB*, pages 426–435, 1997.
- [6] Dijkstra E. *A Discipline of Programming*. Prentice-Hall, 1976.
- [7] Furmanski W. Fox G. *High performance commodity computing*, chapter 10. Morgan Kaufman Publishers, 1998.
- [8] Kiczales G., Lamping L., Mendhekar A., Maeda C., Lopes C., Loingtier J., and Irwin J. Aspect-oriented programming. In *In Proceedings ECOOP 97*, Finland, 1997.
- [9] Esteban Gesto, Sandra Casas, and Osiris Sofia. Implementación de aplicaciones paralelas bsp usando aspectos. In *XXIII Brazilian Symposiums on DataBases (SBDD) and XXII Brazilian Symposiums Software Engineering (SBES) II Latin American Workshop on Aspect-Oriented Software Development - LA-WASP 2008*, pages 96–97, october 2008.
- [10] Foster I. and Kesselman C. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufman Publishers, 1998.
- [11] Novotny J. Grid computing environments special issue of concurrency and computation practice and experience. *The Grid Portal Development Kit*, pages 1129–1144, 2002.
- [12] Thomas M., Mock S., and Boisseau J. Development of web toolkits for computational science portals: The npaci hotpage. In *9th IEEE International Symposium on High Performance Distributed Computing*, pages 308–309, 2000.
- [13] Gonzalo Navarro. Searching in metric spaces by spatial approximation. *The Very Large Databases Journal (VLDBJ)*, 11(1):28–46, 2002.
- [14] Proyecto Globus. <http://www.globus.org/>.
- [15] Albert Osiris Sofia and Sandra Isabel Casas. Survey de tecnologías grid. *Revista Ciencia y Técnica Administrativa*, 07(04), 2008. <http://www.cyta.com.ar/ta0704/v7n4a1.htm>.
- [16] Roberto Solar-Gallardo, Roberto Uribe-Paredes, Esteban Gesto, and Osiris Sofia. Implementación de un digesto digital paralelo para búsquedas por similitud sobre documentos. In *XIV Congreso Argentino de Ciencias de la Computación - CACIC 2008*, october 2008.
- [17] Hans-Georg Stork. Webs, grids and knowledge spaces, - programmes, projects and prospects. In *Journal of Universal Computer Science*, volume 8, pages 848–868, 2002.
- [18] Natalia Trejo and Juan C. Fabero. Aplicación de multicast ipv6 a servicios de información en entornos grid. In *XXXIV Conferencia Latinoamericana de Informática - CLEI 2008*, pages 183–192, september 2008.
- [19] Natalia Trejo and Juan C. Fabero. Grid resource discovery using signed ipv6 multicast. In *Jornadas Chilenas de Computación 2008. XII Workshop de Sistemas Distribuidos y Paralelismo. JCC 2008*, pages 20–21, november 2008.
- [20] Natalia Bibiana Trejo. Aplicación de multicast ipv6 seguro a servicios de información en entornos grid. Tesis de máster en investigación en informática, Facultad de Informática, Universidad Complutense de Madrid, 2008.
- [21] Natalia Bibiana Trejo and Juan Carlos Fabero Jiménez. Descubrimiento de recursos grid utilizando multicast ipv6 seguro. *Revista de Ingeniería Informática, JCC 2008. Selección de los mejores artículos del Workshop de Sistemas Distribuidos y Paralelismo*, 01(01), 2008. <http://revista.inf.udp.cl/index.php>.
- [22] Roberto Uribe-Paredes. Manipulación de estructuras métricas en memoria secundaria. Master’s thesis, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile, Abril 2005.
- [23] P. Yianilos. Data structures and algorithms for nearest neighbor search in general metric spaces. In *4th ACM-SIAM Symposium on Discrete Algorithms (SODA ’93)*, pages 311–321, 1993.

Sistemas Paralelos sobre Arquitecturas Distribuidas. Cluster / Multicluster / Grid / Cloud Computing.

Armando De Giusti, Marcelo Naiouf, Laura De Giusti, Franco Chichizola, Mónica Denham, Ismael Rodríguez, Adrián Pousa, José E. Pettoruti, Diego Montezanti, Diego Encinas, Luciano Iglesias, Horacio Villagarcía Wanza, Leandro Bertogna, Emmanuel Frati, Victoria Sanz, Fabiana Leibovich .

Instituto de Investigación en Informática LIDI (III-LIDI)

Facultad de Informática – UNLP

{degiusti, mnaiouf, ldgiusti, francoch, mdenham, ismael, apousa, josep, dmontezanti, dencinas, li, hvw}@lidi.info.unlp.edu.ar, mlbertog@uncoma.edu.ar, {fefrati, vsanz, fleibovich}@lidi.info.unlp.edu.ar,

CONTEXTO

Esta línea de Investigación forma parte de dos de los Subproyectos dentro del Proyecto “Sistemas Distribuidos y Paralelos” acreditado por la UNLP y de proyectos específicos apoyados por CyTED, CIC, Agencia, IBM, Telefónica y Fundación YPF.

En el tema hay cooperación con varias Universidades de Argentina y se está trabajando con la red de Universidades iberoamericanas del proyecto CyTED “Tecnología grid como motor de desarrollo regional”. También el III-LIDI forma parte (desde la Facultad de Informática) de la iniciativa LAGrid (Latin American Grid) de IBM y el proyecto EELA2 (E-infrastructure shared between Europe and Latin America).

RESUMEN

Esta línea de I/D se enfoca en la especificación, desarrollo y evaluación de sistemas paralelos utilizando esquemas multiprocesador.

Se parte del estudio de las arquitecturas de múltiples núcleos (multicore) y se trabaja también en configuraciones distribuidas como Clusters, Multiclusters, Grids y Clouds.

Los temas fundamentales se relacionan con el estudio de los esquemas multiprocesador

como soporte para diferentes paradigmas de programación paralela, las técnicas de implementación de algoritmos paralelos sobre tales arquitecturas (desde multicores a grid), el desarrollo de modelos para predicción de performance en estas configuraciones, el desarrollo y utilización de middleware para el uso de esquemas multicluster y la resolución de aplicaciones concretas.

Keywords: *Sistemas Paralelos. Paradigmas de programación paralela. Multicores. Cluster, Multicluster, Grid y Cloud Computing. Modelos y predicción de performance. Scheduling. Virtualización. Aplicaciones.*

1. INTRODUCCION

La investigación en Sistemas Distribuidos y Paralelos es una de las líneas de mayor desarrollo en la Ciencia Informática actual [1][2]. En particular la utilización de arquitecturas multiprocesador configuradas en clusters, multiclusters y grids y soportadas por redes de diferentes características y topologías se ha generalizado, tanto para el desarrollo de algoritmos paralelos como para el de servicios WEB distribuidos [3][4][5]. En la misma línea están los desarrollos incipientes en cloud computing [6] El cambio tecnológico, fundamentalmente a partir de los procesadores multicore, ha im-

puesto la necesidad de investigar en modelos mixtos o híbridos, en los cuales coexisten esquemas de memoria compartida con mensajes [7] [8].

Es importante en este contexto investigar la modelización del comportamiento de esta clase de sistemas paralelos, así como desarrollar nuevos paradigmas para la programación eficiente de aplicaciones.

Por otra parte la heterogeneidad que caracteriza las redes y las arquitecturas Grid, se puede extender también a los procesadores multicore (buscando optimizar la funcionalidad de cada uno de los núcleos), lo que genera restricciones y desafíos para la modelización y también para problemas clásicos como el scheduling y la optimización de los algoritmos [9]. Aquí reaparece el tema del multithreading interno de los procesadores y la posibilidad de su manejo desde la aplicación. [10].

Asimismo los problemas de virtualización crecen en complejidad al considerar sistemas paralelos con componentes multicore. [11].

1.1 Definiciones básicas

Un *procesador multicore* surge a partir de integrar dos o más núcleos computacionales dentro de un mismo chip [12]. La motivación de su desarrollo se basa en los problemas de consumo de energía y generación de calor que aparecen al escalar la velocidad de un procesador. Un procesador multicore incrementa el rendimiento de una aplicación si la misma divide su trabajo entre los núcleos del mismo, en vez de utilizar un único procesador más potente.

Un *cluster* es un sistema de procesamiento paralelo compuesto por un conjunto de computadoras interconectadas vía algún tipo de red, las cuales cooperan configurando un recurso que se ve como “único e integrado”, más allá de la distribución física de sus componentes. Cada “procesador” puede tener diferente hardware y sistema operativo, e incluso puede ser un “multiprocesador” [13]. Cuando se conectan dos o más clusters sobre una red tipo LAN o WAN, se tiene un *multicluster* [14]. La configuración

más simple a considerar es la conexión de clusters homogéneos sobre una red LAN o WAN, utilizando un sistema operativo común [15].

Un *Grid* es un tipo de sistema distribuido que permite seleccionar, compartir e integrar recursos autónomos geográficamente distribuidos [16]. Un Grid es una configuración colaborativa que se puede adaptar dinámicamente según lo requerido por el usuario, la disponibilidad y potencia de cómputo de los recursos conectados. El Grid puede verse como un “entorno de procesamiento virtual”, donde el usuario tiene la visión de un sistema de procesamiento “único” y en realidad trabaja con recursos dispersos geográficamente [17] [18].

Algunos autores consideran que un Grid es un “*Cluster de Clusters*”, lo que resulta una definición algo restrictiva pero útil para el desarrollo de un proyecto relacionado con Sistemas Paralelos que migren aplicaciones paralelas de Clusters a Grid.

1.2 Aspectos de interés

- En un cluster normalmente se configura una única máquina paralela virtual que puede estar ejecutando una aplicación dedicada. Un Grid permite configurar múltiples máquinas paralelas virtuales para varios usuarios/aplicaciones simultáneas. Sin embargo, este concepto de virtualización se está extendiendo a clusters (incluso a multicores) al aumentar el número de núcleos involucrados. [19]
- Tanto clusters como Grids se basan en procesadores heterogéneos. Sin embargo en Grid esta heterogeneidad se extiende a la red de comunicaciones y al tipo de componentes en cada nodo que pueden ser procesadores (mono o multi core), instrumentos, sensores, etc. La heterogeneidad aparece reforzada al considerar el caso de procesadores multicore que son internamente heterogéneos.
- El middleware necesario para Grid es más complejo que el de los clusters. Fundamentalmente, para configurar la máquina paralela virtual es necesario una

etapa de identificación de recursos físicos y su ubicación, así como la caracterización de los procesos usuarios por prioridades. Estos problemas básicos para el scheduling se están extendiendo a clusters y multicores.

- Las herramientas para el desarrollo de aplicaciones requieren un mayor nivel de abstracción en Grid, por la complejidad y variedad de los múltiples usuarios que pueden utilizar la arquitectura [20].
- La definición de los paradigmas de programación paralela y de los compiladores de lenguajes para sistemas paralelos está fuertemente impactada por el desarrollo de los multicores: el tema de la optimización de los recursos a nivel procesador se vuelve esencial, más allá que la aplicación sea de muy alto nivel (tipo transacción WEB contra un servidor multicore).

Es interesante notar que una estructura de multicluster, visualizada como un *número limitado de clusters dedicados que cooperan en una única aplicación paralela*, es un punto intermedio entre clusters y Grid y requiere algunos servicios especiales en su middleware.

Este tipo de configuración es especialmente interesante para las aplicaciones a investigar desde las Universidades que forman parte de este proyecto.

Más aún cuando cada uno de los clusters puede estar formado por computadoras multicore, homogéneas o heterogéneas.

1.3 Sistemas Paralelos sobre arquitecturas distribuidas

El desarrollo de sistemas paralelos sobre arquitecturas distribuidas (normalmente débilmente acopladas y comunicadas por una red heterogénea) presenta algunos desafíos, entre los que pueden mencionarse:

- La heterogeneidad de las comunicaciones y su costo variable según los nodos a conectar dificulta la asignación óptima de tareas a procesadores y el balance dinámico de la carga.

- La incorporación de procesadores con diferente potencia y velocidad, que en algunos casos aparecen como multiprocesadores con memoria compartida, hace especialmente complejo el problema del balance de carga y muy crítica la localidad de datos y/o procesos.
- Los modelos para predicción de performance son complejos y agregan la incertidumbre del ancho de banda efectivo en el caso de emplear una conexión vía red WAN no dedicada.
- La granularidad óptima a emplear depende de la relación entre potencia de cómputo local y remota. Muchas veces la configuración efectiva de los nodos remotos a utilizar no es conocida a priori.
- El modelo cliente-servidor (paradigma muy empleado en algoritmos paralelos sobre clusters) se torna ineficiente al incrementar el número de nodos. Esto requiere la reformulación de algoritmos que ejecutan sobre clusters y grids. En particular el paradigma “peer to peer” resulta atractivo para investigar, así como esquemas mixtos dentro del mismo sistema paralelo [21].
- Las herramientas de software más generalizadas en clusters están evolucionando para ser utilizables en multicluster y Grid (ej. MPI o PVM) ya que tienen restricciones al tratar de emplearlas en topologías que conectan diferentes redes.
- A su vez, a nivel de procesadores multi núcleo es necesario replantear los temas de scheduling, migración de procesos, comunicaciones sincrónicas/asincrónicas y desarrollo de lenguajes y compiladores que exploten la arquitectura.

2. LINEAS DE INVESTIGACION y DESARROLLO

- ❖ Arquitectura de procesadores multicore y su software de base. Modelos y Caracterización de performance.
- ❖ Sistemas distribuidos y paralelos. Cluster, multicluster, grid y cloud

computing. Modelos y Caracterización de performance.

- ❖ Paradigmas para el desarrollo de algoritmos paralelos sobre arquitecturas multiprocesador distribuidas.
- ❖ Optimización de algoritmos sobre arquitecturas distribuidas heterogéneas.
- ❖ Lenguajes y compiladores para procesamiento paralelo.
- ❖ Modelos de arquitectura híbrida (memoria compartida localizada con mensajes sobre una red). Predicción de performance.
- ❖ Análisis (teórico y práctico) de los problemas de migración y asignación óptima de procesos y datos a procesadores. Métricas de paralelismo.
- ❖ Administración y monitorización de recursos en arquitecturas distribuidas sobre las que se implementan sistemas paralelos. Análisis de rendimiento.
- ❖ Tolerancia a fallas y seguridad en cluster, multicluster, grid y cloud.
- ❖ Optimización de las arquitecturas y procesadores, considerando la migración de algoritmos a hardware.
- ❖ Aplicaciones de tratamiento masivo de datos y simulación paramétrica sobre arquitecturas multiprocesador distribuidas.

3. RESULTADOS OBTENIDOS /ESPERADOS

- ✓ Estudiar el modelo de comportamiento de arquitecturas multicore (homogéneas y heterogéneas), su inserción en sistemas paralelos y la optimización de algoritmos sobre ellas.
- ✓ Simulación de arquitecturas (tipo multicore) y análisis de performance, incluyendo el software de bajo nivel. Estudio de la migración de software a hardware en tales arquitecturas.
- ✓ Modelizar el comportamiento de clusters y multiclusters homogéneos y hete-

rogéneos, sobre redes LAN y WAN, haciendo predicción de performance.

- ✓ Estudiar los temas de scheduling y virtualización a todo nivel de arquitecturas multiprocesador.
- ✓ Estructurar un GRID vinculando Universidades del país y del exterior (está activo)
- ✓ Estudiar predicción de performance de sistemas paralelos sobre arquitecturas distribuidas, realizando estudios experimentales para optimizar los mismos.
- ✓ Desarrollar primitivas de comunicaciones orientadas a cómputo paralelo en multicluster.
- ✓ Estudiar el overhead introducido por el middleware de Grid y como reducirlo para aplicaciones específicas.
- ✓ Investigar aplicaciones concretas de procesamiento masivo de datos y simulación paramétrica sobre Grid.
- ✓ Investigar la especificación e implementación de WEB services eficientes sobre Grid.

4. FORMACION DE RECURSOS HUMANOS

En esta línea de I/D existe cooperación a nivel nacional e internacional. Hay 7 Investigadores realizando su Doctorado en Argentina y 1 en el exterior. Asimismo hay 2 Tesistas de Maestría y 2 de Especialista en curso, y 4 alumnos avanzados están trabajando en su Tesina de Grado de Licenciatura.

En cooperación con las Universidades miembros del proyecto CyTED "Tecnología Grid como motor de desarrollo regional" se ha implementado una Especialización en Cómputo de Altas Prestaciones y Tecnología Grid que se inició en 2007 en la UNLP con Profesores europeos y de las 4 Universidades vinculadas al proyecto.

5. BIBLIOGRAFIA

1. Grama A, Gupta A, Karypis G, Kumar V. "Introduction to parallel computing". Second Edition. Pearson Addison Wesley, 2003.
2. Jordan H, Alagband G. "Fundamentals of parallel computing". Prentice Hall, 2002.
3. Dongarra J, Foster I, Fox G, Gropp W, Kennedy K, Torczon L, White A. "The Sourcebook of Parallel Computing". Morgan Kauffman Publishers. Elsevier Science, 2003.
4. Z. Juhasz (Editor), P. Kacsuk (Editor), D. Kranzlmuller (Editor). "Distributed and Parallel Systems: Cluster and Grid Computing" (The International Series in Engineering and Computer Science). Springer; 1 edition (September 21, 2004).
5. Di Stefano M. "Distributed data management for Grid Computing". John Wiley & Sons Inc (29 Jul 2005).
6. Miller Michael Web Based applications that change the way you work online". Que Publishing. USA 2009.
7. Mc. Cool Michael "Programming models for scalable multicore programming" <http://www.hpcwire.com/features/17902939.html> (2007)
8. Lei Chai, Qi Gao, Dhabaleswar K. Panda. "Understanding the Impact of Multi-Core Architecture in Cluster Computing: A Case Study with Intel Dual-Core System". IEEE International Symposium on Cluster Computing and the Grid 2007 (CCGRID 2007), pp. 471-478 (May 2007).
9. Suresh Siddha, Venkatesh Pallipadi, Asit Mallick. "Process Scheduling Challenges in the Era of Multicore Processors" Intel Technology Journal, Vol. 11, Issue 04, November 2007.
10. Marek Olszewski, Jason Ansel, Saman Amarasinghe. "Kendo: Efficient Deterministic Multithreading in Software". Architectural Support for Programming Languages and Operating Systems (ASPLOS '09)
11. Lei Chai Qi Gao Dhabaleswar K. Panda. "Understanding the Impact of Multi-Core Architecture in Cluster Computing: A Case Study with Intel Dual-Core System" 7 th. IEEE International Symposium on Cluster Computing and the Grid. CCGRID 07 – Brasil Mayo 2007.
12. Thomas W. Burger. Intel Multi-Core Processors: Quick Reference Guide. http://cachewww.intel.com/cd/00/00/23/19/231912_231912.pdf
13. Foster I., Kesselman C., Kaufmann M. "The Grid 2: Blueprint for a New Computing Infrastructure". The Morgan Kaufmann Series in Computer Architecture and Design. 2 edition (November 18, 2003).
14. A. E. De Giusti. "Tutorial Grid Computing". June 2006.
15. Grid Computing and Distributed Systems (GRIDS) Laboratory - Department of Computer Science and Software Engineering (University of Melbourne). "Cluster and Grid Computing". 2007. <http://www.cs.mu.oz.au/678/>.
16. Berman F., Fox G., Hey A. "Grid Computing: Making The Global Infrastructure a Reality". John Wiley & Sons (April 8, 2003).
17. The Globus Alliance: <http://www.globus.org>
18. Grid Computing Infocentre: <http://www.gridcomputing.com/>
19. Bertogna M., Grosclaude E., Naiouf M., De Giusti A., Luque E. "Dynamic on Demand Virtual Clusters in Grids" 3rd Workshop on Virtualization in High-Performance Cluster and Grid Computing. VHPC 08 – España. Agosto 2008. (Springer Verlag en prensa).
20. Berman F., Fox G., Hey A. "Grid Computing: Making The Global Infrastructure a Reality". John Wiley & Sons (April 8, 2003).
21. Ghosh S. "Distributed System. An Algorithmic Approach". Chapman & Hall/CRC Computer and Information Science Series.

Procesamiento distribuido y paralelo. Fundamentos y aplicaciones.

Marcelo Naiouf, Armando E. De Giusti, Laura C. De Giusti, Franco Chichizola, Adrian Pousa, Waldo Hasperué, Victoria Sanz, Fabiana Leibovich

Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP
{mnaiouf, degiusti, ldgiusti, francoch, apousa, whasperue, vsanz, fleibovich}@lidi.info.unlp.edu.ar

CONTEXTO

La línea de Investigación que se presenta es parte del Proyecto “Algoritmos Distribuidos y Paralelos. Aplicación a Sistemas Inteligentes y Tratamiento Masivo de Datos” del Instituto de Investigación en Informática LIDI acreditado por la UNLP, y de proyectos apoyados por Cyted, CIC, IBM, Fundación YPF y Telefónica.

RESUMEN

El eje central de esta línea de I/D lo constituye el estudio de los temas de procesamiento paralelo y distribuido, tanto en lo referente a los fundamentos como a las aplicaciones. Esto incluye los problemas de software asociados con la construcción, evaluación y optimización de algoritmos paralelos y distribuidos sobre arquitecturas multiprocesador.

Los temas de interés abarcan: paralelización de algoritmos, paradigmas paralelos, métricas, escalabilidad, balance de carga, y modelos de computación paralela para la predicción y evaluación de performance sobre diferentes clases de arquitecturas de soporte. Tales arquitecturas pueden ser homogéneas o heterogéneas, como multicore, cluster, multicluster y grid.

Se trabaja en la concepción de aplicaciones paralelas numéricas y no numéricas sobre grandes volúmenes de datos que requieren cómputo intensivo, y en el desarrollo de laboratorios remotos para el acceso transparente a recursos de cómputo paralelo.

Palabras clave: *Sistemas paralelos. Algoritmos paralelos y distribuidos. Clusters. Multicluster. Grid. Balance de carga. Evaluación de performance.*

1. INTRODUCCION

El procesamiento paralelo y distribuido se ha convertido en un área de gran desarrollo actual dentro de la Ciencia de la Computación, produciendo profundas transformaciones en las líneas de I/D [1][2][3][4][5][8][9][10][11].

Interesa realizar investigación en la especificación, transformación, optimización y evaluación de algoritmos distribuidos y paralelos. Esto incluye el

diseño y desarrollo de sistemas paralelos, la transformación de algoritmos secuenciales en paralelos, y las métricas de evaluación de performance sobre distintas plataformas de soporte (hardware y software). En esta línea de I/D la mayor importancia está dada en los *algoritmos paralelos*, y en los métodos utilizados para su construcción y análisis [6][7][26].

Si bien utilizar múltiples procesadores para resolver problemas, en general de complejidad creciente y para obtener resultados en menor tiempo, resulta un concepto intuitivo, los fundamentos subyacentes a los mecanismos y soportes de paralelización presentan numerosas variantes. Pueden encontrarse diferentes formulaciones paralelas para un problema, y el rendimiento y la eficiencia de cada una dependerá del algoritmo y de la arquitectura utilizada.

1.1. Algoritmos y Sistemas Paralelos

La creación de algoritmos paralelos y distribuidos sobre arquitecturas multiprocesador, o la transformación de un algoritmo secuencial en paralelo, no es un proceso directo. El costo del paralelismo puede ser alto en términos del esfuerzo de programación, ya que debe pensarse en la aplicación de técnicas nuevas reescribiendo totalmente el código secuencial, y las técnicas de debugging y tuning de performance no se extienden de manera directa al mundo paralelo [8][9][10][11].

Un *sistema paralelo* (SP) es la combinación de un algoritmo paralelo y la máquina sobre la cual éste se ejecuta; ambos factores poseen numerosas variantes y de un adecuado “matching” entre ellos depende el éxito de la solución [16][17].

Las arquitecturas para procesamiento paralelo y distribuido han evolucionado, y en la actualidad las redes de computadoras constituyen una plataforma de cómputo paralelo muy utilizada por sus ventajas en términos de la relación costo/rendimiento. La noción de sistema distribuido como máquina paralela es común a las denominaciones redes de computadoras, NOW, redes SMP, clusters, multiclusters y grid. En estos casos, se deben identificar las capacidades de procesamiento, interconexión, sincronización y escalabilidad [13].

El concepto de multicluster es una generalización que permite que redes dedicadas a una aplicación paralela se interconecten y puedan cooperar en un algoritmo, compartiendo recursos e incrementando la potencia de cómputo. Esto implica clusters dedicados interconectados, a diferencia de *grid computing* en que cada procesador puede realizar otras tareas independientes del algoritmo, brindando alta disponibilidad de procesamiento y/o de almacenamiento [14][15][16][17].

La caracterización y estudio de rendimiento del sistema de comunicaciones es de particular interés para la predicción y optimización de performance de los algoritmos, así como la homogeneidad o heterogeneidad de los procesadores que componen la arquitectura [18].

Muchos de los problemas algorítmicos en arquitecturas multiprocesador se han visto fuertemente impactados por el surgimiento de las máquinas *multicore* (que integran dos o más núcleos computacionales dentro de un mismo chip). Esto obliga al estudio de performance en sistemas híbridos, en los que se tiene una combinación de memoria compartida y distribuida.

1.2. Métricas del paralelismo

La diversidad de opciones en los SP vuelve complejo el análisis de performance, ya que los ejes sobre los cuales pueden compararse dos sistemas son varios (relacionados con el algoritmo y/o la arquitectura).

Existe un gran número de métricas para evaluar el rendimiento de los SP. Las tradicionales y más conocidas son el tiempo de ejecución paralelo, el speedup y la eficiencia. Otras características de los SP pueden ser analizadas por medio del costo, overhead paralelo, grado de concurrencia, etc. [19]

La *escalabilidad* de un SP permite capturar las características de un algoritmo paralelo y de la arquitectura en la que se lo implementa. Permite testear la performance de un programa paralelo sobre pocos procesadores y predecir su performance en un número mayor, y caracterizar la cantidad de paralelismo inherente en un algoritmo. Una de las principales métricas para medir la escalabilidad es la isoeficiencia.

Al tratar con multiclusters y grid, los problemas clásicos que caracterizan el análisis de los algoritmos paralelos, reaparecen potenciados por las dificultades propias de la interconexión a través de una red que en general es no dedicada. Esto se torna más complejo aún si se considera que cada uno de los nodos puede ser en sí mismo una máquina multicore.

1.3 Balance de carga

El *balance de carga* consiste en, dado un conjunto de tareas que comprenden un algoritmo y un conjunto de procesadores donde ejecutarlas, encontrar el mapeo de tareas a procesadores que resulte en que cada una tenga una cantidad de trabajo que demande aproximadamente el mismo tiempo.

Un mapeo que balancea la carga de trabajo de los procesadores incrementa la eficiencia global y reduce el tiempo de ejecución. Este objetivo es particularmente complejo si los procesadores (y las comunicaciones entre ellos) son heterogéneos, y deben tenerse en cuenta las distintas velocidades.

El problema general de asignación es *NP-completo* para un sistema con n procesadores, y por lo tanto la tarea de encontrar una asignación de costo mínimo es computacionalmente intratable salvo para sistemas muy pequeños (métodos *óptimos*). Por esto pueden utilizarse enfoques que brindan soluciones subóptimas aceptables, como relajación, desarrollo de soluciones para casos particulares, optimización enumerativa, u optimización aproximada (métodos *heurísticos*) [20].

Si el tiempo de cómputo de una tarea dada puede determinarse “a priori”, se puede realizar el mapeo antes de comenzar la computación (balance de carga *estático*). En muchas aplicaciones la carga de trabajo para una tarea particular puede modificarse en el curso del cómputo, y no puede estimarse de antemano; en estos casos el mapeo debe cambiarse durante el cómputo (balance de carga *dinámico*), realizando etapas de balanceo durante la ejecución de la aplicación.

El balance estático, en general, es de menor complejidad que el dinámico, pero también menos versátil y escalable. Los métodos dinámicos requieren alguna forma de mantener una visión global y algún mecanismo de análisis para la migración de procesos y/o datos, lo que agrega overhead de cómputo y comunicaciones.

No puede establecerse un método efectivo y que sea eficiente en *todos* los casos. Siempre la elección depende de la aplicación y la plataforma de soporte, y en muchos casos es necesario adaptar o combinar métodos existentes para lograr buena performance [24][27].

1.4 Modelos de representación, predicción y análisis de performance

Es importante referirse a un algoritmo paralelo mencionando el modelo de computación para el que

fue diseñado. Uno de los objetivos en la definición de un modelo de computación es la posibilidad de *predicción de performance* que brinde el mismo, teniendo en cuenta conceptos tales como comunicación, sincronización y arquitectura física. Estos factores impiden que alguno de los modelos existentes pueda ser usado para *todas* las máquinas paralelas.

Respecto de la representación de las aplicaciones paralelas en arquitecturas distribuidas, existen diferentes modelos basados en grafos para caracterizar el comportamiento de las mismas [21]. Entre los modelos se pueden mencionar el modelo TIG (Grafo de Interacción de Tareas), TPG (Grafo de Precedencia de Tareas) y TTIG (Grafo de Interacción Temporal de Tareas) [22]. Sin embargo, estos modelos consideran que la arquitectura es homogénea, situación que en general no se da en cluster, multicluster y grid, lo que hace necesaria la investigación en esta área. El desarrollo de nuevos modelos de predicción y análisis de performance para estas arquitecturas requiere caracterizar el contexto de comunicaciones entre los procesadores y la asociación entre los algoritmos de aplicación, el paradigma de cómputo paralelo elegido y la arquitectura de soporte.

1.5 Evaluación de performance. Aplicaciones

Es de interés la evaluación de performance de distintas clases de aplicaciones sobre las arquitecturas disponibles. Muchos sistemas paralelos no alcanzan su capacidad teórica, y las causas de esta degradación son muchas y no siempre fáciles de determinar. El análisis permite estudiar el impacto que tienen algunos de estos factores sobre las implementaciones, y adecuar las métricas a las mismas. Interesa estudiar la influencia de las estrategias de distribución de procesos y datos, y la carga (estática o dinámica) asignada a cada procesador sobre el speedup, la eficiencia y la escalabilidad.

Desde el punto de vista de la relación costo/rendimiento, el cómputo paralelo sobre arquitecturas distribuidas ha ganado rápidamente espacio en el campo de las aplicaciones reales dado el bajo costo de los procesadores y estaciones de trabajo estándares junto con su alto rendimiento. Si bien existe una amplia gama de posibilidades estudiadas y publicaciones con todo tipo de aplicaciones resueltas, se acepta que es necesario continuar con la investigación en esta área [13][24].

Entre las aplicaciones de interés se encuentran las numéricas y no numéricas, el tratamiento de imágenes y video, reconocimiento de patrones en secuencias de ADN, bases de datos distribuidas, sistemas inteligentes, data mining, etc.

Por otro lado, interesa el desarrollo de laboratorios remotos para el acceso a recursos de cómputo paralelo. Esto implica software de administración de recursos físicos, comunicaciones y software disponible en clusters, multiclusters y grid, con el objetivo de permitir acceso transparente.

2. LINEAS DE INVESTIGACION y DESARROLLO

- Paralelización de algoritmos secuenciales. Diseño de algoritmos paralelos. Optimización de algoritmos.
- Lenguajes y bibliotecas para procesamiento paralelo y distribuido. Comparación de performance.
- Multithreading en multicore.
- Sistemas paralelos como combinación de software y arquitectura.
- Modelos y paradigmas de computación paralela. Combinación de pasaje de mensajes y memoria compartida en cluster de multicore.
- Modelos de representación y predicción de performance de algoritmos paralelos. Mapping y scheduling de aplicaciones paralelas.
- Métricas del paralelismo. Speedup, eficiencia, rendimiento, isoeficiencia, granularidad, superlinealidad. Escalabilidad de algoritmos paralelos en arquitecturas distribuidas.
- Balance de carga estático y dinámico. Técnicas de balanceo de carga.
- Análisis (teórico y práctico) de los problemas de migración y asignación óptima de procesos y datos a procesadores. Migración dinámica.
- Implementación de soluciones sobre diferentes modelos de arquitectura homogéneas y heterogéneas (multicores, clusters, multiclusters y grid). Ajuste del modelo de software al modelo de hardware, a fin de optimizar el sistema paralelo. Evaluación de performance.
- Laboratorios remotos para el acceso transparente a recursos de cómputo paralelo

3. RESULTADOS OBTENIDOS/ESPERADOS

- Formar recursos humanos en los temas del Subproyecto, incluyendo tesinas de grado y tesis de postgrado.
- Desarrollar y optimizar algoritmos paralelos sobre los diferentes modelos de arquitectura multiprocesador.
- Realizar la migración de soluciones paralelas en cluster a multicluster y grid.
- Evaluar la eficiencia, rendimiento, speedup y escalabilidad de las soluciones propuestas.

- Mejorar y adecuar las técnicas disponibles para el balance de carga (estático y dinámico).
- Estudiar y proponer optimizaciones para los modelos de predicción/evaluación de performance con diferentes paradigmas de interacción entre procesos, en esquemas multicluster y grid.

En este marco, pueden mencionarse los siguientes resultados:

- Se han utilizado diferentes tipos de arquitecturas homogéneas o heterogéneas:
 - Clusters comunicados en la misma LAN (cada cluster con un nodo principal que maneja las comunicaciones hacia los otros)
 - Clusters en redes diferentes con una conexión directa entre los nodos principales de cada uno vía fibra óptica.
 - Clusters en redes distintas conectadas por Internet en una WAN compartida y por Internet 2 con ancho de banda asegurado.
 - Infraestructura grid experimental en el marco del despliegue del Proyecto CyTEDGrid.
 - Máquinas multicore y cluster de multicore.
- En cuanto a modelos de representación y predicción de performance:
 - Desarrollo de dos modelos (TTIGHa y MPAHA) para representar aplicaciones paralelas de manera más realista considerando la heterogeneidad de los procesadores y de la red de interconexión [23].
 - Desarrollo de tres algoritmos de scheduling (MATEHA, MATEHAIB y AMTHA) [23], que permiten obtener una mejor distribución de procesos en procesadores.
 - Extensión del algoritmo de scheduling AMTHA para realizar la asignación de múltiples aplicaciones paralelas a una misma arquitectura. Dió origen a dos algoritmos de scheduling (SchCE_AMTHA y SchCD_AMTHA).
 - Se está trabajando sobre las modificaciones necesarias a los modelos y algoritmos de scheduling para adecuar sus usos sobre clusters de multicore y grid [27].
- Respecto de los algoritmos implementados, básicamente se trabajó con soluciones paralelas previamente tratadas en clusters:
 - ***N-Queens***: consiste en ubicar N reinas en un tablero de $N \times N$ sin generar un ataque (esto ocurre cuando dos reinas están en la misma fila, columna o diagonal). Para

resolver este problema se utilizó un modelo asincrónico, en el cual un porcentaje del trabajo es distribuido inicialmente considerando la potencia de cómputo de cada procesador/cluster. Cuando los procesadores terminan su trabajo le piden más al master. Se estudiaron speedup, eficiencia y balance de carga [24]. Se migró la aplicación a un grid con el objetivo de estudiar el overhead generado por el middleware incorporado. Y se estudia su migración a cluster de multicore.

- ***N-Puzzle***: es un problema de optimización discreta en el cual se debe llegar a un tablero “final” a partir de uno “inicial”. Para esto deben realizarse intercambios entre una pieza y el lugar libre en el tablero (hueco) [25]. Se utilizó una variante de la heurística clásica de predicción de trabajo a realizar para llegar a una solución (tablero “final”) y se demostró que su empleo mejora notoriamente el tiempo del algoritmo secuencial A^* . Posteriormente se realizó una solución paralela sobre una arquitectura distribuida para analizar el speedup en función del número de procesadores, la eficiencia y la superlinealidad al escalar el problema. Este es un problema de interés por su aplicación principalmente en el campo de la robótica. Se está trabajando sobre la migración a cluster de multicore y la generalización del problema a multiobjetivos.
- ***Análisis de Secuencias de ADN***: el análisis de secuencias de ADN tiene múltiples aplicaciones, una de ellas es la búsqueda de similitud entre una secuencia *test* y las almacenadas en grandes bases de datos. El gran tamaño de las bases de datos, y la complejidad computacional para comparar dos secuencias (Smith-Waterman) hacen necesaria la paralelización del algoritmo [12]. Se diseñaron dos algoritmos de acuerdo a la centralización y/o replicación de la Base de Datos. Estos algoritmos se ejecutaron sobre arquitecturas de tipo cluster y grid, estudiando el speedup y la eficiencia alcanzable, la escalabilidad de las soluciones y el overhead introducido en grid por las comunicaciones y el middleware requerido. Se pretende extender las experiencias sobre GRID utilizando procesadores geográficamente distribuidos sobre diferentes redes ubicadas en Argentina, Brasil y España.
- ***Extracción de conocimiento en grandes bases de datos utilizando estrategias***

adaptativas: la tecnología actual posibilita el almacenamiento de enormes volúmenes de información. Es importante encontrar métodos y técnicas de minería de datos que sean capaces de generar conocimiento útil, produciendo resultados que sean de provecho al usuario final. Para esto, se estudian métodos de clustering y clasificación de patrones para lograr asociar respuestas dinámicas con los datos de entrada obtenidos. Dado el gran volumen de información a procesar (que puede estar físicamente distribuida), se requiere analizar la arquitectura y los paradigmas de programación paralela utilizables de modo de minimizar el tiempo de cálculo del proceso adaptativo.

- En cuanto a los laboratorios para acceso remoto a recursos de cómputo paralelo:
 - Se ha desarrollado una aplicación que permite el acceso a los clusters de la Facultad de Informática de manera remota vía WEB (para investigadores y alumnos). Esto implica una capa de software de administración de los recursos y acceso transparente.
 - Se está trabajando en la generalización de la herramienta a clusters que no se encuentran en la misma red

4. FORMACION DE RECURSOS HUMANOS

Existe cooperación con grupos de otras Universidades del país y del exterior.

Dentro de la temática de la línea de I/D se ha concluido una tesis doctoral en 2008 y se espera concluir este año otras 3 tesis que se encuentran en curso y 2 tesis de maestría. También se concluyeron dos Tesinas de Grado de Licenciatura y se espera finalizar otra.

Además, se participó en la definición y el dictado de la carrera de Especialización en Cómputo de Altas Prestaciones y Tecnología Grid de la Facultad de Informática de la UNLP.

5. BIBLIOGRAFIA

[1] Basney J., Livny M.. "Deploying a High Throughput Computing Cluster". Buyya R. Ed., High Performance Cluster Computing: Architectures and Systems, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, USA, pp. 116-134, 1999.

[2] Vijay K. Garg. "Elements of Distributed Computing". Wiley-IEEE Press, 2002.

[3] Liu M. L. "Distributed Computing: Principles and Applications". Addison Wesley; 1st edition, 2003.

[4] Grama A., Gupta A., Karypis G., Kumar V. "Introduction to Parallel Computing", Pearson Addison Wesley, 2nd Edition, 2003.

[5] Attiya H., Welch J. "Distributed Computing: Fundamentals, Simulations, and Advanced Topics", (Wiley Series on Parallel and Distributed Computing). Wiley-Interscience; 2nd edition, 2004.

[6] Berman K. A., Pau J. L. "Algorithms: Sequential, Parallel, and Distributed". Course Technology; 1st edition, 2004.

[7] Wilkinson B., Allen M. "Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers (2nd Edition)". Prentice Hall, 2004.

[8] Akl S. "Parallel Computing. Models and Methods". Prentice Hall. 1997.

[9] Andrews G. "Foundations of Multithreaded, Parallel and Distributed Programming", Addison Wesley, 2000

[10] Leopold C. "Parallel and Distributed Computing. A survey of Models, Paradigms, and Approaches", Wiley Series on Parallel and Distributed Computing. Albert Zomaya Series Editor, 2001

[11] Jordan H. F., Alaghband G., Jordan H. E. "Fundamentals of Parallel Computing", Prentice Hall, 2002"

[12] Chichizola F., Naiouf R. M., De Giusti L. C., Rodríguez I., De Giusti A. E. "Parallel Processing ADN Sequences on Multicluter and Grid Architectures. Software Overhead". Proceedings del IX Workshop de Procesamiento Distribuido y Paralelo. La Rioja, Argentina. ISBN: 978-987-24611-0-2. Octubre de 2008.

[13] Tinetti F. "Cómputo Paralelo en Redes Locales de Computadoras". Tesis Doctoral. Universidad Autónoma de Valencia. 2004.

[14] IEEE Task Force on Cluster Computing (www.ieeetfcc.org)

[15] Foster I., Kesselman C. "The Grid 2: Blueprint for a New Computing Infrastructure". (The Morgan Kaufmann Series in Computer Architecture and Design). Morgan Kaufmann; 2nd edition, 2003.

[16] Juhasz Z., Kacsuk P., Kranzlmuller D. (Eds), "Distributed and Parallel Systems: Cluster and Grid Computing". (The Intl Series in Engineering and Computer Science). Springer; 1st edition, 2004

[17] Silva V. "Grid Computing For Developers" (Programming Series). Charles River Media; 1st edition, 2005.

[18] Kurmann C., Rauch F., Stricker M. "Cost/Performance Tradeoffs in Network Interconnects for Clusters of Commodity PCs". Technical Report 391, Swiss Federal Institute of Technology Zurich, Institute for Computer Systems, January 2003

[19] Sun X-H . "Scalability versus Execution Time in Scalable Systems". Journal of Parallel and Distributed Computing, Number 12, 2002, pp 173-192

[20] Naiouf R. M. "Procesamiento paralelo. Balance dinámico de carga en algoritmos de sorting". Tesis doctoral. Universidad Nacional de La Plata, 2004.

[21] Kalinov A., Klimov S.. "Optimal Mapping of a Parallel Application Processes onto Heterogeneous Platform". Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05), April 2005

[22] Roig C., Ripoll A. Guirado F. "A New Task Graph Model for Mapping Message Passing Applications". IEEE Transactions on Parallel and Distributed Systems, vol 18 (12), pp.740-1753. Diciembre 2007.

- [23] De Giusti L. "Mapping sobre Arquitecturas Heterogéneas". Tesis Doctoral, Universidad Nacional de La Plata (2008).
- [24] Naiouf R. M., De Giusti L. C., Chichizola F., De Giusti A. E. "Dynamic Load Balancing on Non-homogeneous Clusters". G.Min et al. (Eds.): ISPA 2006 Ws, LNCS 4331, pags. 65-73, 2006. Springer – Verlag, Berlin Heidelberg 2006.
- [25] Sanz V., De Giusti A. E., Chichizola F., Naiouf R. M., De Giusti L. C. "Parallel Processing Puzzle N^2-1 on Cluster Architectures. Performance Analysis". 30th International Conference on Information Technology Interfaces, 2008. ITI 2008. 23-26 Page(s):879 – 884. June 2008.
- [26] Qiu X., Fox G. G., Yuan H., Bae S., Chrysanthakopoulos G., Nielsen H. F. "Performance of Multicore Systems on Parallel Data Clustering with Deterministic Annealing". LNCS 4331, pags. 407-416. ISBN 978-3-540-69383-3. Springer Berlin / Heidelberg 2008.
- [27] Yi Liu, Xin Zhang, He Li, Depei Qian. "Allocating Tasks in Multi-core Processor based Parallel Systems". Proceedings of the 2007 IFIP International Conference on Network and Parallel Computing – Workshops. IEEE Computer Society. 2007.

Metodología de Sincronización de Relojes para Instrumentación

Fernando L. Romero, Fernando G. Tinetti¹
Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP
fromero@lidi.info.unlp.edu.ar, fernando@info.unlp.edu.ar

CONTEXTO

Esta línea de Investigación forma parte de dos de los Subproyectos dentro del Proyecto “Sistemas Distribuidos y Paralelos” acreditado por la UNLP y de proyectos específicos apoyados por CyTED, CICPBA, Agencia Nacional de Promoción Científica y Tecnológica e IBM.

RESUMEN

En el presente trabajo se expone la línea de investigación sobre sincronización de relojes en ambientes distribuidos. Como parte de este trabajo se implementaron herramientas que permiten realizar evaluación de rendimiento a través de la instrumentación de código. Se desarrollaron estas herramientas en virtud de las falencias detectadas en las existentes. Si bien los experimentos se llevaron a cabo en ambientes de clusters, se estima en el futuro extender la herramienta a otros ambientes distribuidos utilizados para cómputo paralelo. Asimismo, las máquinas de dichos clusters son mono procesadores, pero se planifica investigar el problema aplicado a multiprocesadores y múltiples núcleos.

Keywords: *Sincronización de Procesos, Relojes Distribuidos, Rendimiento e Instrumentación, Sistemas Paralelos y Distribuidos, Paralelismo en Clusters e Intercluster, Sincronización Interna y Externa .*

1. INTRODUCCION

Los sistemas de cómputo y de comunicación que conforman un sistema distribuido disponen de relojes. En caso de querer realizar mediciones de eventos que se produzcan en diferentes nodos de dicho sistema, se hace necesaria la sincronización de estos relojes. En el caso de medir tiempos de comunicaciones, los requerimientos de exactitud con que se realizan esta sincronización y las mediciones han ido creciendo desde el segundo, milisegundo [6] [7], hasta llegar a exigir microsegundos [10] [9] [8]. Los tópicos que se estudian del problema comprenden:

- **Resolución**
- **Precisión**
- Tolerancia a fallas
- **Sobrecarga del sistema**
- **Exactitud**
- Fiabilidad
- Interoperabilidad
- Seguridad

La tarea de sincronizar relojes implica un compromiso entre estas variables, lo cual hace que por optimizar un aspecto, se degrada otro. Por ejemplo, si se incluye seguridad habría que encriptar los paquetes de comunicación de referencias horarias, lo cual degrada el aspecto de la sobrecarga, tanto de comunicaciones como el procesamiento de los mismos. Así es que en la lista anterior se resaltan

¹ Investigador Adjunto, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

en negrita los aspectos que se privilegian en este trabajo, esenciales al problema a resolver, la instrumentación. Por ello, se plantea la sincronización de relojes [3] [7] [6] pero con restricciones en los aspectos a resolver, de tal manera que cumplan los siguientes requerimientos:

- Pueda usarse como herramienta de instrumentación para programas paralelos
- Uso inicial en cluster de PC's, y posibilidad de ser extendido a clusters en general y a plataformas distribuidas aún más generales
- Sea de alta resolución, es decir, para medir tiempos cortos, del orden de microsegundos.
- No altere el funcionamiento de la aplicación a probar, o que la alteración sea mínima y conocida por la aplicación.
- Utilice en forma predecible la red de interconexión, determinando desde la aplicación los intervalos de tiempo en los cuales se usará la red, y así *desacoplar* el uso de la red entre tiempos para sincronización y tiempos para ejecución de programas paralelos.
- Se conozca el tipo (o al menos magnitud) de error con que se sincroniza.
- Sin incluir hardware adicional al del sistema. Para comunicaciones se usará la red entre computadoras que ya existe, y como sistema de medición el de cada computadora.

Como producto se obtuvo la biblioteca *st*, que realiza las tareas de sincronización dentro de los requerimientos previstos. También se obtuvo la biblioteca *timings*, para registro de tiempos, la cual provee la resolución adecuada, con bajo uso de recursos, sin usar entrada/salida ni llamadas al sistema, y con el menor uso posible de ciclos de CPU.

La exactitud fue uno de los tópicos priorizados, a fin de lograr cotas de error aceptables que hagan posible estimaciones de rendimiento en ambientes de red local de aplicaciones paralelas. Se debe tener en cuenta que en una red local Ethernet los tiempos de comunicación son del orden de decenas de microsegundos, por lo que determinaciones de tiempo transcurrido entre la partida y el arribo de un mensaje requieren sincronizaciones con diferencias de microsegundos en las máquinas que intercambian mensajes, para poder contar con una medida con un error aceptable.

En todas las circunstancias es deseable que el registro de tiempos implique la menor carga posible de procesamiento, como también conocer la magnitud del error que se comete en la medición. En el caso de mediciones de rendimiento, sería deseable que la tarea de sincronización se lleve a cabo fuera del tiempo en que se ejecuta el programa que se está monitorizando. Por otro lado, la cantidad de máquinas a sincronizar puede llevar a que el tiempo que se insume en la sincronización sea excesivo. Por ello se han llevado a cabo experimentos a fin de evaluar la escalabilidad de la herramienta en términos prácticos.

2. LINEAS DE INVESTIGACION Y DESARROLLO

La evaluación de rendimiento introduciendo instrumentación, exige que cada computadora cuente con un reloj físico, a partir del cuál se derivan los relojes lógicos que son los que se sincronizan [4]. De las limitaciones de estos relojes dependerá lo que se pueda lograr, por lo que se realizaron experimentos a fin de caracterizar estos relojes en cuanto a exactitud, estabilidad, confiabilidad, resolución y demás atributos.

Los relojes lógicos deben sincronizarse en una *hora inicial* a partir de la cual se empiezan a actualizar y en *frecuencia* de actualización de la misma. Las referencias son intercambiadas entre las máquinas a través de la red existente. Esto adiciona una demora que debe adicionarse a la referencia. Debido a la varianza en los tiempos de comunicaciones, es difícil determinar este valor más allá de cierta exactitud. Para la medición de este valor se puede recurrir a la estadística de los valores de ida-vuelta (*round trip time*) de un mensaje y asumir que solo serán válidos los valores de referencia transmitidos en un tiempo de transmisión igual a la *moda* de los tiempos de transmisión. Ya que la información es provista y centralizada en el servidor, al aumentar la cantidad de clientes, el servidor se transforma en un cuello de botella. De tal manera que la escalabilidad del modelo debió probarse.

3. RESULTADOS OBTENIDOS/ESPERADOS

A partir del estudio de los algoritmos básicos e implementaciones existentes [1] [2] [6], se realizaron un conjunto de experimentos a fin de poder compararlas y ver sus limitaciones [9] [10] [11]. Como conclusión de dichos experimentos se desarrolló una biblioteca *timings* para mediciones de tiempo a nivel local y una biblioteca *synchro*, que funciona en forma distribuida y permite sincronizar en hora y frecuencia los relojes de *timings*. Estas bibliotecas además proveen una caracterización del error, dada por sus diversos componentes:

- Error en la estimación del tiempo de comunicación entre máquinas de las referencias.
- Error debido a la precisión del reloj local de la biblioteca desarrollada (*timings*).
- Error debido a latencia propia del sistema operativo.

Como extensiones futuras, siempre es deseable la sincronización externa de los relojes [3]. Este paso está muy ligado también a la posibilidad de utilizar más de un cluster de computadoras para cómputo paralelo y en este contexto la sincronización de los relojes va más allá del análisis de rendimiento con el objetivo de optimizarlo.

Una de los experimentos fue medir la escalabilidad del modelo [10] [11], midiendo el tiempo en microsegundos que lleva sincronizar los relojes de 2 a 16 computadoras. Los resultados resumidos se muestran en la siguiente tabla:

	2	4	6	8	10	12	14	16
Mínimo	37972	39902	40819	38596	41910	37522	38525	41774
Máximo	43362	46933	52908	48233	59397	53436	56029	52791
Promedio	40475	42444	45314	43585	47353	45433	47645	48312

Tabla I: Tiempos de ejecución para 20 corridas de Synchro en 2 a 16 máquinas.

En forma gráfica se puede apreciar la evolución de estos tiempos en la Fig. I, donde, además, se muestran los valores intermedios que se omitieron en la tabla I.

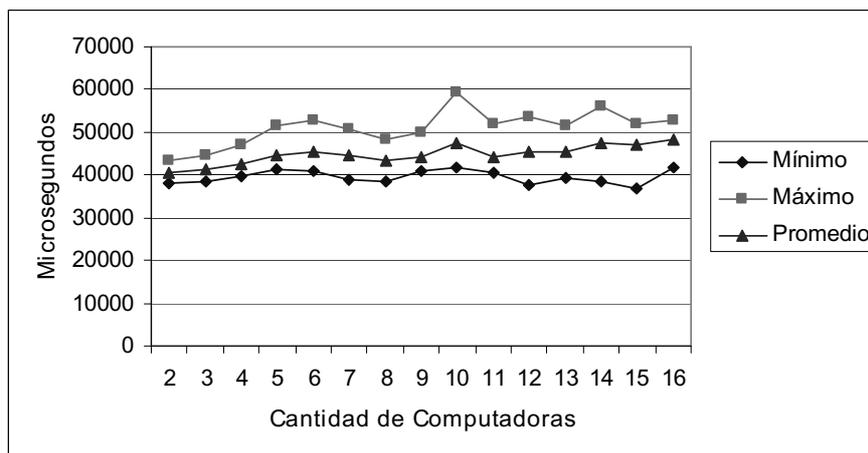


Figura I: Tiempos de ejecución para 20 corridas de Synchro en 2 a 16 máquinas.

Estos resultados permiten concluir que la escalabilidad de este algoritmo, a pesar de basarse en un modelo cliente-servidor, no representa mayores inconvenientes en los tamaños típicos de clusters. La sincronización se lleva a cabo entre el servidor y cada cliente, en forma secuencial. Como se desprende de restar al tiempo de sincronizar 16 el de sincronizar 2, y dividir dicho tiempo por 14 máquinas, corresponderían 560 microsegundos por máquina agregada a la sincronización. Los tiempos máximos medidos son menores a 60 milisegundos en total. Teniendo en cuenta la cantidad de computadoras de un cluster, son valores despreciables, sobre todo si se piensa que son tiempos

previos al proceso de medición, y absolutamente aislados del mismo.

La uniformidad y linealidad de los tiempos, permite estimar extrapolaciones. Sería posible estimar en $560 \times 100 = 56000$, o sea, 56 milisegundos el tiempo agregado a los 40 milisegundos iniciales para sincronizar un cluster de 100 computadoras, llevando el total del tiempo de sincronización a menos de una décima de segundo para un cluster de 100 computadoras. En experimentos posteriores, se comprobó que los aproximadamente 40 milisegundos iniciales se deben a falta de sincronización entre clientes y el servidor. Al sincronizarse los clientes con un *delay*, desaparece dicho tiempo inicial, ya que solicitan sincronizarse recién cuando el servidor ya terminó de inicializarse.

Con respecto a sobrecarga de tiempo de medición durante la ejecución de un programa paralelo, cada consulta de tiempo implica un tiempo de entre 230 y 340 ciclos de reloj de CPU [7], con lo cual, en una computadora con velocidad de reloj 2Ghz, estaríamos en el orden de 1 a 2 décimas de microsegundo por cada medición, contra 615 ciclos para `gettimeofday()`, que es una llamada al SO, opcional al método utilizado por la biblioteca *timings*.

Otro experimento que se llevó a cabo, es la utilización de una única referencia, a cargo del servidor, para determinar la constante de *sincronización de frecuencia* MHz. La tabla II muestra los valores que se obtienen al sincronizar relojes con cálculo de MHz totalmente independientes (con los relojes locales) y con cálculo de MHz con referencia única.

Cant. Máq.	Ref. única		Ref. local	
	Mín.	Máx.	Mín.	Máx.
	Us	us	Us	us
2	-1	3	42	103
4	-3	3	55	98
8	-3	4	56	325
16	-4	3	43	1109

Tabla II: Errores de Sincronización por cálculo de MHz.

En el cálculo de MHz con referencia única (mostrado en la Tabla II) una sola computadora proporciona las referencias de tiempo para el cálculo de MHz. El experimento se llevó a cabo sincronizando desde 2 hasta 17 máquinas, y luego se midió la diferencia 1 segundo después. En todos los casos, son valores mínimos y máximos luego de 50 corridas.

Se pueden comparar estos valores con los obtenidos en NTP. La tabla III muestra dicha comparación. Los datos evitan cualquier comentario sobre la superioridad de la biblioteca *st* respecto de NTP [6] [7] [8] para sincronizar la hora en un cluster, sobre todo, teniendo en cuenta que la experiencia se realizó sobre el mismo hardware:

	NTP		Biblioteca st	
	Máximo	Estimado	Máximo	Estimado
Min	239673	31	32	1
Max	989368	2434	33	4

Tabla III: Comparación de error en sincronización de NTP contra la Biblioteca st.

También se realizaron pruebas tendientes a verificar la estabilidad de frecuencia [4] de los relojes de cuarzo en el tiempo similares a las reportadas en [5]. El experimento requeriría de una referencia de tiempo perfecta o al menos con la menor deriva de frecuencia posible, tal como un reloj atómico o la hora UTC. Debido a que no se cuenta con estos elementos, se toma como referencia el reloj de cuarzo del *timer* de una PC, y con el se miden las variaciones del reloj TSC (*Time Stamp Counter*). Si bien no se puede medir el error en términos absolutos, se puede determinar cuánto varía la diferencia de deriva. Dado que en la línea de investigación que se sigue no se requiere de referencias externas, sino solo que las referencias estén ajustadas a la escala de tiempos que

determine el servidor, es válido medir esta variación de frecuencia y pensar que el error será proporcional a dicha diferencia. Los resultados obtenidos son similares a los de [5], de menos de 10 ppm (partes por millón) en un período de varias horas.

Para la definición de futuras extensiones, se tienen planificados algunos experimentos para estimar:

- La mejora proporcionada por la comunicación de la referencia de tiempo utilizando comunicación *broadcast*.
- Más exhaustivamente la estabilidad de los relojes de las máquinas y su deriva, a fin de determinar la relación del intervalo entre sincronizaciones con el error de sincronización esperado.
- La utilidad de un servidor de hora con sistema operativo de tiempo real para disminuir la latencia impuesta por el SO al menos del lado del servidor.

4. FORMACION DE RECURSOS HUMANOS

En esta línea de I/D existe cooperación a nivel nacional e internacional. Se ha completado una tesis de maestría y está abierta la posibilidad para varias Tesinas de Grado de Licenciatura.

5. BIBLIOGRAFIA

- [1] G. Coulouris, Dollimore J., Kindberg T., *Sistemas Distribuidos. Conceptos y Diseño*, 3ª edición. Pearson Educación, 2001, ISBN: 8478290494.
- [2] F. Cristian. "Probabilistic Clock Synchronization", *Distributed Computing*, 3: 146–158, 1989.
- [3] Cristian F., Fetzer C., "The Time Asynchronous Distributed System Model", *IEEE Transactions on Parallel Systems*, June 1999, pp. 603618.
- [4] X. Luo, "TSC-I2: A Lightweight Implementation for Precision-Augmented Timekeeping", reporte técnico de University of Illinois at Chicago, http://tsc-xluo.sourceforge.net/TSC-I2_Tech_Report.pdf
- [5] D. L. Mills, "Measured performance of the Network Time Protocol in the Internet System". *ACM Computer Communication Review* 20, Jan. 1990. pp. 6575.
- [6] D. L. Mills, "A Brief History of NTP Time: Confessions of an Internet Timekeeper". *ACM Computer Communications Review* 33, 2 (April 2003), pp 922.
- [7] Mills D.L., "Network Time Protocol (Version 3) specification, implementation and analysis", DARPA Networking Group Report RFC1305, University of Delaware, March 1992.
- [8] Mills D. L., Kamp P.H., "The Nanokernel", Proc. Precision Time and Time Interval (PTTI) Applications and Planning Meeting (Reston VA, November 2000).
- [9] F. L. Romero, W. Aróztegui, F. G. Tinetti, "Sincronización de Relojes en Ambientes Distribuidos", XII Congreso Argentino de Ciencias de la Computación (XII CACIC) Octubre 2006.
- [10] F. L. Romero, Armando E. De Giusti, Fernando G. Tinetti. "Sincronización de Relojes para Evaluación de Rendimiento: Experiencias en un Cluster Utilizado para Cómputo Numérico". XIV Congreso Argentino de Ciencias de la Computación (XIV CACIC) Octubre 2008
- [11] Fernando G. Tinetti, Fernando L. Romero, Armando E. De Giusti "Clock Synchronization in Clusters for Performance Evaluation: Numeric/Scientific Computing". World Congress on Computer Science and Information Engineering (CSIE 2009), Abril 2009.

SISTEMAS DE CÓMPUTO DE ALTAS PRESTACIONES CON ALTA DISPONIBILIDAD: EVALUACIÓN DE LA PERFORMANCE EN DIFERENTES CONFIGURACIONES

Cecilia M. Lasserre¹, Emilio Luque Fadón², Dolores I. Rexachs del Rosario², Nilda M. Pérez Otero¹, Sandra Adriana Mendez¹, C. Marcelo Pérez Ibarra¹, Héctor P. Liberatori¹, Luis M. Valdiviezo¹, Fernando E. Flores¹, Rodrigo Cachambe¹

¹Facultad de Ingeniería – Universidad Nacional de Jujuy – Argentina

{lasserre, nilperez, smendez, cmperezi, hliberatori, mlvaldiviezo, efloresc, rcachambe}@fi.unju.edu.ar

²Departamento de Arquitectura de Computadores y Sistemas Operativos (DACSO-CAOS) – Universidad Autónoma de Barcelona – España

{dolores.rexachs, emilio.luque}@uab.es

CONTEXTO

El presente proyecto pertenece al GIS (Grupo de Ingeniería del Software) de la Facultad de Ingeniería de la UNJu y pertenece al Programa 08/D058. Se encuadra en el campo del Cómputo de Altas Prestaciones, específicamente en Tolerancia a Fallos.

RESUMEN

La Computación de Altas Prestaciones a través de clusters de computadores basados en Workstation y redes convencionales, posibilitó la construcción y uso de computadores paralelos. En las arquitecturas paralelas el objetivo principal es el aumento de prestaciones, utilizando el potencial ofrecido por gran número de procesadores.

La construcción de un cluster tiene tres retos: Alto Rendimiento, Alta Disponibilidad y Alta Productividad. Manejar eficientemente un número elevado de computadores en ambientes heterogéneos no es trivial, por ello requiere un cuidadoso diseño de la arquitectura y funcionalidad. Respecto a la alta disponibilidad, debemos considerar la probabilidad de los fallos o desconexión de nodos, reduciendo el tiempo medio entre fallos del computador paralelo como un todo.

Este proyecto pretende definir la configuración adecuada de tolerancia a fallos para diferentes tipos de aplicaciones, teniendo en cuenta los requerimientos de rendimiento y prestaciones del usuario, definir y validar un modelo genérico de aplicación-prestación-tolerancia a fallos.

Palabras clave: Computación de Altas Prestaciones, HPC, Alto Rendimiento, Alta Disponibilidad, Alta Productividad, Tolerancia a Fallos

1. INTRODUCCION

1.1 Clusters

El número de computadores que se interconectan en red en la actualidad está creciendo, por otro lado, el software que se desarrolla cada vez exige más y más recursos de los computadores, como procesadores, memoria, almacenamiento,... [1]. La llegada de la Computación de Altas Prestaciones (HPC - *High Performance Computing*) a través de *clusters* de

computadores basados en *Workstation* (WS) y redes convencionales (local o remota) o COTS (*Commodity Off The Shelf*), ha posibilitado que muchas instituciones y organizaciones vean viable la construcción y el uso de computadores paralelos. En las llamadas arquitecturas paralelas el objetivo principal es el aumento de la capacidad de procesamiento, utilizando el potencial ofrecido por un gran número de procesadores (alto aumento del *speedup* y alta disponibilidad) [2].

La construcción de un cluster tiene tres retos: Alto Rendimiento (*High Performance: HP*), Alta Disponibilidad (*High Availability: HA*) y Alta Productividad (*High Throughput: HT*) [1] [3]. Manejar eficientemente un número elevado de computadores en ambientes heterogéneos (heterogeneidad de nodos de cómputo, de redes, etc.) no es trivial, por ello requiere un cuidadoso diseño de la arquitectura y su funcionalidad. Respecto a la alta disponibilidad, cuando se considera un elevado número de nodos funcionando durante un tiempo elevado, debemos considerar que los fallos en los nodos o desconexión son eventos probables, reduciendo el MTBF (*Mean Time Between Failures* - tiempo medio entre fallos) del computador paralelo como un todo [4].

Los clusters se construyen a partir de varios nodos procesadores independientes conectados por alguna tecnología de red. Esos sistemas se diferencian de las máquinas masivamente paralelas por el débil acoplamiento entre sus nodos, o sea, los elementos del cluster no tienen acceso a una memoria común compartida. Otra característica que marca los clusters es la inexistencia de un reloj común que pueda ser utilizado para ordenar los eventos. Toda la comunicación entre los nodos se da a través de paso de mensajes a través de canales de comunicación. Además, son generalmente construidos con elementos heterogéneos y asíncronos. La tolerancia a fallos tiene el objetivo de proveer al cluster de una capacidad de operación continuada, intentando que sólo provoque una caída pequeña de prestaciones ocasionada por la prevención y la detección de fallos, tanto en la presencia como en la ausencia de

fallos. A pesar de conocerse un buen conjunto de técnicas de tolerancia a fallos, su aplicación en los clusters es muy compleja y sus resultados son muchas veces insatisfactorios debido al *overhead* introducido en el sistema [5].

1.2 Tolerancia a Fallos

Los clusters presentan una redundancia física intrínseca debido a la existencia de numerosos nodos de cómputo, que puede ser utilizada para el empleo de tolerancia a fallos. La ocurrencia de fallo en algún nodo procesador no tiene que significar necesariamente la interrupción del suministro del servicio. Una solución posible para esto es que el sistema sea reconfigurado, utilizando solamente los nodos disponibles.

Con el objetivo de aumentar la garantía de funcionamiento de los clusters, son esenciales, técnicas de prevención de fallos, detección de errores, diagnóstico, recuperación y reconfiguración.

Es necesario definir una arquitectura que permita ejecutar aplicaciones paralelas que requieren HP, HA y HT. Esta arquitectura debe garantizar que se realice la ejecución eficiente de una aplicación paralela cuando se utiliza un entorno con un elevado número de nodos, durante un largo período de tiempo. Por otro lado, es necesario considerar que la probabilidad de fallo aumenta, y puede llegar a ocurrir que el fallo ocasione la pérdida total del trabajo realizado. Un cluster de alta disponibilidad intenta mantener en todo momento la prestación de servicio encubriendo los fallos que se pueden producir.

Por lo tanto para proveer HA es necesario utilizar técnicas de tolerancia a fallos [6] que permitan garantizar (confiabilidad o garantía de servicio) Fiabilidad (*Reliability*), Disponibilidad (*Availability*) y Facilidad de mantenimiento (*Serviceability*) del sistema.

La disponibilidad de un sistema se puede ver afectada por diferentes averías de sus componentes, a pesar de utilizar componentes cotidianos sin características especiales de fiabilidad nos interesa disponer de un sistema fiable.

Si ocurre un error durante la ejecución y ocasiona una ejecución incorrecta de las funciones del sistema, se tiene una avería [7] [8]. Estos pasos, no se producen de forma simultánea en el tiempo, sino que existe un tiempo de inactividad, llamado "latencia del error" desde el instante en que se produce el fallo hasta que se manifiesta el error. [9]

En caso de aparición de fallo en algún componente, la tolerancia a fallos busca que el sistema siga trabajando, aunque esté funcionando en modo degradado, hasta que acabe la ejecución, lo que **podrá incidir en mayor o menor medida en sus prestaciones**. El número de fallos que pueden estar

presentes en un momento dado dependerá del número de componentes del sistema, del *MTBF* y del tiempo de ejecución de la aplicación. La probabilidad de que dos o más fallos ocurran simultáneamente decrece. La tolerancia a fallos en un sistema se logra mediante la inclusión de técnicas de redundancia [10]. Puede haber redundancia en cualquier nivel: utilización de componentes hardware extra (redundancia en el hardware), repetición de las operaciones y comparación de los resultados (redundancia temporal), codificación de los datos (redundancia en la información) e incluso la realización de varias versiones de un mismo programa (redundancia de software). Las técnicas más utilizadas en computación paralela son el uso de técnicas de replicación de datos (redundancia de información) [8] o de *checkpoint* (redundancia temporal) [11].

Para evitar la interrupción en el suministro del servicio, debido algún fallo en sus componentes, los fallos deben ser detectados lo más rápidamente posible: latencia del fallo.

El nodo en que ha ocurrido el fallo debe ser identificado a través de diagnóstico apropiado y finalmente reparado o aislado a través de reconfiguración del sistema. Esa reconfiguración se hace reasignando tareas y seleccionando caminos alternativos de comunicación entre los nodos.

Los fallos en clusters pueden ser causados por problemas en el hardware, sistema operativo (OS), aplicaciones, o por el propio software de tolerancia a fallos [12].

Como vemos es necesario utilizar alguna técnica o mecanismo que proteja al sistema, utilizamos redundancia temporal, protocolos de *rollback-recovery*, debemos considerar la prevención (generar redundancia), la detección (monitorización), el diagnóstico (latencia del error) y la recuperación y reconfiguración del sistema (retornar el sistema a condiciones operativas razonables después un fallo) como un todo. Para lograr esto la tolerancia a fallos incluye las etapas:

- Protección: redundancia
- Detección del error
- Estimación de daños: diagnosis
- Recuperación del error: llevar al sistema a un estado correctos, desde el que pueda seguir funcionando (tal vez con funcionalidad parcial)
- Tratamiento del fallo y continuación del servicio del sistema: consistencia global del sistema: reconfiguración y enmascaramiento del fallo.

Se debe tratar de mantener el equilibrio entre la latencia y el *overhead* que introducen las técnicas de tolerancia a fallos. Además de aumentar el coste del sistema, desde el punto de vista del usuario, las actividades realizadas por el mecanismo de

tolerancia a fallos reducen las prestaciones del sistema.

1.3 RADIC

Se parte de una arquitectura tolerante a fallos RADIC (*Redundant Array of Distributed Independent Fault Tolerance Controllers*) escalable, flexible, transparente y descentralizada para sistemas de paso de mensajes, cuyo objetivo principal es asegurar que una aplicación paralela-distribuida finalice correctamente aun si ocurrieron fallos en algunos nodos del computador paralelo.

RADIC establece un modelo de arquitectura que define la interacción de la arquitectura tolerante a fallos y la estructura del computador paralelo, implementa una capa entre el nivel de paso de mensajes y la estructura del computador.

El núcleo de la arquitectura RADIC es un controlador completamente distribuido para tolerancia a fallos que trata automáticamente los fallos en los nodos del cluster. Tal controlador comparte los recursos del computador paralelo usados en la ejecución de la aplicación paralela.

RADIC adopta el mecanismo de *rollback-recovery* basado en protocolo *log* pesimista, realiza automáticamente un grupo de actividades requeridas para la Tolerancia a Fallos. Cada actividad está dentro de alguno de los cuatro procedimientos generales requeridos por un mecanismo automático de tolerancia a fallos. La estructura de la arquitectura de RADIC usa un grupo de procesos que colaboran para crear un controlador distribuido para tolerancia a fallos. Hay dos clases de procesos: protectores y observadores. Cada nodo del computador paralelo tiene un protector dedicado y hay un observador dedicado asociado a cada proceso de la aplicación paralela. En la Tabla 1.1 se presentan las Fases Operativas de RADIC y el componente RADIC implicado en cada fase.

Tabla 1.1: Fases Operativas de RADIC

Fase Funcional	Actividad	Componente RADIC
Protección: Salvar el estado	<ul style="list-style-type: none"> Realizar y enviar los <i>checkpoints</i> Almacenar <i>checkpoints</i> Realizar y enviar <i>logs</i> Almacenar <i>logs</i> de mensajes Gestionar basura (<i>checkpoints</i> obsoletos). 	<ul style="list-style-type: none"> Observador Protector Observador Protector Protector
Detección de fallos	<ul style="list-style-type: none"> Monitorizar los nodos usando un mecanismo de <i>heartbeat/watchdog</i> Detectar los fallos de comunicación entre procesos. 	<ul style="list-style-type: none"> Protector Observador
Recuperación	<ul style="list-style-type: none"> Reiniciar los procesos fallidos desde los <i>checkpoints</i> y <i>logs</i>. Reejecutar y procesar los <i>logs</i>. 	<ul style="list-style-type: none"> Protector Observador
Reconfiguración y enmascaramiento de fallos	<ul style="list-style-type: none"> Asegurar que los procesos que no se vieron afectados por el fallo puedan encontrar los procesos recuperados. 	<ul style="list-style-type: none"> Observador y Protector

Estas fases suceden concurrentemente con la ejecución de la aplicación paralela y no interfiere en sus resultados [13].

2. LINEAS DE INVESTIGACION y DESARROLLO

- Definir la configuración adecuada de tolerancia a fallos para diferentes tipos de aplicaciones, teniendo en cuenta los requerimientos de rendimiento y prestaciones del usuario.
- Definir un modelo genérico de aplicación-prestación-tolerancia a fallos.
- Validar experimentalmente el modelo propuesto.
- Formar recursos humanos en HPC en general y Tolerancia a fallos, en particular.

3. RESULTADOS OBTENIDOS/ESPERADOS

Debido a que el presente proyecto está previsto para el trienio 2009-2011, actualmente se encuentra en la transición entre una etapa de recopilación, selección y análisis de la bibliografía existente sobre tolerancia a fallos, RADIC y simuladores de protocolos y tecnologías de red y la selección de una aplicación piloto, selección y configuración de un simulador de protocolos y tecnologías de red y configuración inicial y prueba de RADIC.

Para los próximos meses se espera iniciar la etapa de configuración de RADIC sobre el cluster real y simulado para la aplicación seleccionada y la experimentación sobre el simulador y cluster real, sin y con tolerancia a fallos con el fin de realizar una contrastación de las prestaciones y medidas de rendimiento obtenidas de la simulación y la ejecución en el cluster.

A futuro se realizarán las etapas de:

- Selección de un conjunto de aplicaciones con diferentes requisitos de rendimiento y prestaciones. Configuración del mecanismo de tolerancia a fallos RADIC para las aplicaciones seleccionadas.
- Caracterización del conjunto de aplicaciones seleccionadas y experimentación sobre el simulador, sin y con tolerancia a fallos (RADIC).
- Formulación del Modelo Aplicación-Prestación-Tolerancia-a-Fallos.
- Validación experimental del Modelo Aplicación-Prestación-Tolerancia-a-Fallos.

4. FORMACION DE RECURSOS HUMANOS

Durante el desarrollo del proyecto se tiene previsto realizar transferencia a los alumnos de la carrera de Ingeniería en Informática de la UNJu y a los alumnos de Maestría en Ingeniería del Software y de Doctorado en Informática de la Red UNJu-UNSL.

Además, tres de los integrantes del proyecto están realizando el trabajo final de la Especialización en Cómputo de Altas Prestaciones y Tecnología GRID y posteriormente trabajarán en su tesis Doctoral sobre esta temática. También se ofrecerán opciones de trabajos finales de carrera en HPC a los alumnos de Ingeniería en Informática.

5. BIBLIOGRAFIA

- [1] Buyya R. High Performance Cluster Computing: Architectures and Systems, Volume 1 [Book]. - [s.l.] : Prentice Hall PTR, 1999.
- [2] Rodrigues de Souza Josemar FTDR: Tolerancia a fallos, en clusters de computadores geográficamente distribuidos, basada en Replicación de Datos [Book]. - Barcelona : [s.n.], 2006.
- [3] Sterling T.L. Launching into the Future of Commodity Cluster [Conference] // IEEE International Conference on Cluster Computing (CLUSTER 2002). - 2002. - p. 345.
- [4] Bosilca G. [et al.] MPICH-V: Toward a Scalable Fault Tolerant MPI for Volatile Nodes [Conference] // SC '02: Proceedings of the Proceedings of the IEEE/ACM SC2002 Conference. - 2002.
- [5] Rodrigues de Souza J [et al.] Fault Tolerant Master-Worker over a Multi-Cluster Architecture [Conference] // ParCo2005. - 2005.
- [6] Avizienis A. Toward systematic design of fault-tolerant systems [Journal] // IEEE, Computer v.30, Iss.4. - 1997. - pp. 51-58.
- [7] Pradhan D.K. Fault Tolerant System Design [Book]. - New Jersey : Prentice Hall, 1996.
- [8] Weissman J.B. Fault Tolerant Computing on the Grid: What are My Options? [Conference] // High Performance Distributed Computing, 1999. Proceedings. The Eighth International Symposium on. - 1999.
- [9] Johnson B.W. Design and Analysis of Fault Tolerant Digital Systems [Book]. - [s.l.] : Addison Wesley, 1990.
- [10] Avizienis A., Laprie J. and Randel B. & Landwehr, C. Basic Concepts and Taxonomy of Dependable and Secure Computing [Journal] // IEEE Transactions on Dependable and Secure Computing v 1, n.1. - 2004. - pp. 11-33.
- [11] Chakravorty S. & Kalé, L.V. A Fault Tolerant Protocol for Massively [Conference] // 18th International Parallel and Distributed Processing Symposium (IPDPS'04) - Workshop 11. Publisher: IEEE Computer Society. - 2004.
- [12] Sun H. and Han J.J. & Levendel, H. A Generic Availability Model for [Conference] // PRDC '01: Proceedings of the 2001 Pacific Rim International Symposium on Dependable Computing. - Washington : IEEE Computer Society, 2001.
- [13] Duarte Angelo Amancio RADIC: A Powerful Fault-Tolerant Architecture [Book]. - Barcelona : [s.n.], 2007.

USO DE THREADS PARA LA EJECUCIÓN EN PARALELO SOBRE UNA MALLA COMPUTACIONAL

Julio Monetti & Oscar Leon /Ingeniería en Sistemas de Información/FR.Mendoza -UTN
jmonetti@frm.utn.edu.ar , oleon@frm.utn.edu.ar

CONTEXTO

El presente trabajo muestra el uso y análisis, desde el punto de vista de los Sistemas de Información, de metodologías no clásicas de computación a través de programación concurrente, observando características estructurales del software y hardware empleado, con el fin de asistir a aquellos usuarios que requieren la computación de alto rendimiento. Una alternativa para optimizar y compartir recursos de computación es la computación basada en ambientes *Grid* [1]. Se discute la administración de recursos a través de *Threads* en Java, para lanzar tareas sobre un *Grid*, con el objeto de paralelizar las mismas. Se presenta la arquitectura de la aplicación, diseñada sobre un modelo de 3 capas, y se comentan aspectos de su implementación, en un caso concreto para resolver un problema numérico.

Palabras clave: *paralelismo, programación concurrente.*

1. INTRODUCCION

En [2] se discutió la migración de una aplicación paralela realizada en MPI [3] sobre un *cluster*, hacia una arquitectura computacional desacoplada y con recursos descentralizados. Bajo esta migración se mantuvo el comportamiento paralelo del sistema, marcado fundamentalmente por la descomposición del dominio de datos a procesar. El nuevo ambiente de ejecución es una malla computacional montada sobre una red de área local, donde no se cuenta con las características que provee un *cluster*: paso de mensajes entre procesadores y facilidad en la

sincronización de procesos que provee MPI. No obstante, la posibilidad de crecimiento se ve sumamente favorecida, ya que se trata de una arquitectura heterogénea y teóricamente sin restricciones de escalamiento. El problema de estudio es la solución de una ecuación diferencial a través del método de diferencias finitas [4]. La migración de la aplicación desde *cluster* a *Grid* presenta nuevos problemas desde el punto de vista informático y la implementación de algoritmos que soporten dichas arquitecturas. Por otro lado, se probó en [2] que el algoritmo secuencial se puede mantener sin cambios luego de la migración, si se respeta un buen diseño en la codificación. La ingeniería de software como disciplina de desarrollo, sugiere la reutilización de código tanto a lo largo de un sistema, como así también a través de diferentes sistemas. El paradigma orientado a objetos toma este principio y lo pone como uno de los principales cimientos en el diseño de grandes sistemas. Esta forma de concebir la reutilización de código se materializa a través de técnicas de programación, donde el producto final de software es concebido en términos de diferentes capas. La programación en tres capas sugiere el desarrollo en forma separada de interfaces e interacción con el usuario en la capa más alta. En otra capa (o conjunto de componentes de software) la lógica principal del sistema. Finalmente en la capa inferior todos aquellos componentes de software que toman forma de servicios a ser explotados por la capa inmediata superior. En aquellos ambientes clásicos de programación, entiendase por tal a la

secuencial, la cual es generalmente utilizada para aplicaciones de cálculo y de carácter científico, no se aplica con frecuencia esta metodología, dando lugar a la duplicación de código, o a una mala estructuración del mismo.

Uno de los objetivos del presente trabajo, es reutilizar el código de cálculo, que será explotado en una arquitectura más enriquecida: la de computación distribuida. De esta forma se puede visualizar el producto como un conjunto de componentes de cálculo ubicados en la capa inferior o de servicios, una capa de software que conforma la lógica del sistema, y está destinada a la coordinación de los diferentes procesos de cálculo, y finalmente el *postproceso* ubicado en la capa de interface.

2. DESARROLLO

2.1 Descomposición del Dominio de Datos. Se emplea un *middleware* o capa de gestión para intercambiar información entre los equipos que constituyen el *cluster*. Este *middleware* es un componente de software que asegura una correcta administración de los procesos: sincronización, mensajes, etc. MPI permite la comunicación entre procesos residentes en diferentes computadoras de un *cluster*. El pasaje de mensajes entre ellas asegura la sincronización de las mismas y el intercambio de datos. Se debe considerar que una aplicación clásica en MPI supone distribuir el mismo programa a lo largo de todos los nodos del *cluster*, donde cada uno de estos ejecutará solo una porción del programa.

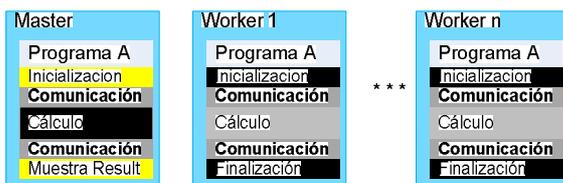


FIGURA 1. Distribución del Programa A al conjunto workers.

En la figura [1] se observa una estación de trabajo, el *master*, responsable de sincronizar el trabajo del resto de las estaciones (*workers*). El *master* es el encargado de particionar el conjunto de datos original, distribuirlo a cada *worker*, sincronizar la ejecución de las tareas de cálculo en cada uno de ellos y recibir los datos procesados. La versatilidad provista por MPI en la utilización de un *cluster* de computadoras permite el paralelismo y sincronismo de tareas llevada a cabo por cada *worker*. Como se mencionó en párrafos anteriores, con este conjunto de librerías se hace posible la ejecución del mismo programa en diferentes computadoras actuando cada uno de ellos sobre una parte diferente del dominio de datos, que en este caso está dado por el conjunto de puntos donde la ecuación diferencial debe ser calculada.

2.2. Multithreading como Solución a la Gestión de Procesos. En un entorno computacional donde existen diversos recursos de cálculo pero no la arquitectura de un *cluster*, la funcionalidad de MPI debe ser reemplazada por una arquitectura alternativa, por ejemplo una *malla computacional*, compuesta de nodos heterogéneos y sin una ubicación geográfica determinada. Se deberán emular en este caso los procedimientos de comunicación, administración y sincronización de tareas. Para satisfacer estas tareas de administración, resulta adecuada la utilización de n procesos ejecutándose en el *master* para mantener independientemente la administración y comunicación con los n procesos de cálculo, ahora distribuidos a lo largo de la malla computacional. Por otro lado, si se garantiza paralelismo entre estos n procesos de administración, se puede lograr un mayor paralelismo entre los nodos *workers*. Se utiliza una aplicación escrita en lenguaje Java, residente en el *master*, y se mantienen los códigos de cálculo codificados en lenguaje C en cada uno de las estaciones de trabajo. Cada hilo de

ejecución, representa un conjunto de instrucciones secuenciales, las cuales son ejecutadas en forma paralela con otros hilos de ejecución, eventualmente bajo la supervisión de un hilo coordinador. Cada hilo es responsable del envío del conjunto de datos correspondiente al proceso *worker* que está administrando. En el caso de descomposición del dominio, cada hilo actúa sobre un conjunto de datos diferente. No obstante ello, se debe tener en cuenta que todos los hilos comparten los mismos recursos dentro de la misma estación de trabajo: compartiendo espacio de direcciones, acceso al disco rígido, procesador, etc. Un programa Java puede ser considerado como un hilo único de ejecución al utilizar un flujo único de control. Dicho hilo tiene la posibilidad de crear distintas instancias de control, las que serán consideradas nuevos hilos. Cuando un determinado hilo modifica un dato en memoria este puede ser transmitido a través de un mensaje a otro hilo, así emulando el comportamiento de MPI. Java oculta la complejidad en el tratamiento de múltiples instancias de ejecución tras una manera muy simple y conveniente de codificación, creando cada hilo de ejecución como un objeto de datos [5]. Estos objetos (clase *Thread*) cuentan con un estado autónomo, contador de programa, estado de la pila e información sobre el estado de CPU. Esto se ve complementado por información en forma de atributos contenidos en dicha clase de datos. Luego, cada hilo comienza y termina su trabajo de forma independiente. El trabajo de cálculo llevado a cabo por las rutinas y/o módulos numéricos se reutiliza con relativa facilidad a partir de la codificación serial en lenguaje C. Como se especificó anteriormente, este código ya compilado reside en cada nodo *worker* (nodo de la malla computacional), y es responsable únicamente del cálculo de la ecuación en cada punto del subdominio de datos local. Por otro lado las tareas comunes llevadas a cabo por el nodo *master* son codificadas en el hilo principal del programa administrador. El objeto Java

empleado para describir una tarea en forma remota cuenta tanto con la información necesaria para ejecutar el programa en la estación de destino (identificador del recurso, tipo de estación, etc.), como así también con información necesaria para el control de ejecución de dicho trabajo. El hilo principal decide sobre la continuidad o no de determinado hilo, su recuperación, migración hacia otra estación, etc. Esto resulta sumamente ventajoso para realizar recuperación ante fallos debido a la salida de servicio de un determinado *worker* en tiempo de ejecución. El cuerpo del programa principal es encargado de crear los distintos hilos (clase *Trabajo* que hereda *Thread*), su iniciación y control. Puesto que existe una determinada cantidad de trabajos a ejecutar, cada objeto es contenido en un único arreglo (clase *Vector*).

2.3 Hilos de Ejecución y Grid Computing. Las mallas computacionales [1] permiten el aprovechamiento de recursos descentralizados, con el objeto de incrementar la capacidad de cómputo, almacenamiento de datos, etc. De esta forma, es posible presentar al usuario final el conjunto de recursos como una estación de cálculo única con gran capacidad de cómputo, requiriendo adecuar sensiblemente las aplicaciones finales. Por otro lado, un rasgo importante de un *cluster* a tener en cuenta es que este representa una herramienta dedicada y fuertemente acoplada, dedicada a resolver problemas de alto rendimiento y, en general, de alta disponibilidad, lo que conlleva una inversión importante en recursos de *hardware* e instalaciones auxiliares. Las mallas computacionales utilizan recursos provistos por equipos distribuidos geográficamente y conectadas por una red para resolver problemas de computación de gran escala. Estas posibilitan la realización de cómputos sobre grandes conjuntos de datos, los que se subdividen en partes más pequeñas, siendo estos procesados en los nodos que constituyen la *Grid*. En este sentido se requiere, al igual que en el caso

de *clusters*, que la aplicación presente concurrencia y en este caso también se aprovecha el paralelismo para realizar cálculos simultáneos. Así, la malla provee la capacidad de realizar computación de alto rendimiento, modelando una arquitectura de computadora virtual que permite, entre otras cosas, distribuir la ejecución de procesos a través de una estructura computacional enriquecida. Cada computadora disponible en el laboratorio, la Universidad, o el escritorio de un miembro del grupo de trabajo, se convierte en un potencial nodo *worker* al contar con el software de base correspondiente. El conjunto de nueve estaciones de cálculo utilizadas para este trabajo, eventualmente en estado ocioso, se convierten en un conjunto de nueve estaciones receptoras de datos, cada una de ellas administrada por un objeto (hilo de ejecución) residente en el *master*. El *middleware* utilizado se ve materializado a través del conjunto de protocolos y servicios provistos por el *Globus Toolkit* [6], el cual tiene la funcionalidad necesaria para asegurar la autenticación de usuarios, envío y recepción de datos entre estaciones, etc. Este software de base provee un servicio denominado *gatekeeper* capaz de recibir la solicitud de ejecución de un trabajo desde un nodo *master* y realizar dicha tarea en forma local en el *worker*. Esto obliga a la instalación de este servicio en cada *worker*, con el objeto de satisfacer la ejecución del código de cálculo en lenguaje C. La infraestructura GRAM provista por el *Globus Toolkit* se utiliza para la localización y administración de recursos de procesamiento en una malla computacional. A través de GRAM la aplicación coordinadora, en este caso la aplicación Java residente en el *master*, solicita la ejecución de la tarea de cálculo residente en un nodo remoto, quedando el hilo responsable de esta tarea a la espera de la finalización de la misma. Cada “ejecución GRAM” es una sentencia más contenida dentro de las funciones de cada hilo. Esta tarea es llevada a cabo con la

asistencia de *GridFtp*, aplicación provista también por el *Globus Toolkit* para la transferencia de ficheros entre nodos.

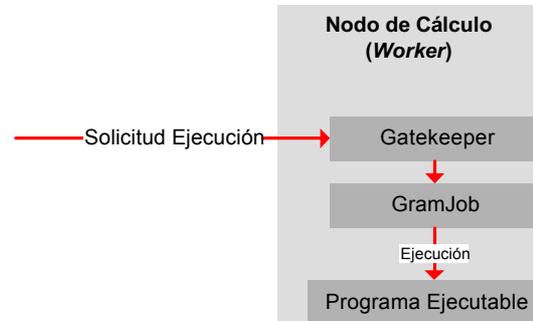


Figura. 3. La aplicación coordinadora invoca la ejecución sobre nodos remotos.

2.4. Implementación de los Hilos. Como se mencionó anteriormente, la solución final fué codificada en un programa Java, la cual tiene la posibilidad de manipular múltiples hilos de ejecución, donde cada uno de ellos representa un *objeto de datos* responsable de la administración de cada nodo *worker*. Al existir paralelismo entre los diferentes hilos de ejecución instanciados, se asegura el paralelismo de ejecución sobre los distintos recursos computacionales. Este paralelismo se obtiene a través del envío simultáneo de un mensaje especial a cada hilo (objeto *Thread*), “ordenándole” a los mismos el inicio de su ejecución. El código de cálculo residente en cada nodo *worker* complementa el sistema y es invocado por el servicio *gatekeeper* en forma local. Cada iteración sobre el conjunto de datos se transforma ahora en una invocación de ejecución; la que supone la ejecución del programa ejecutable en el nodo *worker* a través de los servicios GRAM. Esta dinámica dada por la ejecución del trabajo remoto y espera de los resultados procesados en cada iteración, se ve fortalecida por el manejo de hilos de ejecución de la aplicación coordinadora, permitiendo así definir tanto políticas de administración sobre cada uno de ellos, como así también observar tiempos de

ejecución sobre cada nodo entre otros datos. Previo a la invocación de la ejecución de tareas por parte de la aplicación coordinadora, esta debe descubrir qué recursos se encuentran actualmente disponibles, para lo cual hace uso de una función provista por una librería de funciones de Grid y aplicable en lenguaje java. Dicha función es la encargada de enviar paquetes de información hacia cada nodo *worker* quedando a la espera de la correspondiente respuesta. Cuando esta respuesta llega indica que el nodo *worker* está disponible para recibir solicitudes de ejecución.

Complementa este esquema una interfaz de usuario, plasmada en una aplicación WEB JSP [5], la cual tiene por objeto recolectar la información de entrada y características de ejecución provistas por el usuario final. De esta forma la aplicación *WEB* tiene acceso a la única credencial útil [1] para solicitar el uso de los recursos de la malla. Esta aplicación, se transforma ahora en la coordinadora de los distintos procesadores que calculan cada subconjunto de datos, pudiendo cada usuario final solicitar la ejecución de una instancia de la misma. Por último será esta aplicación la encargada de recolectar los resultados una vez finalizado el proceso iterativo, y presentar los mismos al usuario final. Este enfoque se ve enriquecido con el uso de portales *GRID*, donde se oculta al usuario la complejidad del uso de credenciales digitales, mejorando también la información sobre los recursos disponibles.

3. RESULTADOS OBTENIDOS

Se ha diseñado una aplicación capaz de emular el comportamiento paralelo para el cálculo de diferencias finitas a través de una malla computacional. Se observa que los tiempos se incrementan considerablemente debido a las características de la arquitectura utilizada: muy desacoplada y con una latencia muy alta. Sin embargo, la capacidad de escalabilidad de este modelo de ejecución no merece mayor comparación con la escalabilidad provista por un *cluster*

de computadoras. La malla computacional con respecto al *cluster* presenta un escenario de mayor “libertad de crecimiento”, como así también mayor incertidumbre de trabajo. La programación de aplicaciones en un ambiente de mallas computacionales puede presentar variados matices de acuerdo a la forma que toman los componentes de software y la forma en que son invocados.

Los autores estudian actualmente la migración de partes de la aplicación, las cuales se encuentran actualmente como archivos ejecutables, a *WebServices*. Por otro lado, se complementará el código Java con servicios provistos por Globus Toolkit para el seguimiento continuo de cada una de las tareas remotas.

4. FORMACION DE RECURSOS

Se trabaja en el departamento de Ingeniería en Sistemas de Información con alumnos del tercer año en temas relacionados con la programación concurrente, con el objeto de crear una relación directa con otros grupos de la Universidad dedicados al modelado numérico de aplicaciones; los cuales son potenciales usuarios de computación paralela.

5. BIBLIOGRAFIA

- [1] J.Monetti & C.García Garino. Mallas Computacionales. Conceptos y Experiencias. 2º Congreso Internacional de Informática CIDI-Cuyo. Mendoza. 13-15 de abril de 2005.
- [2] J. Monetti, C. Garcia Garino, O. León, Diseño y programación de aplicaciones para Grid Computing. Encuentro de Investigadores y Docentes de Ingeniería. ENIDI-2006. Mendoza, Argentina.
- [3] MPI. www-unix.mcs.anl.gov/mpi.
- [4] S. Chapra, C. Canale & P. Raymond, Métodos Numéricos para Ingenieros. Mc Graw-Hill. Cuarta Edición. 2004.
- [5] H. Deitel & P.J. Deitel. Java. Compo Programar. Prentice Hall. Quinta Edición. 2004.
- [6] The Globus Alliance. www.globus.org.

Seguridad en Entornos Virtuales

Javier Echaiz*

Jorge Ardenghi

Laboratorio de Investigación de Sistemas Distribuidos (LISiDi)
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur, Bahía Blanca (8000), Argentina
{je,jra}@cs.uns.edu.ar

Resumen

En un entorno virtual, un *monitor de máquina virtual* (VMM) controla múltiples VMs mediante una abstracción de software del hardware subyacente. Esta arquitectura provee algunas ventajas con respecto a la seguridad pero también introduce desafíos únicos. Irónicamente los avances en la potencia de cómputo y la disminución de los costos del hardware fueron los factores que dieron origen a la pérdida de interés en la virtualización, hoy principales contribuyentes de su renacimiento.

La virtualización surgió a finales de la década de 1960, con el objetivo de multiplexar las aplicaciones sobre mainframes de forma tal de poder repartir los escasos y costosos recursos de cómputo entre múltiples procesos. La creación de las VMs hizo posible que múltiples aplicaciones coexistiesen sobre una máquina única.

Esta línea de investigación busca desarrollar nuevas tecnologías centradas en la seguridad de los entornos virtuales, especialmente a nivel del VMM.

Palabras Clave: Virtualización, Detección de Intrusos, Seguridad en Redes, Políticas de Seguridad, Sistemas Distribuidos, Computación Colaborativa, Automatización.

1. Introducción

Cuarenta años después del surgimiento de la virtualización, modernos sistemas operativos (SOs), altas velocidades en los procesadores y bajos costos de hardware resolvieron los problemas para los que se inventaron las VMs. Fue posible entonces instalar nuevas aplicaciones tan fácilmente y barato como instalarlas en un servidor dedicado con su propio procesador, memoria y almacenamiento. Sin embargo, esta tendencia dio origen a nuevos problemas: el hardware barato hizo que proliferasen las máquinas subutilizadas, las cuales a su vez ocuparon espacios significativos y aumentaron el overhead en su administración. Mantener las aplicaciones actualizadas y los patches de los sistemas operativos en cada server se volvió una tarea pesada. Lograr que éstas máquinas fuesen seguras implicó que se determinasen claramente las responsabilidades de las organizaciones sobre cada server. Por ende, mover aplicaciones nuevamente a las VMs dentro de unas pocas máquinas físicas y administrándolas mediante monitores de VMs (VMMs) se volvió una solución factible.

Un VMM es una capa de software que usualmente corre directamente sobre el hardware. En los sistemas actuales el VMM puede correr junto a un sistema operativo host. Las VMMs exportan una abstracción de máquina virtual que se asemeja al hardware subyacente. Cada abstracción de máquina virtual es un *guest* que encapsula el estado completo del SO que corre dentro de ella. El sistema operativo guest interactúa con la abstracción virtual del hardware, la cual es a su vez manejada por el VMM como si fuese hardware real. La VMM es entonces un “sistema operativo de sistemas operativos”. La VMM usualmente corre en modo privilegiado mientras que el sistema operativo guest lo hace en modo usuario.

*Becario CONICET, Argentina.

2. Máquinas virtuales

Básicamente existen dos tipos de entornos de máquinas virtuales. En el primero, el VMM se conoce con el nombre de *hypervisor* y se implementa directamente entre el hardware y los sistemas guest. El ejemplo clásico de este tipo de máquinas virtuales es el hypervisor de Xen (<http://www.xen.org>). Este tipo de virtualización es especialmente aplicable a entornos de servidores tipo UNIX.

En el otro tipo de monitor de máquina virtual, el VMM corre como una aplicación más dentro del SO guest y depende de él para proveer drivers de I/O y código *bootstrap* en lugar de implementarlo directamente desde cero. Este tipo de VMM se popularizó en PCs comunes (siendo posiblemente VMware (<http://www.vmware.com>) el más difundido), cuya plataforma x86 no fue originalmente diseñada para soportar completamente la virtualización por hardware. Este tipo de entorno suele ser considerado menos seguro que el anterior debido a que el VMM depende de los servicios provistos por el sistema operativo host.

La virtualización actual se ha desviado de sus raíces originales como una herramienta de multiplexado para convertirse actualmente en una solución para hacer frente a los problemas de seguridad, confiabilidad y administración. Este hecho presenta consecuencias positivas y negativas, pues por un lado la virtualización junto con sus técnicas asociadas ayudan a resolver múltiples problemas de seguridad, especialmente porque las máquinas virtuales corren sobre un único sistema, el cual puede implementar un sistema seguro multinivel con sistemas virtuales separados en cada nivel. Por otro lado la virtualización fabrica espacios para nuevas vulnerabilidades, donde los mecanismos de seguridad tradicionales no están preparados para resolver estos problemas. Esta línea de investigación se suma entonces a la búsqueda de nuevos mecanismos para combatir estos nuevos vectores de ataque.

2.1. Monitores de máquinas virtuales (VMMs)

Los VMMs soportan tres atributos en entornos virtuales [1]: aislamiento, interposición e inspección. Los VMMs propician el aislamiento debido a que las VMs no comparten memoria física. Gracias a la abstracción de la memoria virtual el VMM puede crear la ilusión de que cada VM tiene su propio espacio de direcciones. De esta forma cada VM corre sin saber de la existencia de otras VMs, pues todas sus acciones están confinadas a su propio espacio de direcciones.

Para soportar interposición los VMMs gestionan todas las operaciones privilegiadas a nivel del hardware físico. Los SOs guest transfieren todos los traps e interrupciones al VMM para procesar los eventos. El VMM intercepta todos los pedidos de I/O provenientes de los dispositivos virtuales de las VMs y los mapea al dispositivo físico de I/O correspondiente. Gracias a esta abstracción el VMM gestiona y planifica todas las VMs simultáneamente. Los sistemas operativos guest no saben de la existencia del propio VMM, ni saben que se encuentran compartiendo recursos con otras VMs; por el contrario la abstracción les presenta la ilusión de estar interactuando directamente con los dispositivos físicos.

Por último los VMMs tienen acceso a todos los estados de una VM, incluyendo el estado del CPU, de la memoria y de los dispositivos. Esta visión provee las capacidades de inspección, propiciando *checkpoints*, *rollbacks* y *replays*. Adicionalmente estos atributos se combinan para brindar otra característica deseable: *portabilidad*. Los administradores pueden guardar, copiar, mover e instanciar entornos con estados encapsulados desde una máquina a otra.

Por todo ello decimos que el VMM es en sí mismo un pequeño sistema operativo que atrapa los eventos que surgen en los sistemas operativos guest, gestionando sus I/Os y mapeando datos a memoria y discos virtuales. Al mismo tiempo la implementación física está separada y oculta de los SOs guest pero colectivamente encapsulada permitiendo replicación y migración eficiente.

2.2. Virtualización: lo bueno

Las características inherentes de la virtualización y los atributos de las VMMs que arriba resumimos simplifican la gestión de los recursos de cómputo. La reducción del overhead administrativo facilita el proceso de hacer seguro un sistema.

2.2.1. Pool de hardware

Los VMMs brindan una vista uniforme del hardware subyacente y por ello una plataforma de hardware puede contener múltiples entornos virtuales. Por lo tanto los administradores pueden ver a una computadora como parte de un pool de recursos de hardware genéricos [2]. Esta característica permite disminuir costos de hardware y bajar los requerimientos de espacio. Adicionalmente, dado que el VMM puede mapear máquinas a recursos de hardware disponibles se simplifica el balance de carga y la escalabilidad, volviendo triviales las fallas de hardware.

2.2.2. Encapsulamiento

La propiedad de encapsulamiento mejora la seguridad desde varios frentes. Las VMs son fácilmente encapsulables y replicables, por lo tanto los administradores pueden sistemáticamente agregar en tiempo de ejecución nuevos servicios y aplicaciones al entorno replicado.

Cada nuevo servicio o aplicación puede correr independientemente sin el riesgo de corromper o interferir con otros. Si un sistema cae (*crash*) u ocurre un ataque, los administradores pueden suspender a la VM afectada, hacer un rollback a un estado de ejecución estable y recomenzar. Por último, analizar mediante replay puede exponer configuraciones defectuosas o proveer información acerca del método de ataque, vulnerabilidades, etc.

2.2.3. Logs seguros

Los servicios de seguridad, como los logs seguros y la protección de intrusos a nivel del SO puede implementarse también en el VMM. El hecho de que estos servicios se ejecuten separados de todos los procesos en un entorno virtual mejora sus capacidades.

La implementación de logs seguros a nivel del SO tiene la desventaja de que un atacante puede deshabilitar o modificar los logs una vez comprometido el sistema y por ello tampoco los logs proveen información útil para un efectivo análisis de seguridad forense. Estos problemas se solucionan con la implementación a nivel del VMM, pues el atacante no puede modificar los logs y con ello se vuelven una herramienta útil para el análisis forense.

Además la virtualización mejora la prevención y detección de intrusos mediante el uso de *clones*. Los sistemas de detección de intrusos suelen basarse en firmas de patrones de ataques conocidos, tornándolas prácticamente inútiles frente a nuevos eventos sospechosos. Por otro lado los IDSs basados en anomalías corren el riesgo de marcar como ataques acciones legítimas; o peor aún pueden aceptar eventos maliciosos repetidos como actividad normal. Los clones por otra parte pueden correr eventos sospechosos capturados en una copia virtual del sistema real y observar los cambios sin comprometer el sistema real.

Por último, la protección a intrusos en el nivel virtual permite detecciones imposibles para los sistemas tradicionales, pues permiten monitorear todos los eventos que ocurren en el entorno virtual y por lo tanto hacer cumplir una política preestablecida detectando por ejemplo *cracking* de passwords a partir de una lectura de bloques de discos que contienen un archivo con passwords seguido de una gran actividad de CPU.

2.2.4. VMs en cuarentena

La virtualización permite que los administradores puedan poner una o más VMs en cuarentena (fuera de la red) y buscar vulnerabilidades, evidencias de ataques, prevención de diseminación de código malicioso hacia otros nodos, etc.

2.2.5. Distribución de software

Para la mayoría de los sistemas complejos, las posibles combinaciones de hardware, versiones de sistemas operativos y librerías vuelven impráctico que los desarrolladores de software contemplen cada posible combinación. La virtualización alivia estos problemas permitiendo que los desarrolladores distribuyan máquinas virtuales completas que contengan su software.

Los dispositivos portables de almacenamiento flash, por ejemplo *pen drives*, permiten extender aún más este concepto. Los usuarios suelen emplear estos dispositivos para llevar consigo documentos, imágenes, etc. Con la virtualización es posible que un usuario lleve consigo una copia de su VM junto a sus archivos de trabajo y llevar consigo su máquina completa en un bolsillo.

2.3. Virtualización: lo malo

En la subsección anterior vimos ventajas de seguridad a partir del uso de entornos VM: infraestructuras que automáticamente realizan balance de carga, detectan fallas independientes de hardware que hacen que las VMs migren, se creen y se destruyan según la demanda de servicios particulares. Sin embargo, algunas propiedades de la virtualización hacen que lograr un nivel de seguridad aceptable en entornos VM sea más difícil.

2.3.1. Proliferación de VMs

La escalabilidad constituyó uno de los puntos fuertes desde la creación de los entornos virtuales, pues crear una nueva VM es tan simple como encapsular y generar mediante copia una nueva instancia. Sin embargo esto no siempre es bueno, pues los usuarios pueden tener demasiadas VMs de propósitos especiales, e.g., una para testing, otra para demostraciones, otra como *sandbox* para probar nuevas aplicaciones, y otra por cada sistema operativo. Esta situación torna inmanejable el tema desde el punto de vista de la performance y desde el gasto considerable de memoria física.

Además, desde el punto de vista de la administración la situación se vuelve compleja, pues las actualizaciones, configuraciones, etc. deben hacerse en cada máquina. Desde la perspectiva de la seguridad, la proliferación de VMs puede abrumar a los administradores, haciendo inseguras a algunas VMs (exponiendo a amenazas a la organización).

Finalmente la virtualización imposibilita algunos procedimientos tradicionales de seguridad, por ejemplo dado que múltiples VMs corren en el mismo host físico, deshabilitar un *port* en la máquina host para hacer segura una aplicación especial en una VM específica presenta el indeseable efecto de deshabilitarlo también para el resto de las VMs que podrían necesitarlo.

2.3.2. Comportamiento transitorio

El beneficio de la portabilidad es a la vez un problema grave de seguridad. Dado que las VMs se replican fácilmente puede aparecer una legión de “máquinas transitorias” que aparecen y desaparecen de la red. Esto puede llevar a que vulnerabilidades que no existían ahora aparezcan brevemente, infecten otras máquinas y luego desaparezcan antes de detectarlas, complicando en gran medida el trabajo de los administradores de la seguridad, no sólo respecto a la gestión de patches sino también dificultando los *penetration tests*.

Si bien los checkpoints, rollbacks y replays de las VMS ayudan a recuperar fallas en entornos virtuales también presentan el problema de la reexposición a vulnerabilidades, reactivación de servicios riesgosos, rehabilitación de cuentas desactivadas, etc.

2.3.3. Aspectos sociales

Los aspectos sociales relacionados con la virtualización son sociales por naturaleza. La facilidad de instanciación de las VMs puede llevar a que los administradores simplemente remuevan y reinstalen una VM comprometida en lugar de que analicen lo sucedido en un ataque.

2.3.4. Nuevos riesgos

Los entornos virtuales traen aparejados nuevos riesgos, como una máquina virtual robada que alguien llevaba en un pen drive. Éste es un riesgo análogo al de una laptop robada pero en este caso con un dispositivo de menor tamaño y por ende más fácil de sustraer. Adicionalmente un cracker podría tratar de robar máquinas virtuales completas atacando *file servers*.

3. Línea de investigación

En el presente artículo se analizaron brevemente algunos aspectos a tener en cuenta a la hora de emplear virtualización como alternativa segura a las máquinas físicas. Vimos que no todas son ventajas, y por ello nos proponemos investigar en profundidad en este campo, área que resurgió a partir de las promesas de una seguridad mejorada, confiabilidad y ventajas administrativas.

En particular nuestro interés se centra en la implementación de servicios de seguridad avanzados a nivel del VMM, explotando vistas que le son imposibles a un sistema de seguridad a nivel del SO guest.

Las promesas de la virtualización arriba mencionadas son viables pero debemos ser concientes de los riesgos inherentes que plantea esta tecnología, pues si bien es cierto que se disminuyen overheads, se facilita la administración y se combaten las vulnerabilidades de seguridad a nivel del SO, no es menos real que se introducen nuevos riesgos que van en detrimento de la seguridad del sistema.

Referencias

- [1] T. Garfinkel and M. Rosenblum, "A virtual machine introspection based architecture for intrusion detection," in *NDSS*, The Internet Society, 2003.
- [2] M. Rosenblum and T. Garfinkel, "Virtual machine monitors: Current technology and future trends," *IEEE Computer*, vol. 38, no. 5, pp. 39–47, 2005.
- [3] P. M. Chen and B. D. Noble, "When virtual is better than real," in *HotOS*, pp. 133–138, IEEE Computer Society, 2001.
- [4] T. Garfinkel and M. Rosenblum, "When virtual is harder than real: Security challenges in virtual machine based computing environments," in *Proceedings of HotOS X: The 10th Workshop on Hot Topics in Operating Systems*, USENIX, June 2005.
- [5] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Nuegebauer, I. Pratt, and A. Warfield, "Xen and the art of virtualization," in *Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP'03)*, (Bolton Landing, NY, USA), ACM, Oct. 2003.
- [6] R. Sailer, T. Jaeger, E. Valdez, R. Perez, S. Berger, J. L. Griffin, and L. van Doorn, "Building a MAC-based security architecture for the Xen open-source hypervisor," in *21st Annual Computer Security Applications Conference*, ACM, Dec. 2005.
- [7] S. Weber, P. A. Karger, and A. Paradkar, "A software flaw taxonomy: aiming tools at security," *SIGSOFT Softw. Eng. Notes*, vol. 30, no. 4, pp. 1–7, 2005.

AP-SIG: un SIG con funciones específicas para Agricultura de Precisión

Marcelo Ariel Uva

uva@dc.exa.unrc.edu.ar

Depto. de Computación

Facultad de Ciencias Exactas Físico Químicas y Naturales

Universidad Nacional de Río Cuarto

Oswaldo Campanella

ocampanella@exa.unrc.edu.ar

Depto. de Geología

Facultad de Ciencias Exactas Físico Químicas y Naturales

Universidad Nacional de Río Cuarto

Contexto

Desde hace ya un tiempo nuestro grupo de trabajo viene desarrollando tareas de investigación en temas vinculados a la aplicación de los Sistemas de Información Geográfica (SIG) en diversas áreas. La utilización de SIG en la actualidad es muy variado. Se utilizan para la producción cartográfica, evaluaciones ambientales y de recursos naturales, estudio y evaluación de redes de servicios (electricidad, telefonía, emergencias médicas, etc.) y transportes, sistemas de catastro, etc. Este grupo interdisciplinar, está conformado por profesionales y estudiantes provenientes del campo de las ciencias geológicas, geográficas, de la computación, agrónomas, etc. Uno de los trabajos más importantes desarrollados por este grupo ha sido la *Construcción de un mapa de riesgo por colapso de la ciudad de Río Cuarto usando la tecnología de los SIG*. La imperiosa necesidad del aprovechamiento eficiente de los recursos naturales junto con el contexto socio-económico en el cual está inmersa nuestra ciudad, ha motivado a este grupo al desarrollo de herramientas SIG aplicables a las tareas que comprenden la Agricultura de Precisión.

Resumen

La Agricultura de Precisión (AP) es un concepto agronómico de gestión de lotes o parcelas agrícolas, basado en la variabilidad espacial de los factores que controlan el rendimiento de un cultivo y en la posibilidad humana de alterarlos mediante la aplicación controlada de agroquímicos. La tecnología de los SIG aporta funciones de análisis de datos que permiten hacer más eficiente la aplicación de técnicas y productos para optimizar el tratamiento de cada sector del campo. El desarrollo de un SIG específico aplicable a la AP es sumamente necesario para lograr un máximo aprovechamiento de la información que el mismo campo brinda. Actualmente existen aplicaciones comerciales que proporcionan algunas funcionalidades limitadas para el área de la AP. En nuestra línea de investigación proponemos el desarrollo de un SIG, AP-SIG que contará con herramientas específicas para asistir al profesional a cargo (ingeniero agrónomo u otro) en las tareas de AP. Como resultados parciales se muestra una de las extensiones desarrolladas hasta el momento.

Palabras Clave: *Sistemas de Información Geográfica (SIG), Agricultura de Precisión, Polígonos de Thiessen.*

1. Introducción

La Agricultura de Precisión (AP) [1] es un concepto agronómico de gestión de lotes o parcelas agrícolas, basado en la variabilidad espacial de los factores que controlan el rendimiento de un cultivo y en la posibilidad humana de alterarlos mediante la aplicación controlada de agroquímicos. El uso de las nuevas tecnologías, como los Sistemas de Posicionamiento Global (GPS), sensores, monitores de rendimiento del cultivo, imágenes satelitales, y Sistemas de Información Geográfica (SIG) son incorporadas al campo. Los SIG vinculados a tareas de AP están limitados a la visualización de mapas de cosecha, cálculo del rendimiento promedio parcial o total de un lote, entre otras funcionalidades. Carecen, principalmente, de herramientas que permitan la vinculación de los datos, como por ejemplo la comparación de mapas de un mismo lote en campañas diferentes. Actualmente, ingenieros agrónomos vinculados con la AP recopilan un conjunto de datos: análisis de suelos, mapas de cosecha, datos referidos a como ha sido trabajado el campo, fertilizantes aplicados en campañas anteriores, información brindada por el dueño o encargado del campo, etc. Es necesario que el profesional a cargo de la toma de decisiones estratégicas de los tratamientos a seguir, tenga disponible toda esta información de manera integrada. Son los SIG quienes deben encargarse de esta labor permitiendo vincular datos existentes y posibilitando la extracción de nueva información. En nuestra línea de investigación proponemos el desarrollo de SIG-AP, un SIG que contará con herramientas específicas aplicables a la AP. Como resultado parcial se muestra una de las herramientas desarrolladas hasta el momento, la misma permite la generación de polígonos de Thiessen teniendo en cuenta restricciones espaciales como lo es el loteo del campo. Los desarrollos parciales han sido integrados como extensiones al SIG ArcView 3.2. Se tiene planificado en los próximos meses migrar este desarrollo a la plataforma SIG libre Gvsig.

2. Líneas de Investigación y Desarrollo

Desde hace tiempo nuestro grupo de trabajo viene desarrollando tareas de investigación en temas vinculados a la aplicación de SIG en diversas áreas. En nuestro trabajo, proponemos el desarrollo de SIG-AP, un SIG que contará con herramientas específicas aplicables a la AP. En las subsecciones siguientes se realiza una reseña general acerca de los SIG y se presenta nuestra propuesta, SIG-AP.

2.1. Sistemas de Información Geográfica (SIG)

La tecnología de los SIG (Sistemas de Información Geográfica) está basada en componentes de hardware y software que permite el análisis e integración de información geográfica, permitiendo visualizar los datos obtenidos en un mapa[2]. Los SIG integran la información a través de capas o coberturas de datos georeferenciados que se van superponiendo. De la integración por superposición se obtienen mapas temáticos. El uso de los SIG en la actualidad es muy variado. Se utilizan para la producción cartográfica, evaluaciones ambientales y de recursos naturales, estudio y evaluación de redes de servicios y transportes, sistemas de catastro, etc. Para el caso de la agricultura, los SIG brindan la posibilidad de obtener un mapa digital del campo y al mismo tiempo la interacción con el mismo. Un ingeniero agrónomo podría determinar los puntos exactos para la realización de las extracciones de muestras de suelo para un posterior análisis. En esta línea de investigación se ha utilizado el SIG ArcView 3.2 de la empresa ESRI [2]. Se optó por este SIG en un primer momento debido a su alto rendimiento, liderazgo en el mercado y además, por ser altamente configurable y extensible.

2.2. SIG-AP

La maquinaria agrícola moderna está equipada con la tecnología que permite la extracción de una gran cantidad de capas de información de cada

sector o lote del campo. Una de estas capas son los "mapas de rendimiento". Un mapa de rendimiento puede definirse como un modelo digital del lote conformado por un conjunto de puntos de mapeo registrados durante la cosecha del lote. Cada uno de estos puntos posee un conjunto de atributos tales como la humedad del suelo, rendimiento en quintales, etc. La cantidad de puntos de mapeo que posee un mapa de rendimiento varía según el tamaño del lote y ronda en un promedio de 22.000 puntos. En la figura 1 podemos observar un mapa de rendimiento.

A medida que se suceden las diferentes campañas, se va generando una historia del campo (por lotes). Además de los mapas de rendimiento, se recolectan mapas de aplicación de fertilizantes, de refertilización, de velocidad de siembra, de puntos de extracción de muestras de suelo con sus respectivos resultados, de ambientación, etc. El volumen de la información crece de manera exponencial a medida que pasa el tiempo, tornándose cada vez más compleja de administrar y manipular eficientemente. Por todo lo anterior que surge

la idea de desarrollar SIG-AP, un SIG formado por un conjunto de herramientas específicas para asistir al profesional agropecuario no sólo en la administración de los datos, sino también en la generación de nueva información. Para su desarrollo se siguieron los lineamientos establecidos por la metodología de desarrollo *Extreme Programming*[3]. SIG-AP a sido implementado en Avenue, el lenguaje de scripts de ArcView 3.2.

3. Resultados obtenidos / esperados: Extensión areal de la validez de un análisis de suelo delimitado por un lote

Previo a la siembra de una campaña, se efectúan extracciones de muestras de suelo en cada lote y se envían a analizar al laboratorio. Este proceso es realizado luego de seleccionar los lugares de manera estratégica y observado en paralelo, los mapas de rendimiento recolectados de las campañas anteriores.

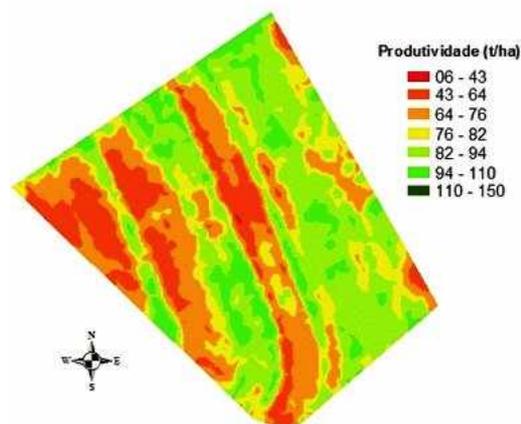


Figura nro. 1 - Mapa de rendimiento

El problema a tratar aquí es ¿hasta dónde se puede asegurar la validez de un dato?, es decir, ¿cuál será la distancia máxima en que es posible alejarse del punto de extracción de la muestra y seguir asegurando que esta tiene validez?. Para responder a estas preguntas hemos utilizado el concepto de polígonos de Voronoi o Thiessen[4].

Los Polígonos de Thiessen son una construcción geométrica que divide al plano euclidiano en secciones. Los polígonos de Thiessen establecen un método de interpolación basado en la distancia. Los polígonos son generados al unir los puntos entre sí, trazando las mediatrices de los segmentos unión. Las intersecciones de estas mediatrices determinan una serie de polígonos en un espacio bidimensional alrededor de un conjunto de puntos de control, de manera que el perímetro de los polígonos generados sea equidistante a los puntos vecinos y estableciendo su área de influencia. [4]. El cálculo los polígonos de Thiessen es posible de ser aplicado en este caso debido a que los datos obtenidos de los análisis de suelos son datos continuos. La metodología para extender arealmente la validez de un dato puntual depende principalmente de la forma en que el dato varía a lo largo de una línea horizontal. Aquellos datos que muestran una variación continua, tales como altitud o profundidad del nivel freático son interpolados por diversos métodos. Uno de los métodos más comúnmente usados es la generación de polígonos de Thiessen o Voronoi.

La generación de polígonos de Thiessen o Voronoi [5] ya es provista por Arcview, pero no de la forma que es necesaria. Generalmente, la recolección de las muestras de suelo es realizada en todo el campo, el problema aquí es que una muestra obtenida en un lote puede diferir enormemente con otra ubicada a los pocos metros por el sólo hecho de estar ubicadas en lotes diferentes. Esto tiene sentido si pensamos que cada lote es trabajado de forma diferente, es decir, puede ser cultivado con cereales distintos, fertilizado en diferentes proporciones y con diferentes agroquímicos, puede ser regado en forma diferentes, etc. Es por todo esto que a la hora de realizar una extrapolación de la validez de un resultado de análisis de suelo, es necesario tener en cuenta los límites espaciales impuestos por

el loteo del campo.

Una de las herramientas incorporadas a SIG-AP permite la discriminación de éstos límites. En la figura nro 2 es posible observar la generación de los polígonos de Thiessen sin tomar en cuenta las restricciones del lote y tomando en cuenta dichas restricciones espaciales.

4. Formación de Recursos Humanos

Durante el desarrollo de este trabajo, tres alumnos han realizado sus trabajos finales, correspondiente a la carrera de Analista en Ciencias de la Computación y otros tres los han iniciado. Además, uno de los autores de este trabajo ha obtenido el título de Dr. Geología.

5. Conclusiones y trabajos futuros

El incremento de la población mundial junto con una demanda creciente de alimentos y productos agrícolas, exige un uso racional y eficiente de los recursos naturales. La agricultura de precisión es un concepto agronómico que permite a través de nuevas herramientas tecnológicas dar a cada zona del campo cultivado el tratamiento agronómico más apropiado, permitiendo reducir costos en la producción, aumentando la productividad y haciendo un uso más eficiente de los insumos. Dentro de estas nuevas tecnologías, son los SIG los que poseen un potencial enorme para aportar al crecimiento de técnicas vinculadas a la tareas de AP. Actualmente, es muy poco el desarrollo llevado a cabo en esta área. Los SIG utilizados hoy en día no proveen las funcionalidades específicas para aprovechar al máximo todo el caudal de información obtenida del campo. En nuestra línea de investigación, se ha trabajado ,y se continúa haciéndolo, en el desarrollo de SIG-AP, un conjunto de herramientas específicas para ser aplicadas en tareas de AP. En los meses siguientes, se tiene pensado completar el desarrollo de SIG-AP y posteriormente migrar a Gvsig un SIG libre y muy potente.

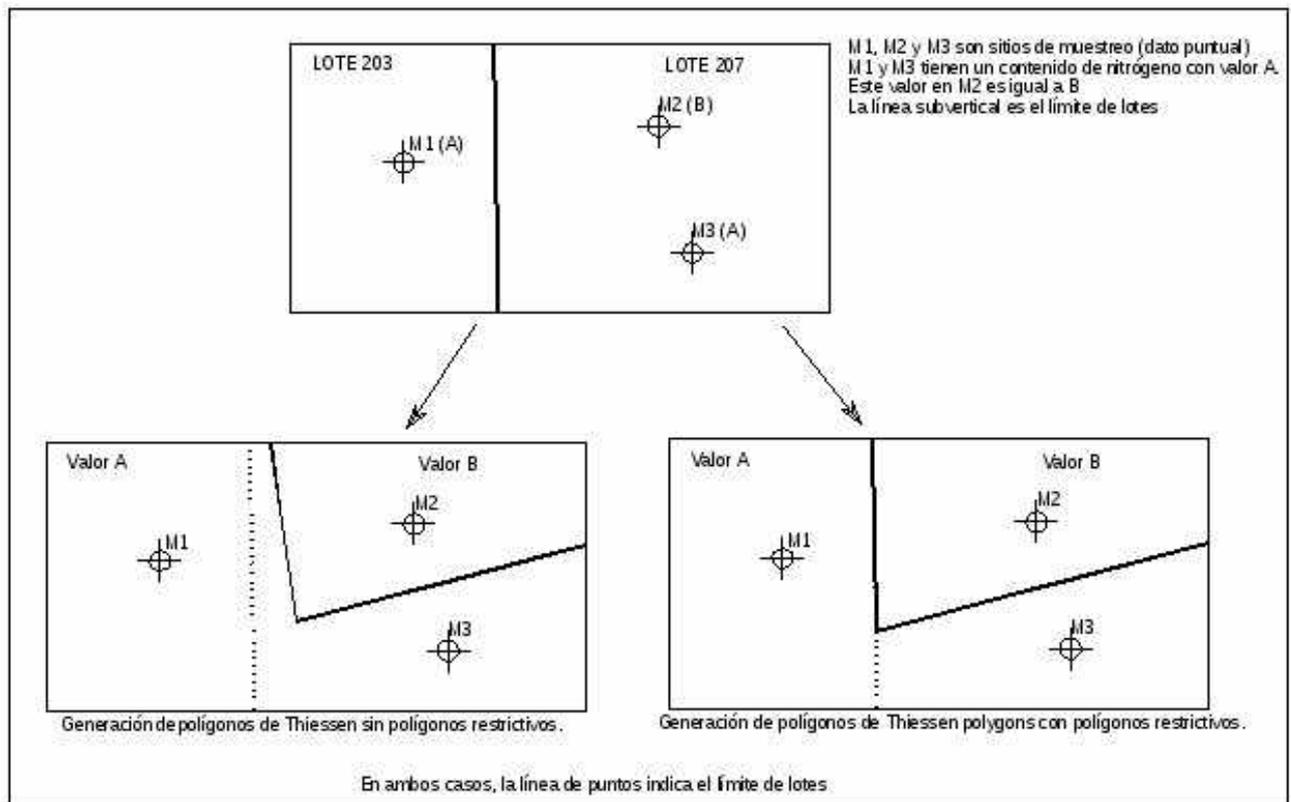


Figura no. 2 - Gráfico comparativo de la generación de los polígonos de Thiessen sin y con las restricciones espaciales

Referencias

- [1] Situación actual y posibilidades de la Agricultura de Precisión. Emilio Gil. Escuela Superior de Agricultura de Barcelona. Universidad Politécnica de Catalunya.
- [2] Sistemas de información geográfica: Prácticas con ArcView. Nieves Lantada Zarzosa, M. Amparo Núñez Andrés. UPC, 2002.
- [3] Extreme Programming and Agile Processes in Software Engineering: 6th International Conference, XP 2005, Sheffield, UK. Proceedings. Hubert Baumeister y otros. Publicado por Springer, 2005.
- [4] OKABE A., BOOTS B., SUGIHARA K. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. John Wiley and Sons. 1999.
- [5] BERG, KREVELD, OVERMARS, SCHWARZKOPF. Computational Geometry, Algorithms and Applications. Springer. 1997.

Un modelo de validación automático para la definición y mantenimiento de procesos de desarrollo de software

Pablo Szyrko, Mauricio Silclir, Gonzalo García Favre, Diego Rubio

{pablo.szyrko, msilclir, gonzalo.garcia.favre, rubiodiego}@gmail.com

Grupo de Investigación en Ingeniería y Calidad de Software

<http://www.institucional.frc.utn.edu.ar/sistemas/gidicalso/>

Departamento de Ing. en Sistemas de Información

Universidad Tecnológica Nacional

Maestro M. López esq. Cruz Roja Argentina

(X50165ZAA) Ciudad Universitaria, Córdoba, Argentina

CONTEXTO

Este trabajo surge como parte del proyecto de Investigación “Implementación de un modelo de validación para la definición y mantenimiento de procesos de desarrollo de software” el cual toma como base los conocimientos adquiridos y los resultados obtenidos en el proyecto "Implementación de modelos de madurez en empresas locales" realizado durante el período 2007-2008. Los mismos se realizan como parte de las actividades del Grupo de Investigación y Desarrollo en Ingeniería y Calidad del Software (GIDICALSO) de La Universidad Tecnológica Nacional – Facultad Regional Córdoba. Este proyecto se alinea en forma directa con los objetivos planteados por GIDICALSO con el fin de estudiar el cuerpo de conocimientos vinculados a la Ingeniería y Calidad del Software y los estándares, modelos, sistemas de excelencia y guías, junto con el estudio y desarrollo de herramientas vinculadas. Las actividades de investigación están paralelamente alineadas a las premisas definidas por el departamento de Ingeniería en Sistemas de Información y de la Secretaría de Ciencia y Tecnología de la UTN-FRC en pos de fomentar el desarrollo del área de la Ingeniería y Calidad del Software y sus servicios y poder contribuir a la industria del Software local y nacional.

RESUMEN

Diversos modelos, estándares y normas¹ han sido creados con el objetivo de asistir a las organizaciones en la definición y mejora de sus procesos. La premisa fundamental subyacente es que la calidad de un producto es determinada en gran medida por la calidad del proceso utilizado para desarrollarlo y mantenerlo. Si bien estos modelos se actualizan permanentemente para reflejar las mejores prácticas de ingeniería de software disponibles, las organizaciones enfrentan permanentemente la dificultad de asegurar que sus procesos cumplen con dichas prácticas.

Este proyecto de investigación pretende contribuir a la solución de este problema mediante el desarrollo de un modelo general (metamodelo) que mapee las prácticas identificadas en estos modelos, y su implementación en el proceso definido organizacional. Conjuntamente, se plantea desarrollar una herramienta que implemente dicho metamodelo, permitiendo analizar el impacto ante cualquier cambio tanto del modelo de referencia como del proceso organizacional. Finalmente se plantea la incorporación de trazabilidad bidireccional entre el proceso implementado por los proyectos y el proceso organizacional para permitir un análisis de impacto general dado cambios en cualquiera de los componentes mencionados (modelos, proceso definido, proceso implementado).

¹ Utilizaremos la palabra Modelo como referencia genérica a Normas, modelos y estándares para facilitar la narración.

Palabras clave: Modelo - Metamodelo – Framework – Validación – Proceso – Arquitectura de proceso

1. INTRODUCCION

A lo largo de la industria del software se han identificado una importante cantidad de grandes ideas y conocimientos disponibles acerca de cómo desarrollar efectivamente software, partiendo de la programación estructurada tradicional [Dahl:1972] hasta llegar a las actuales tecnologías de desarrollo [Pfleeger:2002].

Hoy en día, los equipos de desarrollo necesitan tener acceso a una amplia gama de información [Schaaf:2007]. No sólo es necesario adquirir información detallada sobre tecnologías de desarrollo específicas, tales como Java, Java EE, Eclipse, las tecnologías SOA, .NET, así como diversas herramientas y ambientes de desarrollo, sino que también es necesario averiguar la forma de organizar el trabajo a través de las mejores prácticas de desarrollo modernas, tales como metodologías ágiles, modelos iterativos, y desarrollo de software dirigido por el riesgo y la calidad [SPEM:2008]. En este contexto se presentan algunos problemas al pensar cómo definir el proceso de desarrollo en una organización:

1. Los miembros del equipo no tienen un acceso fácil y centralizado al mismo cuerpo de información para la ejecución del proceso cuando lo necesitan [SPEM:2008].
2. Se deben combinar e integrar contenidos y procesos de desarrollo que están disponibles en formato propietario, junto con las diferencias en los estilos y formas de presentación de cada uno de ellos [SPEM:2008].
3. Cada organización debe definir un enfoque sistemático y organizado que sea apropiado para sus necesidades [SPEM:2008].

A partir de estos problemas se desarrollaron meta-modelos que proveen los conceptos necesarios para modelar, documentar, presentar, administrar, intercambiar y publicar métodos y procesos de desarrollo [SPEM:2008]. Además representan procesos de ingeniería de negocios y software que ayuda a implementar un proceso de desarrollo de software efectivo [MSF:2002].

Cada organización que desarrolla software tiene definido un proceso de desarrollo, basado generalmente en alguna metodología estándar de la industria [PFLEEGER:2002]. Como una consecuencia de esta situación, se han creado diferentes herramientas que facilitan el modelado de dichos procesos de desarrollo, teniendo cada uno de ellas ventajas y limitaciones.

Paralelamente, el gran crecimiento de la industria del Software a nivel global ha estado sustentado, entre otros componentes, por la aplicación de modelos vinculados a las Tecnologías de la Información y la Comunicación (TIC's), y, entre ellos, los orientados a la Gestión de la Calidad, entre los cuales se encuentran CMMI [SEI:2006] e ISO [ISO:2008], entre otros. Esto implica que las organizaciones definen sus procesos de desarrollo en pos de cumplir con la aplicación múltiples modelos [Siviy:2008], pudiendo utilizar para ello las herramientas previamente explicadas.

2. LINEAS DE INVESTIGACION y DESARROLLO

La siguiente ilustración proporciona un mayor entendimiento de la problemática que representa la definición e implementación de procesos de desarrollo de software (Figura 1).

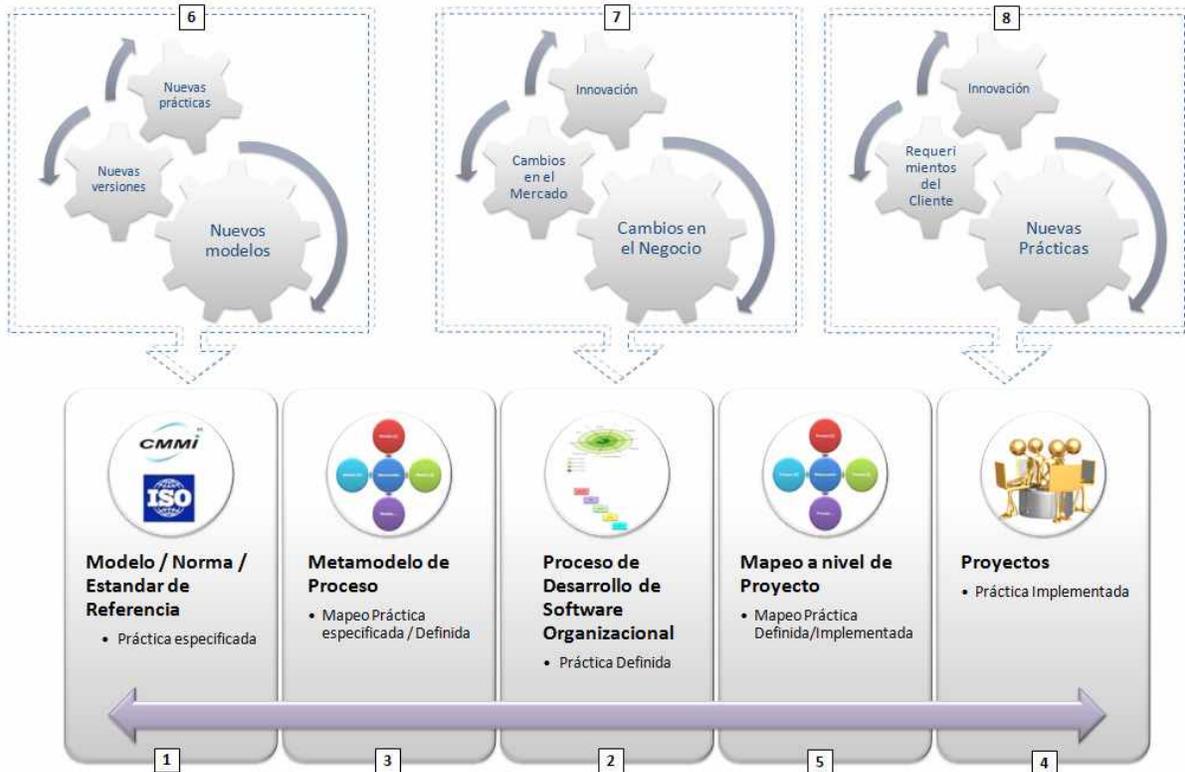


Figura 1: Procesos de desarrollo de SW

Cada modelo define un conjunto de **prácticas**, las cuales son la **referencia** para aquellas organizaciones que desean que su proceso de desarrollo esté conforme a dichos modelos (1).

Al momento de establecer el proceso de desarrollo de software se establecen aquellas **prácticas específicas definidas** para la organización, tomando como base las prácticas de referencia (2).

De esta forma se establece una relación directa entre la práctica de referencia y la práctica definida, denominada **Mapeo**. El grado en que las prácticas específicas estén mapeadas a las prácticas definidas determinará si el proceso es **conforme** a la norma, modelo o estándar (3).

Los proyectos que se desarrollan en la organización **implementan las prácticas** establecidas en el proceso de desarrollo de software especificado (4).

Nuevamente se establece un **Mapeo**, pero en este caso entre las prácticas definidas como estándar a nivel organizacional y las prácticas implementadas en un proyecto particular (5). El grado en que las prácticas del proyecto se mapean con las prácticas del proceso de desarrollo estándar determina si el proyecto está en conformidad con dicho proceso estándar.

El dinamismo, el cambio constante, es un factor clave al momento de analizar los procesos de desarrollo en una organización. Diversos factores determinan que los **modelos de referencia evolucionen** a lo largo del tiempo (6). Paralelamente la **definición del proceso de desarrollo de**

software a nivel organizacional también sufre **cambios** (7). De la misma forma, los proyectos no son ajenos a esta realidad, determinando que las **prácticas implementadas** sean **modificadas constantemente** (8).

Se hace visible entonces la necesidad de disponer de algún tipo de **mecanismo de validación** que verifique que el proceso implementado en la organización sigue cumpliendo con las reglas y prácticas definidas en el modelo estándar, tanto al momento de efectuar la especificación inicial (el mapeo entre el modelo estándar y la implementación de dicho modelo en la organización) como al momento de aplicar los cambios, y que esta conformidad se vea a su vez reflejada en cada implementación de las prácticas organizacionales a nivel de proyecto.

En la actualidad las validaciones y evaluaciones de la definición del proceso de desarrollo de software en término de mapeo con lo especificado en los modelos es realizado manualmente por un especialista (y en general distintos especialistas y tareas de evaluación por cada modelo utilizado).

Este proyecto pretende contribuir a la reducción y/o eliminación de los problemas de definición del proceso de desarrollo de software en las organizaciones a través de la **validación de dicha definición en relación a las prácticas definidas en los modelos que se pretenden implementar**.

A tal efecto, se postulan como ejes sobre los cuales se desarrolla la investigación el **desarrollo de un modelo** que permita generar el **mapeo entre las prácticas identificadas** de múltiples modelos y la **forma** en que éstos son **definidos en una organización**, y también el **mapeo a nivel de implementación de proyecto**. Paralelamente se pretende desarrollar una **herramienta que valide que la definición de un proceso de desarrollo de software implementado** en una organización cumpla con los lineamientos establecidos en la norma/estándar tomado como base, y también a nivel de proyecto. Esta herramienta permitirá también realizar un **análisis de impacto** que tendría cualquier **cambio** en el proceso implementado en la organización.

3. RESULTADOS OBTENIDOS/ESPERADOS

Hasta la fecha se realizaron investigaciones exploratorias sobre las herramientas de definición de proceso de desarrollo EPF [EPF:2007] y TFS [MSF:2002], como base para el desarrollo del proyecto.

El proyecto de investigación tiene planteados los siguientes objetivos que guían las actividades en curso:

1. Obtener conocimientos teóricos y prácticos relacionados a las herramientas de definición y modelado de procesos de desarrollo de software con el fin de conocer sus principales características y comprender la forma en que están implementadas.
2. Obtener conocimientos teóricos y prácticos relacionados a meta modelos de procesos de desarrollo de software.
3. Desarrollar un modelo general que permita generar el mapeo entre las prácticas identificadas de una norma/estándar y la forma en que éstos son implementados en una organización, en términos de una definición de proceso de desarrollo de software.

4. Desarrollar una herramienta que valide que la definición de un proceso de desarrollo de software especificado cumpla con los lineamientos establecidos en la norma/estándar tomado como base. Esta herramienta permitirá también realizar un análisis de impacto que tendrá cualquier cambio dado en el proceso especificado.
5. Implementar el modelo generado para un subconjunto de componentes de un modelo estándar de proceso de desarrollo de software.

4. FORMACION DE RECURSOS HUMANOS

Se planea la transferencia de los resultados del proyecto en dos niveles: a nivel académico y a nivel industrial. A nivel académico, a través del desarrollo inicial de investigadores relacionados a la temática y el dictado de seminarios y conferencias, introduciendo a alumnos y docentes en los temas relacionados a la implementación de modelos de calidad en las organizaciones y la validación de los mismos. En particular, se planea la participación de un Tesista (actualmente finalizando la Maestría en Ingeniería de Sistemas de Información de la UTN-FRC) y 3 becarios en las actividades de este proyecto.

Y a nivel industrial a través de la confección de materiales para el dictado de cursos de capacitación en relación al modelo de validación de implementación de procesos de desarrollo, el desarrollo de un modelo general que permita generar el mapeo entre las prácticas identificadas de los modelos y la forma en que éstos son implementados en diversas organizaciones y la implementación de una herramienta que permita validar la implementación de un proceso de desarrollo de software para un conjunto de prácticas de los modelos de referencia seleccionados.

6. BIBLIOGRAFIA

[Dahl:1972]	O.J. Dahl, E. W. Dijkstra, C. A. R. Hoare; Structured Programming; Academic Press; England; 1972.
[EPF:2007]	Eclipse Process Framework Composer - Part 1 Key Concepts; [Online]; 2007. http://www.eclipse.org/epf/general/EPFComposerOverviewPart1.pdf
[ISO:2008]	International Organization for Standarization. <i>ISO9001:2008 Sistemas de gestión de la calidad - Requisitos</i> . s.l. : ISO copyright office, 2007. ICS 01.040.03.
[Juran:1988]	J.M.Juran; Juran on Planning for Quality; New York: MacMillan, 1988.
[MSF:2002]	Microsoft Solution Framework Process Model v.3.1; [online]; 2002. http://download.microsoft.com/download/2/3/f/23f13f70-8e46-4f44-97f6-7dfb45010859/MSF%20Process%20Model%20v.%203.1.pdf
[Pfleeger:2002]	J.L.Pfleeger; Ingeniería del Software: Teoría y Práctica; Buenos Aires: Prentice Hall, 2002.
[SEI:2006]	CMMI Product Team; CMMI for Development, version 1.2. Pittsburgh, Pennsylvania, USA : Software Engineering Institute (SEI), August 2006. CMU/SEI-2006-TR-008
[Siviy:2008]	Jeannine Siviy, Pat Kirwan, Lisa Marino, and John Morley; [online]; 2008. http://www.sei.cmu.edu/prime/documents/multimodelSeries_wp4_processArch_052008_v1.pdf
[SPEM:2008]	Software & Systems Process Engineering Meta-Model Specification; [online], 2008. http://www.omg.org/cgi-bin/apps/doc?formal/08-04-01.pdf

Verificación estática de confidencialidad en un sistema de múltiples niveles de seguridad basado en Java bytecode

Ricardo Medel, César Martínez Spessot, Juan Carlos Vázquez, Ignacio Giagante

Departamento de Ingeniería de Sistemas de Información
Facultad Regional Córdoba
Universidad Tecnológica Nacional
Maestro M. López esq. Cruz Roja Argentina
Ciudad Universitaria (X5016ZAA)
Córdoba, Argentina

CONTEXTO

La línea aquí presentada es la línea principal de investigación del grupo, centrado en la seguridad informática a través del análisis estático del código. Esta línea de investigación también tiene como objetivo la consolidación del grupo de investigación y su integración en el sistema de ciencia y técnica de la UTN, debido a que dicho grupo es de reciente creación. Este grupo, junto con otros grupos de reciente creación, forman parte de un programa de fomento de la investigación científica puesto en marcha hace un par de años por del Departamento de Ingeniería de Sistemas de Información de la UTN-FRC.

RESUMEN

Debido al incremento en la cantidad e importancia de la información manejada por sistemas informáticos, también se incrementa la importancia de controlar la difusión de información privada de los usuarios. Los mecanismos de control del acceso y las técnicas criptográficas aseguran que la información sólo es accedida por entidades autorizadas, pero resultan insuficientes puesto que una vez que la información es accedida no existe control sobre su distribución.

Extender el sistema de tipos de un lenguaje de programación agregando niveles de seguridad permite analizar estáticamente, antes de su ejecución, cómo un programa accede y utiliza la información

confidencial. El creciente uso de código móvil –software obtenido de la red y ejecutado localmente, generalmente en Java *bytecode*– resalta la importancia de aplicar estos análisis de confidencialidad a lenguajes de bajo nivel.

En este proyecto desarrollaremos un sistema de tipos para Java *bytecode* extendido con información de seguridad e implementaremos un chequeador de tipos que verifique la confidencialidad de la información local accedida por código móvil. También estudiaremos la desclasificación controlada de información, a fin de incluir en nuestro análisis a programas que inevitablemente revelan información confidencial, por ejemplo, al responder negativamente ante una contraseña inválida

Palabras clave: Seguridad informática, Confidencialidad, Sistemas de tipos, Lenguaje Ensamblador, Java *bytecode*

1. INTRODUCCION

En un sistema distribuido, el software puede ser importado de la red y ejecutado localmente. Este tipo de software es llamado "código móvil" y usualmente necesita acceder a los datos almacenados localmente. A fin de proteger los datos locales de accesos no autorizados y de modificaciones ilegales, es necesario establecer mecanismos de seguridad que

aseguren que las políticas de confidencialidad e integridad de los datos locales se cumplen cuando son accedidos por código móvil. Distintos mecanismos han sido desarrollados para proteger distintos aspectos del acceso a la información, incluyendo mecanismos de control de acceso y técnicas criptográficas. Sin embargo, estos mecanismos no pueden asegurar la confidencialidad de la información una vez que los datos son accedidos, ya que no hay ningún control sobre su distribución posterior.

A fin de controlar el manejo que un programa hace de la información confidencial, se pueden utilizar técnicas de análisis del flujo de la información [1]. Para ello, los sistemas con múltiples niveles de seguridad (MLS, por sus siglas en inglés) asignan a cada dato, objeto o fragmento de información manipulado por el programa un rótulo de seguridad indicando su nivel de confidencialidad, mientras que a cada entidad computacional que puede acceder o almacenar dicha información se le asigna un rótulo de seguridad indicando su nivel de acceso. De esta manera, los rótulos permiten analizar si los programas cumplen con las políticas de seguridad que indican a qué información pueden acceder las entidades y cómo pueden distribuirla [2, 3]. Las políticas de seguridad se pueden dividir entre “de confidencialidad” (información con nivel de confidencialidad c sólo pueden ser accedidas por entidades con nivel de acceso a si $a \geq c$) y “de integridad” (la información en entidades con nivel de acceso $a1$ puede ser modificada sólo por entidades con nivel de acceso $a2$ si $a2 \geq a1$).

La formalización de la noción de “no-interferencia” [4], que establece que los resultados públicos (con bajo nivel de confidencialidad) de un programa no puede depender de los datos secretos (con alto nivel de confidencialidad) a los que el programa accede, permitió el desarrollo de sistemas de tipos para analizar el flujo de la información en programas escritos en

lenguajes de programación de alto nivel [5, 6]. Estos sistemas de tipos permiten realizar un análisis estático del código del programa, antes de que éste sea ejecutado, por medio de simples programas verificadores de tipos. Los cuales a su vez pueden ser verificados completamente e incluidos en el sistema básico de seguridad (llamado TCB – *Trusted Computing Base*) de la máquina local sin incurrir en costos computacionales elevados.

Aunque algunos sistemas de tipos para lenguajes de bajo nivel (ensamblador y *bytecode*) han sido desarrollados en la última década [7, 8], el análisis del flujo de información en dichos lenguajes era un problema abierto [1] hasta recientemente, cuando se desarrollaron sistemas de tipos para la confidencialidad de datos accedidos por programas en lenguajes de bajo nivel [9, 10, 11, 12]. Sin embargo, todos los trabajos sobre Java *bytecode* asumen que el código es analizado previamente por el *Java Verifier*, una herramienta que asegura, entre otras cosas, que la pila de ejecución es utilizada correctamente. Nuestra investigación apunta a eliminar ese requisito de modo de evitar la necesidad de incluir el *Java Verifier* en la TCB.

2. LINEAS DE INVESTIGACION y DESARROLLO

Nuestro proyecto aplicará las técnicas estáticas de verificación de confidencialidad desarrolladas para lenguajes ensambladores [11, 12, 20] a la creación de un sistema de tipos para Java *bytecode* que asegure, sin intervención de verificadores externos, que los programas descargados desde la red siguen las políticas de confidencialidad del usuario de dichos programas.

Además, estudiaremos cómo permitir que cierto monto de información confidencial pueda ser observada por entidades de menor nivel de seguridad mientras se mantiene la confidencialidad del sistema. Este fenómeno se denomina “desclasificación”, y su inclusión en las técnicas de

verificación permite analizar programas que por su naturaleza no poseen la propiedad de no-interferencia, tales como los programas que verifican contraseñas, los cuales responden negativamente si la palabra ingresada no es la contraseña esperada, mostrando información de bajo nivel de confidencialidad en base a información de alto nivel (la contraseña).

Las técnicas aplicables a MLS fueron desarrolladas originalmente para sistemas militares, pero con el aumento exponencial de la cantidad de programas distribuidos por Internet y ejecutados en computadoras locales, cada vez es más probable que la información confidencial de usuarios, empresas y gobiernos quede expuesta por ataques malintencionados o errores en el software [13]. Nuestro proyecto apunta a disminuir la incidencia de estos ataques, mediante la incorporación de un simple verificador de tipos en el sistema computacional local que verifique que el software bajado de la red cumple con las políticas de confidencialidad del usuario.

El grupo de trabajo original, sito en el Stevens Institute of Technology (Hoboken, New Jersey, EE. UU.), ha desarrollado técnicas para la verificación de la confidencialidad de los datos locales accedidos por programas en lenguaje ensamblador [11, 19, 20]. Dichas técnicas han sido implementadas en los lenguajes SIF [19] y SIFTAL [11], ambos basados en tecnología RISC. Se ha desarrollado un simple compilador y verificador de tipos, escrito en el lenguaje de programación funcional Haskell, para el lenguaje SIF. Además, ha sido definida una función de compilación utilizando un lenguaje imperativo simple similar a C como lenguaje fuente y SIFTAL como lenguaje objeto. Esta función está en proceso de ser implementada. Cabe acotar que un grupo de investigación en el Laboratorio de Investigación y Formación en Informática Avanzada (LIFIA) de la Universidad Nacional de La Plata, ha implementando un

verificador de tipos para SIFTAL, el cual está siendo utilizado en nuestro proyecto.

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo de desarrollo inmediato es crear un compilador de un lenguaje simple a SIFTAL para alimentar el verificador desarrollado en el LIFIA. Luego se aplicará esa experiencia en el desarrollo de un verificador de tipos de seguridad para Java *bytecode*.

Paralelamente, se está estudiando la posibilidad de aplicar esta técnica a una parte del desarrollo de un sistema de “*e-mall*”. Este sistema, en desarrollo por otro grupo de I&D de la UTN-FRC y Microsoft, permite a un proveedor de *e-mall* ofrecer a comercios instalados en un centro de compras un *framework* para proveer a sus clientes información y compras virtuales a través de un sistema *wireless* instalado en el centro comercial. Los comercios utilizan este *framework* para crear muy fácilmente sus servicios web y ponerlos a disposición de los clientes. A su vez, los clientes acceden a esos servicios por medio de sus dispositivos de Internet móviles al entrar al centro comercial. Nuestra técnica podría ser utilizada por el proveedor del *framework* a fin de asegurar que los servicios en línea creados por los comercios cumplen con las políticas de confidencialidad de los datos de los clientes, las cuales han sido establecidas por el proveedor en el contrato aceptado por el cliente cuando se inscribe por primera vez en el servicio.

Como una extensión al proyecto en curso, este desarrollo para “*e-mall*” podría ser utilizado para una comparación empírica de la técnica estática con el reciente renacimiento de las técnicas dinámicas a través de monitores. Se ha postulado [21] que los monitores tienen un poder de análisis de confidencialidad equivalente a los sistemas de tipos. La aplicación de nuestro enfoque al sistema de “*e-mall*” y la creación de un sistema de monitoreo

dinámico para el mismo permitirán un análisis empírico que permita contribuir al estudio de las similitudes y diferencias de estos dos enfoques desde el punto de vista práctico.

4. FORMACION DE RECURSOS HUMANOS

Dado que este es un grupo de investigación de reciente creación, se realizará una primera etapa de estudio dirigido de las técnicas de análisis estático basadas en tipos de datos, a fin de nivelar los conocimientos de los nuevos integrantes y consolidar el grupo. Como resultado de esta etapa se implementará un compilador de un lenguaje imperativo simple al lenguaje SIFTAL, mencionado anteriormente, y se integrará dicho compilador con el verificador desarrollado en el LIFIA.

En lo que respecta a la relación de la formación de recursos humanos con las asignaturas de la carrera asociada al Departamento, este proyecto ampliará los conocimientos del grupo de investigación en temas de las cátedras de Sintaxis y Semántica del Lenguaje, Tecnología de Software de Base y Paradigmas de Programación. Dichos conocimientos serán transferidos a estas cátedras a través de seminarios y cursos sobre los temas tratados en el proyecto de investigación.

También se formarán estudiantes avanzados de la carrera de Ingeniería en Sistemas de Información, quienes serán invitados a iniciar su formación en investigación científica y tecnológica, profundizando sus conocimientos en temas de seguridad de sistemas de información y técnicas de seguridad a través de lenguajes de programación.

A fin de transferir los resultados de este proyecto al ámbito de la industria de desarrollo de software, se escribirán artículos científicos para ser presentados en congresos de la especialidad, se presentarán pósters en jornadas profesionales y

estudiantiles, se escribirán artículos de difusión general que se publicarán en revistas técnicas, y se organizarán cursos abiertos a estudiantes, docentes, investigadores y profesionales.

5. BIBLIOGRAFIA

- [1] A. Sabelfeld y A. Myers, "Language-based information-flow security", IEEE Journal on Selected Areas in Communications, 21(1), 2003.
- [2] David Bell y Len LaPadula, "Secure computer systems: Mathematical foundations and model", Technical Report MTR 2547 v2, MITRE, November 1973.
- [3] Dorothy E. Denning y Peter J. Denning, "Certification of programs for secure information flow", Communications of the ACM, 20(7):504-513, July 1977.
- [4] Joshep A. Goguen y Jose Meseguer, "Security policy and security models", en Proceedings of the Symposium on Security and Privacy, págs. 11-20, IEEE Press, 1982.
- [5] Dennis Volpano, Geoffrey Smith y Cynthia Irvine, "A sound type system for secure flow analysis", Journal of Computer Security, 4(3):167-187, 1996.
- [6] Aninda Banerjee y David Naumann, "Secure information flow and pointer confinement in a java-like language", en Proceedings of 15th IEEE Computer Security Foundation, págs. 253-267, June 2002.
- [7] G. Morrisett, D. Walker, K. Crary y N. Glew, "From System F to Typed Assembly Language", ACM Transactions on Programming Languages and Systems, 21(3):528-569, May 1999.
- [8] David Aspinall y Adriana B. Compagnoni, "Heap bounded assembly language", Journal of Automated Reasoning, Special Issue

- on Proof-Carrying Code, 31(3-4):261-302, 2003.
- [9] G. Barthe, A. Basu y T. Rezk, "Security types preserving compilation", en Proceedings of Verification, Model-Checking and Abstract Interpretation - VMCAI '04, LNCS 2937, Springer-Verlag, 2004.
- [10] Samir Genaim y Fausto Spoto, "Information Flow Analysis for Java Bytecode", en Proceedings of the 6th International Conference VMCAI '05, Paris, France. January 2005, LNCS 3385, págs. 346-362.
- [11] E. Bonelli, A. Compagnoni y R. Medel, "Information-Flow Analysis for Typed Assembly Languages with Polymorphic Stacks", CASSIS 2005, Nice, France. March 8-11 2005, LNCS 3956, Springer-Verlag, 2006.
- [12] Dachuan Yu y Nayeem Islam, "A typed assembly language of confidentiality", en Proceedings of the 2006 European Symposium on Programming (ESOP'06), Vienna, Austria, March 2006, LNCS 3924.
- [13] Ross Anderson, "Security Engineering: A Guide to Building Dependable Distributed Systems", Wiley, 2001.
- [14] Edward Amoroso, "Fundamentals of Computer Security Technology", Prentice-Hall PTR, 1994.
- [15] Jean-Francois Bergeretti y Bernard A. Carré, "Information-flow and data-flow analysis of while-programs", ACT Trans. Program. Lang. Syst., 7(1):37-61, 1985.
- [17] Karl Crary, Aleksey Kliger y Frank Pfenning, "A monadic analysis of information flow security with mutable state", Journal of Functional Programming, 15(2):249-291, 2005.
- [18] Mads Dam y Pablo Gambiagi, "Confidentiality for mobile code: The case of a simple payment protocol", en 13th IEEE Computer Security Foundations Workshop, 2000, CFW-13, págs. 233-244, 2000.
- [19] R. Medel, A. Compagnoni y E. Bonelli, "Non-Interference for a Typed Assembly Language", en 9th Italian Conference ICTCS '05, Certosa di Pontignano, Italy, 12-14 October 2005. LNCS 3701, Springer Verlag, 2005.
- [20] Ricardo Medel, "Typed Assembly Languages for Software Security", PhD Thesis, Stevens Institute of Technology, January 2007.
- [21] Alejandro Russo y Andrei Sabelfeld, Chalmers University of Technology, comunicación personal, 2009.

Desarrollo y Evaluación de Ontologías en Áreas de la Informática Teórica y Aplicada

Margarita M. Álvarez, Diana Palliotto, Graciela E. Barchini
Secretaría de Ciencia y Técnica
Departamento de Informática - Facultad de Ciencias Exactas y Tecnologías
Universidad Nacional de Santiago del Estero (UNSE)
e-mail: {alvarez, dpalliot, grael}@unse.edu.ar

CONTEXTO

En este trabajo se presenta los avances del subproyecto denominado "Ontologías en los Sistemas de Información/Conocimiento", cuya finalidad es analizar el uso de las ontologías en distintos ámbitos de la Informática teórica y aplicada y abordar el rol de las ontologías en los Sistemas de Información. Este subproyecto es parte del proyecto de investigación denominado "Herramientas Conceptuales, Metodológicas y Técnicas de la Informática Teórica y Aplicada", avalado y subvencionado por la Secretaría de Ciencia y Técnica (SeCyT) de la UNSE.

RESUMEN

En este trabajo se presentan dos líneas de investigación: a) Desarrollo y evaluación de Ontologías en distintas áreas de la Informática Teórica y Aplicada; y b) Desarrollo y evaluación de los Sistema de Información Basado en Ontologías (SIBO).

En la primera línea se desarrollan dos ontologías de soporte (para la ingeniería de requerimientos y para la ingeniería de la usabilidad) dentro del ámbito de la Informática teórica, y una ontología en el área de la Informática médica. Además, se establecen las dimensiones e indicadores para evaluar la calidad de una ontología y analizar la usabilidad y reusabilidad de ontologías.

En la segunda línea de investigación se conceptualiza a los SIBO, se definen sus principales características, se realiza una propuesta metodológica para su construcción y se ofrece un esquema conceptual para la evaluación de los SIBO.

Los resultados obtenidos/esperados representan una continuidad de los presentados en [1].

Palabras clave

Ontologías, Ontologías de soporte, Calidad de una ontología, SIBO, Metodología y Evaluación de SIBO.

1. INTRODUCCIÓN

Generalmente, las ontologías se usan para especificar y comunicar el conocimiento del dominio de una manera genérica y son muy útiles para estructurar y definir el significado de los términos. A menudo, el término ontología se usa con diferentes significados [11] y con diferentes roles en el diseño y desarrollo de sistemas de información/conocimiento [4].

Las ontologías se están usando productivamente en distintas áreas de la Informática teórica (Ingeniería de Software, Ingeniería del Conocimiento, Ingeniería Web, Sistemas de Información, etc.), e Informática aplicada (Informática Médica, Informática Educativa, etc.).

En este subproyecto, se abordan las cuestiones que se presentan a continuación.

a) La calidad del proceso y del producto software con el uso de ontologías

Las ontologías se pueden usar en la Ingeniería de Software de dos maneras diferentes: como mecanismos o herramientas que facilitan el desarrollo de software, o para conceptualizar distintas áreas de conocimiento de esa disciplina.

b) Desarrollo, mantenimiento y reutilización de los Sistemas de Información

De acuerdo con algunos autores [5, 10, 17, 21], el uso de ontologías en el desarrollo de los Sistemas de Información (SI) contribuye a mejorar la calidad del producto final.

De esta manera, surge el concepto de Sistema de Información Basado en Ontologías (SIBO), que fue introducido por Guarino [10] y, aunque todavía están en una fase preliminar de desarrollo, se visualizan algunas aplicaciones concretas.

Las ontologías proveen los mecanismos para organizar y almacenar los componentes genéricos de los SI que incluyen esquemas de las bases de datos, objetos de interfaz de usuario y programas de la aplicación [10, 14, 20]. Por ello, los SIBO permiten recuperar, distribuir, integrar, interoperar y compartir la información/conocimiento de una manera más eficaz, efectiva y eficiente que los SI convencionales [2].

c) **Reutilización del conocimiento existente en dominios específicos**

Los desarrolladores de ontologías y/o SIBO requieren reducir los tiempos de desarrollo, por lo que precisan encontrar y comparar ontologías existentes con la finalidad de reusarlas.

Durante el proceso de validación y evaluación de ontologías, no se disponen de herramientas formalizadas que permitan determinar si la ontología construida o recuperada cumple con los requerimientos especificados en un dominio determinado.

Por ello, en esta línea se analizan los siguientes temas:

- Evaluación de la calidad de ontologías
- Análisis de usabilidad/reusabilidad de ontologías

En este trabajo se presenta una síntesis de las líneas de investigación y desarrollo, los resultados alcanzados/esperados y la formación de recursos humanos correspondientes al subproyecto denominado "Ontologías en los Sistemas de Información/Conocimiento".

2. LINEAS DE INVESTIGACION Y DESARROLLO

2.1. DESARROLLO Y EVALUACIÓN DE ONTOLOGÍAS EN DISTINTAS ÁREAS DE LA INFORMÁTICA TEÓRICA Y APLICADA

En esta línea se abordan las siguientes actividades:

a. Construcción de ontologías en distintas áreas de aplicación

▪ Ontología de soporte al proceso de la Ingeniería de Requerimientos

Reconociendo los problemas que acarrea una inadecuada Ingeniería de Requerimientos, y teniendo en cuenta los beneficios de usar ontologías [12], se desarrolla una ontología de soporte a este proceso, la cual asistirá a desarrolladores de software durante la ejecución del mismo.

Está basada en la metodología DorCu (Documentación de requerimientos Centrada en el Usuario) [3]. Para la construcción de la ontología se utiliza el entorno de desarrollo de Protégé 2000 y el lenguaje OWL.

▪ Ontología de soporte a la Ingeniería de la Usabilidad

En esta línea se construye un prototipo para la generación de interfaces universales basadas en ontologías. Este prototipo contempla un amplio rango de usuarios y considera múltiples contextos de uso en donde variados tipos de usuarios pueden llevar a cabo múltiples tareas, posiblemente en diversos dominios de interés [1, 7, 15].

b. Ontología de soporte al diagnóstico de trastornos de ansiedad

En el Área de Salud Mental del Hospital Independencia, para el diagnóstico y tratamiento de pacientes con trastornos mentales, cada médico aplica su conocimiento, basado generalmente en su experiencia, el cual es difícil de transferir y compartir. Para lograr un conocimiento compartido y ofrecer a los médicos una herramienta que les permita obtener diagnósticos más probables en base a síntomas, antecedentes, etc. que presenta el paciente, se desarrolla la "Ontología de Soporte al Diagnóstico de Trastornos de Ansiedad (ODTA)" [8].

c. Determinación de métricas e indicadores de la calidad de una ontología

En base a una investigación exploratoria y descriptiva que se realizó, y habiendo examinado diversas propuestas [9, 13, 18, 19], se define y operacionaliza la variable "calidad de una

ontología". Se determinan cuatro dimensiones y, en algunos casos, los indicadores que permitirían evaluar empíricamente la calidad de una ontología en un ámbito específico. Estas dimensiones enfatizan aspectos descriptivos, estructurales, funcionales y operativos de una ontología.

d. Análisis de la usabilidad y reusabilidad de ontologías

Se necesita que los desarrolladores dispongan de herramientas que les permitan seleccionar la mejor ontología entre un conjunto de ontologías candidatas a ser reusadas. En este proceso se han identificado los siguientes problemas específicos:

- Heterogeneidad de los formalismos de representación del conocimiento.
- Problemas semánticos (sinónimos, hiperónimos e hipónimos).
- Suposiciones ocultas (conocimiento que se considera implícito, rutinario o asociado al sentido común y no se incluye en la ontología).

Se plantea la definición de un conjunto de métricas que permitan validar y evaluar la usabilidad y reusabilidad de las ontologías disponibles en la web.

2.2. DESARROLLO Y EVALUACIÓN DE SIBO

a. Conceptualización y caracterización de los SIBO

Según el rol que las ontologías tienen en los SI, desde el punto de vista estructural, se pueden distinguir dos tipos de SIBO: a) la ontología forma parte de cada uno de los componentes estructurales; b) la ontología es un componente más del SI y coopera con los otros componentes para conseguir los propósitos del sistema [10].

En esta línea de investigación se conceptualiza a los SIBO y se definen sus principales características. Además, se realiza un estudio comparativo con los SI convencionales, de modo tal de resaltar las ventajas de los SIBO [2].

b. Metodología para el desarrollo de los SIBO

La necesidad de incluir las ontologías en la construcción de SI, para obtener los beneficios que éstas ofrecen, lleva a los desarrolladores a pensar en una estrategia para poder diseñar y construir SIBO.

Ante la ausencia de metodologías para construir SIBO, en esta línea se realiza una propuesta metodológica.

c. Evaluación de los SIBO

Los diseñadores de los SIBO se enfrentan con un conjunto de obstáculos, ya que la ontología representa el conocimiento de un dominio, generalmente, dinámico y cambiante. Ellos no sólo tienen que enfrentarse con el problema de la integración/interoperabilidad de los sistemas, sino también lograr que sean extensibles y adaptables a los cambios del dominio de aplicación.

Por ello, a los aspectos temporales propuestos por Guarino [10], se les agrega el "tiempo de uso" para soportar explícitamente los cambios en el dominio del sistema.

Para analizar la calidad de un SIBO se examinan diversas propuestas [6, 16], se establecen las dimensiones, las preguntas centrales que guían el proceso de evaluación, los principales responsables, el momento y los recursos mínimos necesarios para llevar a cabo la evaluación. Los recursos representan los elementos para la medición de cada uno de los indicadores de cada dimensión. Estos pueden ser: documentación del SIBO, productos parciales o definitivos del SIBO en desarrollo, o el SIBO en ejecución.

3. RESULTADOS OBTENIDOS/ESPERADOS

En la tabla 1, se sintetizan los resultados obtenidos por líneas de investigación. En la última columna, se incluye la formación de recursos humanos o la divulgación de los resultados y, en ambos casos, el estado de los mismos.

Tabla 1. Resultados obtenidos / esperados

LÍNEAS DE INVESTIGACIÓN / DESARROLLO	RESULTADOS OBTENIDOS / ESPERADOS	FORMACIÓN DE RECURSOS HUMANOS / DIVULGACIÓN
Desarrollo y evaluación de ontologías en distintas áreas de la Informática teórica y aplicada	Ontología de soporte al proceso de Ingeniería de Requerimientos	Tesina de grado. Carrera: Licenciatura en Sistemas de Información. FCEyT, UNSE. Alumnas: Viviana Manzanedo y Rita Abelleira. Anteproyecto en evaluación.
	Ontología para la generación de interfaces universales de usuario Web	Becas de investigación: "Prototipo para la generación interfaces universales de usuario web basada en ontologías". Alumnos: Mercedes Díaz y Luis Gustavo Chanferoni. En desarrollo.
	Ontología para el Diagnóstico de Trastornos Mentales	Tesina de grado. Carrera: Licenciatura en Sistemas de Información. FCEyT, UNSE. Alumna: Carolina Ger. Trabajo aprobado.
	Análisis de la usabilidad y reusabilidad de ontologías	Beca de investigación: "Análisis de la usabilidad y de la reusabilidad de ontologías disponibles en la web". Alumno: Federico Romano. En desarrollo.
	Métricas e indicadores de la calidad de una ontología	Artículo en proceso de evaluación para su publicación.
Desarrollo y evaluación de SIBO	Conceptualización y caracterización de los SIBO	Jornadas Chilenas de Computación. Noviembre de 2008, Punta Arenas, Chile.
	Metodología para el desarrollo de SIBO	Tesina de Grado. Carrera: Licenciatura en Sistemas de Información. FCEyT, UNSE. Alumna: Gladys Noemí Fortea. Anteproyecto en evaluación.
	Evaluación de la calidad de los SIBO	Enviado y aceptado para el IX Congreso ISKO, Capítulo Español. "Nuevas perspectivas para la difusión y organización del conocimiento", 2009. Valencia, España.

4. FORMACIÓN DE RECURSOS HUMANOS

En la tabla 1 del apartado anterior se incluye la formación de recursos humanos correspondiente al año 2008. No se incorporan los alumnos que están elaborando sus anteproyectos de tesinas en el marco de este subproyecto de investigación.

5. BIBLIOGRAFÍA

- [1] Álvarez, M.; Barchini, G.; Díaz, M.; Chanferoni, L. "Sistemas de Información Basados en Ontologías. Un Área Emergente". X Workshop de Investigadores en Ciencias de la Computación (WICC). General Pico, La Pampa. 5 y 6 de mayo de 2008.
- [2] Álvarez, M.; Palliotto, D.; Barchini, G. "Las Ontologías en la Nueva Generación de los Sistemas de Información". Jornadas Chilenas de Computación. Libro de Resúmenes. ISBN 978-956-319-507. Noviembre de 2008, Punta Arenas, Chile. Disponible en: <http://lahuen.dcc.uchile.cl/~jcc2008/libro.pdf>.
- [3] Báez, M. G.; Barba Brunner, S. I. "Metodología DoRCU para la Ingeniería de Requerimientos". IV Workshop em Engenharia de Requisitos. Buenos Aires, Argentina. 22 y 23 de noviembre de 2001. Disponible en: http://wer.inf.puc-rio.br/WERpapers/artigos/artigos_WER01/baez.pdf.

- [4] Barchini, G.; Álvarez, M.; Herrera, S. "Sistemas de Información: Nuevos Escenarios Basados en Ontologías". *Revista de Gestão da Tecnologia e Sistemas de Informação. Journal of Information Systems and Technology Management*, Vol. 3, No. 1, 2006, p. 3-18. ISSN online: 1807-1775.
- [5] Colomb, Robert M.; Weber, Ron. "Completeness and Quality of an Ontology for an Information System". *International Conference on Formal Ontology in Information Systems (FOIS'98)*. Trento, Italy, 6-8 June, 1998. In N. Guarino (ed.) *Formal Ontology in Information Systems*, IOS-Press (Amsterdam), p. 207-217. Disponible en: <http://www.itee.uq.edu.au/~colomb/Papers/Ontology.html>.
- [6] Fonseca, F.; Egenhofer, M. "Ontology Driven Geographic". *Information Systems Proceedings of the 7th ACM International Symposium on Advances in Geographic Information Systems*, p. 14-19. Kansas City, Missouri, United States. 1999.
- [7] Furtado E.; Furtado V.; Silva, W.; Rodrigues, D.; Taddeo, L.; Limbourg Q.; Vanderdonck J. "An Ontology-Based Method for Universal Design of User Interfaces". *Proceedings of Workshop on Multiple User Interfaces over the Internet: Engineering and applications Trends*. Disponible en: <http://www.cs.concordia.ca/~7Efaculty/seffah/ihm2001/program.html>.
- [8] Ger, S. C.; Barchini, G.; Álvarez, M. "Ontología de Soporte al Diagnóstico de Trastornos de Ansiedad". *JAIIO-2007*. Mar del Plata. Argentina.
- [9] Gruber, T. R. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". 1993. Disponible en: <http://citeseer.ist.psu.edu/gruber93toward.html>.
- [10] Guarino, N. "Formal Ontology and Information Systems". In *Formal Ontology in Information Systems, Proceedings of FOIS'98*, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15. Disponible en: <http://www.loa-cnr.it/Papers/FOIS98.pdf>.
- [11] McGuinness, D.L. "Ontologies Come of Age". Fensel, Hendler, Lieberman and Wahlster (eds), *Spinning the Semantic Web: Bringing the World Wide Web to its Full Potential*. MIT Press, 2002. Disponible en: [http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mitpress-\(with-citation\).htm](http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mitpress-(with-citation).htm).
- [13] Mostowfi, F.; Fotouhi, F. "Improving Quality of Ontology: An Ontology Transformation Approach". *Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06)*, 2006. Disponible en: <http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/10810/34089/01623856.pdf>.
- [14] Gangemi, A.; Pisanelli, D. M.; Steve, G. "Ontology Integration: Experiences with Medical Terminologies". *Formal Ontology in Information Systems (FOIS98)*. 1998.
- [15] Savidis, A.; Akoumianakis, D.; Stephanidis, C. "The Unified User Interface Design Method". Chapter 21, in "User Interfaces for All: Concepts, Methods, and Tools ". C. Stephanidis (ed.), Lawrence Erlbaum Associates, Pub. pp. 417-440. Mahwah. 2001.
- [16] Sheth, Amit P. "Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics". *Interoperating Geographic Information Systems*. M. F. Goodchild, M. J. Egenhofer, R. Fegeas and C. A. Kottman (eds). Kluwer, 1998. Disponible en: <http://lstdis.cs.uga.edu/library/download/S98-changing.pdf>.
- [17] Soares, A.; Fonseca, F. "Ontology-Driven Information Systems: At Development Time". *International Journal of Computers, Systems and Signals*, Vol.8, N° 2, 2007. Disponible en: <http://www.fullcycles.org/iaamsad/ijcss/Journals/Vol8No2/IJCSS%202007%202-7.pdf>.
- [18] Tartir, S.; Arpinar B.; Sheth, A. "Ontological Evaluation and Validation". 2007. Disponible en: http://knoesis.wright.edu/library/download/tart_bud_sheth07.pdf.
- [19] Tartir, S.; Arpinar B.; Sheth, A.; Aleman-Meza, B. "OntoQA: Metric-Based Ontology Quality Analysis". *Proceedings of IEEE ICDM 2005 KADASH Workshop*. Disponible en: <http://www.cs.uga.edu/~tartir/pubs/OntoQA.pdf>.
- [20] Viinikkala, M. "Ontology in Information Systems". Disponible en: <http://www.cs.tut.fi/~kk/webstuff/Ontology.pdf>.
- [21] Yildiz, B.; Miksch, S. "Ontology-Driven Information Systems: Challenges and Requirements". 2007. Disponible en: http://www.donauni.ac.at/imperia/md/content/departament/ike/ike_publications/2007/refereedconferenceandworkshoparticles/yildiz_2007_icsd_ontology_requiremen.pdf.

Consumo de Web Service desde dispositivos móviles heterogéneos

Rodríguez, Nelson; Martín, Adriana ; Valenzuela, Adriana; Chávez, Susana
Instituto y Departamento de Informática - F.C.E.F y N. - U.N.S.J.

CONTEXTO

Nuevas Tecnologías en el Desarrollo de Software

RESUMEN

La Web proporciona acceso a contenidos y también ofrece interacción y servicios, es decir distintas aplicaciones de software desarrolladas en diferentes lenguajes de programación y ejecutadas sobre cualquier plataforma pueden utilizar Web Service para consumir información. Los Web Service son una solución adecuada para lograr la interoperabilidad entre aplicaciones. Fueron diseñados originalmente para acceso desde estaciones fijas, y luego surgieron estrategias para permitir la interacción desde el móvil, lo que garantiza el acceso en cualquier momento y desde cualquier punto, dando así lugar a una mayor interacción del usuario con los servicios. No obstante la diversidad de equipos, sistemas operativos y software de desarrollo dificulta la total interoperabilidad.

Esta propuesta tiene por objetivo plantear estrategias para lograr la interoperabilidad de dispositivos móviles con plataformas Web, por medio de Web Service adaptables a diferentes tipos de dispositivos.

Palabras clave: Web service, dispositivos móviles, accesibilidad

1. INTRODUCCION

La Web proporciona acceso a contenidos y además ofrece interacción y servicios (comprar un producto, reservar un asiento en un vuelo, hacer una transferencia bancaria); conocido como Web Service -WS- básicamente una función o procedimiento que puede ser accedida vía Web por cualquier

programa o aplicación sin importar en que plataforma reside el servicio, o en qué lenguaje ha sido desarrollado.

Pueden encontrarse en las bibliografías numerosas definiciones sobre Web Service (WS) que varían desde conceptos altamente técnicos hasta conceptos demasiados simplistas. Será considerada la definición dada por el Consorcio de la World Wide Web (W3C), como punto de partida para entrar al gran mundo de los WS.

"A Web Service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards." [1].

De esta forma distintas aplicaciones de software desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los Web Service para intercambiar información en redes de computadoras como Internet.

La ventaja de utilizar Web Service reside en que estos aportan interoperabilidad entre aplicaciones de software independientemente de sus propiedades o de las plataformas de distintos fabricantes sobre las que se instalen. También los WS fomentan los estándares y protocolos basados en texto, que hacen más fácil acceder a su contenido y entender su funcionamiento. Es así que servicios y software de diferentes compañías ubicadas en diferentes lugares geográficos puedan ser combinados fácilmente para proveer servicios integrados.

Para lograr un alto grado de interoperabilidad, es necesario adoptar una serie de protocolos mediante los cuales se defina como se invoca el servicio, como se pasan los parámetros, como se recibe un resultado y como se manejan los errores, siendo los dos protocolos más usados SOAP (Simple Object Access Protocol) y XML-RPC (XML Remote Producer Call), sobre los cuales se establece el intercambio.

En la actualidad la industria desarrolla una gran cantidad de dispositivos que proveen variadas alternativas de comunicación, los mismos son producidos por distintos fabricantes e incluyen distintas tecnologías. El mercado se comporta de forma cambiante, y como consecuencia de ello, ni los expertos pueden visualizar el rumbo definitivo que tomarán las cosas. Es tal el nivel de sofisticación y de nuevas prestaciones que se ofrecen que la gran duda que surge es: ¿Es posible una definición de estándar que permita lograr la convergencia e integración de las aplicaciones heterogéneas? Si es así, ¿Qué pautas o requisitos debe cumplir?

La cantidad de celulares y smartphones en el mercado móvil según la Comisión Nacional de Comunicaciones es de 46.508.774 (Diciembre 2008). La modalidad de pre-pago involucra el 90% y la post-pago 10% . La teledensidad de terminales móviles en enero de 2009 es de 118 [2]. A estos valores hay que agregarles un número menor pero importante de dispositivos habilitados para Internet como PDAs y Pocket PC. Esto indica un mercado potencial muy grande que requerirá el desarrollo de soluciones móviles.

El no contar con un dispositivo marco sobre el cual realizar las configuraciones necesarias y la variedad de equipos dentro de un mismo segmento, dificulta la construcción de software que pueda ser ejecutado en la mayoría de las plataformas.

Los SMS, MMS, juegos, música, videos y ringtones continuarán generando ingresos para los operadores móviles o los generadores de

contenido. Las aplicaciones de mensajería móvil, como los mensajes instantáneos y el e-mail tienen buen potencial de crecimiento, así como también la integración de las aplicaciones de escritorio con otras soluciones móviles. Por otro lado, la TV móvil y los servicios de localización (GPS) serán soluciones de nicho en el corto plazo, hasta que la integración de las tecnologías, contenidos y modelos de negocio maduren. En cuanto a las soluciones móviles de banca (hoy basados en SMS) , las aplicaciones de pago (m-payment) y el m-learning necesitan sobreponerse de los problemas de usabilidad y la inhibición de los usuarios.

En este sentido, muchos operadores móviles necesitan introducir nuevos servicios para aprovechar la capacidad no utilizada de sus plataformas de tercera generación (3G) y para compensar la volatilidad de los ingresos por llamadas. Algunos de esos servicios, como las video llamadas están siendo utilizadas, pero al no mantener compatibilidad con la segunda generación su alcance está limitado. En este sentido las soluciones que necesitan de mayores anchos de banda, requieren de nuevos modelos de negocio y una substancial inversión en infraestructura y cobertura.

Según el informe publicado sobre smartphones el 11 de marzo de 2009 por Gartner, todavía el mercado durante 2008 tiene el 89% dividido en 4 sistemas operativos diferentes. Debido a ello el grupo de investigación para sus desarrollo ha constituido dos subgrupos: plataforma .NET (en particular Windows Mobile) y plataforma Java (JME) [3].

Cabe destacar que el segmento de mercado que corresponde a las tecnologías asociadas a este proyecto, a pesar la crisis mundial, ha experimentado un crecimiento en ventas de smartphones y ventas por Internet, entre otras.

La convergencia entre teléfonos celulares y dispositivos móviles (handhelds o PDA's), o sea el smartphone es una tendencia que poco a poco ya está siendo una realidad. Por otro lado, la convergencia entre dispositivos

móviles y PCs, es una tarea pendiente que tiene sus primeros alcances con los Tablet PC y más recientemente con las netbooks.

El surgimiento de nuevos tipos de usuario final, tanto como el aumento del número de conexiones de usuarios a servicios de banda ancha 3G, da la oportunidad de proveer acceso a documentos y servicios residentes en el dominio del hogar o del trabajo desde ubicaciones remotas.

Las investigaciones en computación móvil de la última década han favorecido el desarrollo de sistemas inalámbricos de comunicaciones con mejor rendimiento y calidad de servicio, así como la construcción de plataformas de software móvil más amigables, económicas y adaptables. Los avances tecnológicos en dispositivos móviles tales como Pocket PC's, smartphone y teléfonos móviles han modificado notablemente la forma en que los usuarios interactúan con entornos computacionales así como los contextos de uso.

Los dispositivos móviles ocupan un lugar importante en la sociedad. Se ha extendido el uso de celulares a adolescente y niños. La mayoría de las personas se trasladan con su móvil y en algunos casos con su PDA, u otro dispositivo, los cuales se encuentran en el mercado en una variedad de formas y de marcas. Además presentan importantes diferencias como el sistema operativo, el tamaño de la pantalla, la facilidad de comunicación, la disponibilidad de memoria, etc. Al incrementar su capacidad de procesamiento y funcionalidad, estos equipos resultan verdaderas computadoras con algunas limitaciones.

Esto ha propiciado el surgimiento de nuevas perspectivas de negocio lo cual se plasma en el desarrollo de nuevos servicios así como el de arquitecturas que soporten los mismos.

En un informe publicado por el MIT sobre las diez tecnologías que cambiarán el mundo (diciembre 2006) [4] cita entre ellas a redes de sensores sin cables y software fiable que de alguna manera mantiene puntos de contacto

con los objetivos de este proyecto. Mas recientemente la consultora Gartner [5] publicó las que considera las diez (10) tecnologías Top para 2008, entre las que cita a: Web Platform & WOA y la Real World Web (la Web del mundo real). La primera incluye arquitecturas orientadas a servicios y la segunda promete la inclusión de una diversidad de dispositivos inalámbricos, cámaras filmadoras que se combinarán para analizar la ubicación, la intención y hasta las ecuaciones a través de la red. Esto obviamente aumenta la realidad en la Web. Cabe hacer notar que en el informe 2007 esta consultora consideró tecnologías claves a: la computación ubicua y la Web 2.0.

2. LINEAS DE INVESTIGACION y DESARROLLO

El proyecto origen de este trabajo, tiene como unidades ejecutoras el Departamento de Informática y el Instituto de Informática de la F.C.E.F.y N. U.N.S.J, y persigue como propósito lograr la interoperabilidad de dispositivos móviles con plataformas Web, por medio de Web Service adaptables a diferentes tipos de dispositivos.

Representa un desafío importante adecuar los lineamientos para diseñar un esquema para un Web Service que permita satisfacer necesidades, preferencias y situaciones de los diferentes usuarios interesados en obtener este servicio. Para ello, es esencial que este esquema propuesto sea flexible de manera que proporcione igualdad de oportunidades para todas las personas que necesiten percibir, entender, navegar e interactuar con la Web.

Respecto a las tecnologías para móviles, además de contemplar cuestiones inherentes a la usabilidad, accesibilidad y ergonomía, es fundamental investigar las limitaciones que distintas arquitecturas presentan y proponer soluciones alternativas que permitan el acceso y gestión de servicios. Lo citado se consolida como una de las líneas principales de investigación y desarrollo del proyecto marco de este trabajo.

Como pre-requisito para lograr la interoperabilidad del acceso a WS desde el móvil, se están investigando diversas alternativas. Una de ellas es considerar la construcción de un gateway como un acceso móvil a redes de ambientes [11], solo que en este caso no enfocado a la eficiencia sino a la adaptabilidad del servicio a las diferentes plataformas móviles. La otra posibilidad es evaluar la construcción de un framework que permita manejar adecuadamente la diversidad de móviles, como ya existen antecedentes [12] [13].

Se debe destacar que el grupo de trabajo viene desarrollando sus actividades como un grupo consolidado desde 1999 en proyectos afines que sirvieron como antecedentes al presente proyecto haciendo uso de tecnología móvil, Java y aplicaciones Web entre otras.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se han realizado varias publicaciones [6] [7] [8] [9] [10] durante el presente año. Cabe hacer notar que esta nueva línea de investigación tiene recién un año de desarrollo.

Se espera además avanzar en varios planteos sobre la interoperabilidad que deben ser aun considerados. Tal es el caso de manejo de errores en la comunicación de las partes y la necesidad de firmar y encriptar la llamadas a SW (WS-Security).

Con los dispositivos móviles se presenta un problema. Los algoritmos a aplicar deben ser lo suficientemente robustos para proveer seguridad, pero a su vez lo bastante “livianos” para ejecutarse en un móvil. Estrategias como tarjetas de cifrado o esquemas biométricos están disponibles pero solo para un número reducido de dispositivos y el objetivo es lograr una estrategia que sea lo más amplia posible. Los subgrupos de investigación están abocado al estudio de autenticación de servicios y la evaluación de los mismos, en distintas plataformas.

También se está trabajando en la posibilidad de agilizar el proceso de registración de los Web Service.

Como grupo referente en tecnología móvil, el gobierno de la provincia ha seleccionado al equipo de trabajo para capacitar en tecnología móvil .NET y ASP.NET para mantenimiento y modificación del portal de gobierno.

El proyecto marco de este trabajo: “Arquitectura de acceso a Web Service desde dispositivos móviles heterogéneos” ha establecido lazos de colaboración con el grupo de Ingeniería Telemática de la Universidad de Murcia. Este grupo participa de dos proyectos que son:

Popeye (Professional peer environment beyond edge computing) cuyo objetivo es proveer conceptos, métodos y servicios de núcleo para la generación de ambiente de trabajo colaborativo móvil con énfasis en el intercambio P2P.

Ecospace es otro proyecto de la unión europea que tiene entre sus objetivos principales el diseño y desarrollo de un estándar abierto, de una arquitectura orientada a servicios para sistemas semejantes y complementarios.

4. FORMACION DE RECURSOS HUMANOS

En el ámbito del proyecto antes citado se han desarrollado cuatro tesinas de grado vinculadas al desarrollo de una herramienta para marketing de proximidad (A. Ene, C. Martinez ; Lic.en Ciencias de la Información) y Servicios Web aplicados a Salud.(J. Reta, L. Espin; Lic. en Ciencias de la Información). Se encuentran en desarrollo dos tesinas sobre la implementación de un Web Service con funcionalidades en lenguajes de señas, así como el desarrollo de aplicaciones móviles que consuman dicho servicio. Dos tesinas sobre la implementación y utilización de un Web Service de Noticias

Además, está en etapa inicial el desarrollo de una tesis de postgrado que implica investigaciones sobre accesibilidad a dispositivos móviles por personas con limitaciones auditivas.

También, como extensión del proyecto se dará soporte a los profesionales encargados del desarrollo del sistema del juzgado de faltas, a través del dictado de cursos en tecnología .NET y móvil para .NET.

5. BIBLIOGRAFIA

[1] Definición de Web Service – W3C - <http://www.w3.org/TR/ws-arch/>

[2] Estadísticas Telefonía Móvil - Comisión Nacional de Comunicaciones - www.cnc.gov.ar/indicadores/estadisticas/movil.asp

[3] <http://www.gartner.com/it/page.jsp?id=754112>

[4] <http://www.technologyreview.com/specialreports/specialreport.aspx?id=37>

[5] <http://www.gartner.com/it/page.jsp?id=530109>

[6] S. Chavez, A. Martín, L. Rojas, P. Araya - Diseño de Servicios Web en java - Workshop de actualización en telecomunicaciones y Gestión del Conocimiento en Bibliotecas- 2008

[7] N. Rodríguez, L. Espín, J. Reta - Webservice aplicado a la salud. Información globalizada mediante SOAP, PHP y web2.0 – Workshop actualización en telecomunicaciones y Gestión del Conocimiento en Bibliotecas- 2008.

[8] A. Valenzuela, P. Domínguez, M. Varela Desarrollo de un Servicio Web en .NET – Workshop actualización en telecomunicaciones y Gestión del Conocimiento en Bibliotecas - 2008.

[9] N. Rodríguez ; S. Villodre - Posibilidades Tecnológicas de los servicios Web aplicados a M-Learning – 5ta Jornada de Informática y Educación- Univ. de Villa María-Cordoba-2008.

[10] A. Martin, S. Chavez - Desarrollo de Aplicaciones Móviles con NetBeans – 1º Congreso Nacional de Software Libre - 2008.

[11] Reyes, Messeguer, Royo, Homar - Mobile Access to the Service in ambient Networks - Intelligent Environments, 2008 IET 4th International Conference on.

[12] Chen, Liu, Kao - A New Framework for Mobile Web Service - Applications and the Internet (SAINT) Workshops, 2002. Proceedings. 2002 Symposium on.

[13] Luqun Li - An integrate Web Service Framework for Mobile Device Hosted Web Service and Its Performance Analysis - High Performance Computing and Communications, 2008. HPCC '08. 10th IEEE International Conference on

Integración de Bases de Datos Heterogéneas para Servicios Web en Gobierno Electrónico

Lic. Angela Cesetti - Dr. Pablo Fillotrani - Mg. Elsa Estevez

Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Av. Alem 1253 - 8000 Bahía Blanca - ARGENTINA
phone: +54 291 459 5100 - ext 2613, 1031
fax: +54 291 459 5136
e-mail: [abc, prf, ece]@cs.uns.edu.ar

1. Gobierno Electrónico

Gobierno Electrónico (GE) se define como el uso de las Tecnologías de la Comunicación y la Información (TICs) en ambientes de gobierno como una plataforma para el intercambio de información. Para esto, provee servicios y permite realizar transacciones con ciudadanos, negocios y otras ramas de gobierno. El GE ha evolucionado en aplicaciones más transaccionales e integradas y los gobiernos han comenzado a agregar recursos organizativos y tecnológicos.

Los servicios ofrecidos están pensados para tres grupos que interactúan: usuarios, negocios y gobierno. Algunas características que el GE debería proveer, para lograr una comunicación eficiente son:

- mejorar el acceso de los ciudadanos a los servicios de GE y al uso de la información
- ofrecer privacidad y seguridad

Actualmente, es una realidad que la mayoría de los gobiernos ofrecen un portal donde agrupan determinada información: en principio solo aquella de carácter informativo. El siguiente paso, no del todo desarrollado solo en casos muy particulares, sería ofrecer servicios a partir de un portal, que permita realizar diversas transacciones.

Otro aspecto relacionado y que merecerá un buen análisis es como afectará la participación ciudadana en las decisiones de gobierno, a partir de la publicación de la información por parte de las entidades.

A otro nivel, existen preguntas del dominio del GE concernientes al reto que implica lograr interoperabilidad entre las bases de datos heterogéneas existentes, dadas las diferencias semánticas de interpretación; la diversidad de leyes, regulaciones, servicios ciudadanos, procesos administrativos; la educación e incorporación de mas ciudadanos en el uso de estas aplicaciones y como este uso afectará a las aplicaciones web.

Dada la interacción entre los ciudadanos y los servicios de GE, resultará en evolución y beneficios para ambos.

2. Web Semántica

La *Web Semántica* (WS) surge como consecuencia de la necesidad de diferentes grupos (gobiernos, negocios, ciudadanos) de proveer un ambiente que pueda permitir fácil acceso a la información, servicios de transacción e interoperabilidad entre aplicaciones. En un principio, la idea básica detrás de la WS es que las páginas web se identifiquen como recursos informativos, identificando el tipo de información que incluye la página web y clasificarlas utilizando lenguajes de marcado semántico sobre las páginas web. Los beneficios de esta mejora, es que facilitan buscar y encontrar cualquier clase de información.

La WS involucra la publicación de datos en un lenguaje, Resource Description Framework (RDF) y esto permite categorizar los datos según la percepción humana, pero permitiendo ser comprendida por las computadoras. Por ejemplo, la World Wide Web (WWW) se basa principalmente en documentos escritos en Hypertext Markup Language (HTML), un lenguaje de marcado que codifica el cuerpo de un texto intercalado con objetos multimedia.

3. Servicios Web

La W3C define a un *Servicio Web* (SW) como 'un sistema de software diseñado para soportar interacción interoperable entre máquinas sobre la red'. Los SW frecuentemente son solo Interfaces de Programación de Aplicaciones (APIs) web, que suelen accederse desde Internet, y los servicios requeridos se ejecutan sobre un sistema de hosting remoto. Otras alternativas con casi igual funcionalidad son Object Management Group's (OMG), Common Object Request Broker Architecture (CORBA), Distributed Component Object Model (DCOM) de Microsoft o Java/Remote Method Invocation (RMI) de SUN.

Los SW apuntan a la comunicación del tipo cliente/servidor sobre la Web, mediante el uso del protocolo HTTP. Los servicios que se ofrecen caen en uno de dos grandes grupos: el primero utiliza mensajes de Lenguajes de Marcado Extensible (XML), que toma como estándar el Protocolo de Acceso de Objetos Simples (SOAP), donde, en general, existe una descripción machine-readable con las operaciones ofrecidas por el servicio escritas en Lenguaje de Descripción de Servicios Web (WSDL). Este lenguaje no es un requisito para un endpoint de SOAP, pero es un prerequisite para la generación automática de código del lado del cliente en muchos frameworks SOAP, Java y .NET.

4. Integración de Bases de Datos mediante el uso de ontologías

Una de las cuestiones más importantes del GE es como lograr la interoperatividad de las distintas bases de datos. Para esto, analizaremos el uso de las ontologías.

Una *Ontología* es la representación de un conjunto de conceptos dentro de un dominio y la relación que mantienen. Los componentes comunes de las ontologías incluyen: instancias u objetos, clases, atributos y sus relaciones, reglas que describen las aserciones de inferencias lógicas de forma lógica que incluye la ontología de su dominio de aplicación.

Algunos de los componentes para desarrollar aplicaciones exitosas son:

- Lenguajes de marcado: XML, RDF, OWL
- Herramientas de Edición/marcado: para la construcción y uso de las ontologías
- Motores de inferencia: para deducir nuevo conocimiento del conocimiento ya especificado

Las tecnologías de SW y los ambientes de desarrollo relacionados proveen una base para el desarrollo de un sistema de aplicación. Sin embargo, el desarrollo de sistemas basados en semántica aumenta las preguntas concernientes a los aspectos técnicos y organizativos que todavía necesitan resolverse. Existen diferentes respuestas dependiendo del dominio y del tipo de aplicación.

Hay que considerar algunas cuestiones de la SW, como:

- La experiencia frustrante del usuario,
- Los problemas con la interoperabilidad, causado por diferencias en los esquemas conceptuales de dos aplicaciones que intentan cooperar,
- La administración pobre de documentos, y
- Las barreras en la recuperación de información: estandarización de metadatos.

Para una administración exitosa de los proyectos en SW, se debe considerar:

- El costo/beneficio organizativo,
- La participación del usuario,
- La integración técnica, y
- La estrategia de implementación.

Aquí, es importante destacar los avances que se están realizando en el mundo sobre el desarrollo de infraestructuras que soporten lo anteriormente mencionado, por ejemplo:

- *Providing Integrated Public Services to Citizens at the National and Pan-European level with the use of Emerging Semantic Web Technologies. SemanticGov Project* FP6-2004-IST-4-027517 is funded by the European Commission within the INFORMATION SOCIETY TECHNOLOGIES (IST) Programme. **SemanticGov**¹ is a thirty six month EU-funded research and development project that aims at building the infrastructure (software, models, services, etc) necessary for enabling the offering of semantic web services by public administration (PA).
- *Data, Information and Process Integration with Semantic Web Services (Project DIP*²): DIP's objective has been to develop and extend Semantic Web and Web Service technologies in order to produce a new technology infrastructure for Semantic Web Services (SWS) - an environment in which different web services can discover and cooperate with each other automatically. DIP's long term vision is to deliver the enormous potential benefits of Semantic Web Services to e-Work and e-Commerce.
- The *Cabinet Office*³ of the UK government has based its technical guidance in the eGovernment Interoperability Framework (eGIF), which was first issued in 2000, and updated to its version 6.1 in March 2005. eGIF mandates sets of specifications and policies for joined-up and web enabled government. It covers four areas: interconnectivity, data integration, e-services access and content management. The eGIF contains a Technical Standard Catalogue, which is revised and updated every six months.
- The French *ADAE*⁴, formerly known as ATICA, published "Le Cadre Commun d'Interoperabilité" (CCI) in January 2002 and its latest version (2.1) in September 2003. CCI comprises the recommendations for strengthening public electronic systems coherence and for enabling multi-agency electronic service delivery.

¹ <http://www.semantic-gov.org/>

² <http://dip.semanticweb.org/>

³ <http://www.cabinetoffice.gov.uk/e-government/>

⁴ <http://www.adae.gouv.fr/>

5. Portales Web Semántica

Un *portal web* (PW) es un punto de entrada para la presentación de información e intercambio sobre Internet, usado por una comunidad de interés. Es por eso, que es necesario ofrecer un soporte eficiente para la comunicación y que permita compartir la información.

Actualmente, las tecnologías Web presentan diversas limitaciones en la búsqueda, acceso, extracción, interpretación y procesamiento de información. Estas limitaciones atentan contra los procesos de comunicación y el compartir de la información entre los miembros de la comunidad. Para resolver estos inconvenientes, se pueden hacer uso de las tecnologías de SW y esto permitirá que los portales actuales evolucionen en PW mejorados semánticamente. Bajo este contexto, resulta fundamental el uso de las ontologías, ya que permiten modelar la estructura del portal.

Existen algunos temas en los cuales es importante ahondar, por ejemplo:

- los portales semánticos no emplean múltiples ontologías y no están diseñados para interoperar con portales relacionados, con lo cual se hace difícil compartir información. No hay que olvidar que uno de los principales objetivos es lograr la interoperabilidad. Los portales WS deberían evolucionar en portales P2P, para permitir compartir información y mejorar las capacidades de exportación/importación de la información. Es fundamental que estos portales permitan manejar la heterogeneidad en términos de múltiples ontologías y de múltiples portales.
- no existe la capacidad de mantener versiones. En un ambiente tan dinámico como la Web, las ontologías pueden cambiar en el tiempo. Es por eso que resulta necesario lograr que los portales WS utilicen mecanismos de versión para mantener la información consistente, permitiendo que los cambios que se realicen se puedan rastrear.
- actualmente, los portales no presentan suficiente funcionalidad para soportar SW entre portales. Lo que se debería considerar en un portal SW para obtener una funcionalidad completa es:
 - Los portales existentes y las tecnologías de administración de contenidos deberían ser reutilizables,
 - Proveer diferentes niveles de búsqueda: búsqueda basada en palabras claves, basadas en ontologías, agregando diferentes niveles de flexibilidad y haciendo un uso apropiado de las tecnologías de la SW para mejorar dichas búsquedas.
 - Apuntar a la estandarización de los portales SW, para permitir alcanzar un alto grado de comunicación y la reutilización de estructuras de conocimiento.
 - Comenzar a trabajar en la incorporación de Servicios de SW. Esto representará un punto clave para alcanzar interoperatividad entre portales.

Existen herramientas que permitan evaluar los portales, para detectar sus debilidades y luego analizar cuales estrategias se usarán para su mejora. Una aproximación que se puede utilizar es la que se ofrece desde la perspectiva funcional, que muestra un esquema de evaluación de portales de tres capas:

- *Capa de acceso a la información*: representa las características que permiten la interacción usuario-sistema.
- *Capa de procesamiento de la información*: muestra la forma en la cual el portal procesa los ítems de información: creación, publicación, organización, acceso y mantenimiento.
- *Capa base*: abarca todas las tecnologías necesarias para soportar las características de las capas superiores, tanto las tecnologías de Semántica Web, como las tecnologías de sistema.

6. Propuesta de trabajo

En estos momentos, se está trabajando para la incorporación de GE junto con diversos servicios web en la Secretaría de Relaciones Internacionales y Planeamiento, de la Universidad Nacional del Sur. A partir de los resultados que se obtengan, se evaluarán viabilidad y plazos para la implementación paulatina en el resto de los sectores de la Universidad.

A continuación se mostrarán algunos resultados obtenidos.

- Existe hoy en la Universidad el interés y el poder político que se necesita, para llevar a cabo todas las acciones necesarias para avanzar sobre los objetivos que propone el Gobierno Electrónico. En mi humilde entender, es uno de los aspectos mas importantes para poner las cosas en marcha. Además,
 - En estos tiempos resulta de suma importancia concebir nuestra Universidad como un ente competitivo. Esto se puede lograr mediante el uso de las TICs en su amplia definición.
 - Existe hoy ya una infraestructura informática instaurada con diversos niveles de seguridad, que permiten un buen funcionamiento de todas las dependencias (desarrollos SIU).
 - La Universidad posee la inmejorable oportunidad debido a su área de influencia, de poder estar en contacto con muchas localidades que están involucrados distintos emprendimientos. Además, la ciudad posee un puerto muy importante. En estos momentos, se deben aprovechar las posibilidades económicas y académicas que ofrece esta situación, pudiéndose potenciar.
 - Recientemente se ha logrado que el alumnado se inscriba a las materias desde la página web de la Universidad, siendo este un gran logro, puesto que debido al gran aumento del alumnado, el sistema que se utilizaba hasta hace poco se había transformado en un cuello de botella.
- En la actualidad, es fundamental que la Universidad, que goza de gran prestigio académico, no se quede fuera de los avances no solo tecnológicos sino académicos que puede aportarle la incorporación de GE a su estructura funcional completa. Se apunta a lograr obtener gran beneficio en distintos niveles, como por ejemplo,
 - Permitir el acceso a información pública
 - Lograr menores tiempos en la búsqueda de información
 - Lograr disminuir los tiempos de compras y contrataciones, para así además, lograr mayor transparencia y permitir que mas empresas puedan participar de los procesos.
 - Como consecuencia de lo demás, lograr mayor ahorro y eficiencia de los costos operativos (papel), tiempo de los empleados, horarios de acceso puesto que la pagina estaría disponible 24/7. Así, también se favorecería la operatividad de los empleados, pudiéndose reasignar recursos tanto humanos como económicos.
 - Al difundir más información concerniente a la Universidad, se espera, una mayor participación del alumnado, docentes e investigadores en cuestiones que nos atañen y afectan.
 - La Universidad además de proveer instrucción a la comunidad, podría comenzar a ser rentable: al mostrar todas las investigaciones que se están realizando, lograr patentes conjuntas con empresas que permitan, por ejemplo, obtener fondos para equipamiento, pago de becarios, etc. Y de esta manera lograr financiamiento para otros proyectos.
 - Lograr una proyección internacional que permita que la Universidad participe de más redes y proyectos internacionales.
 - Lanzar en mayor escala la plataforma de Educación a Distancia, que posee la Universidad con una plataforma de desarrollo propio.

- Es necesario implementar más servicios los necesarios para poder obtener feedback de todas las personas que utilizan el portal de la Universidad, que puedan dejar sus inquietudes, sabiendo que serán atendidos y tomados en cuenta a la brevedad.
- Resulta de gran necesidad, para la incorporación exitosa de SW, lograr la interoperabilidad de todas las bases de datos que se manejan en la Universidad.

Son muchas las expectativas que se tienen respecto a los resultados que se puedan alcanzar con la migración a un sistema que aplique las ventajas que ofrece el GE. Se espera que se termine con al primera etapa, a fin de este año, que consta de la implementación de la página web de la Secretaría. Luego de analizar las herramientas que se utilizarán para el desarrollo de la misma, más la implementación de la mayor cantidad de lineamientos que atañen al GE, se espera en mediano plazo, alrededor de 5 años, la compra de una infraestructura especial y comienzo de la migración de todo lo que se encuentra implementado, a la nueva plataforma.

A largo plazo, alrededor de 10 años, se espera incrementar los servicios que hoy brinda la Universidad a todos los actores (personal administrativo, alumnos, docentes, investigadores, empresas).

Para que este proceso sea un éxito, será necesario implementar diversos cursos de perfeccionamiento en las nuevas herramientas, para todos los niveles.

Se espera que con la incorporación de las TICs en todos los niveles de la estructura universitaria, se logren mejorar el funcionamiento de la estructura de Universidad y alcanzar mayor transparencia en todos los aspectos, internos y externos. Además, que los tiempos de operación se reduzcan, con una consecuente disminución de los gastos operativos.

Así, se podrá poner más énfasis en el crecimiento parejo de todas las áreas y ampliar la participación de la Universidad no solo a nivel académico sino también a nivel económico. Algunos ejemplos podrían ser, lograr la incorporación de nuestros alumnos a emprendimientos de la ciudad y la región, incrementar la movilidad estudiantil a nivel de grado y posgrado, obtener mas patentes que sean rentables, etc.

7. Bibliografía

- Electronic Government from Ispedia. J. Ramón Gil-Garcia
- Digital government. Gary Marchionini, Hanan Samet, Larry Brandt
- Semantic Web for e-government. Ralf Klischewski
- Semantic Web Technologies for Information Management within e-Government Services. Ralf Klischewski, Martti Jeenicke
- An Infrastructure for E-government Based on Semantic Web Services. Liuming Lu, Guojin Zhu, Jiaxun Chen.
- Infrastructure for E-government Web Services. Brahim Medjahed, Abdelmounaan Rezgui, Athman Bouguettaya, Mourad Ouzzani.
- Customized Delivery of E-Government Web Services. Brahim Medjahed, Athman Bouguettaya.
- Directions in E-Government: Processes, Portals, Knowledge. Roland Traunmüller, María Wimmer.
- Top down or Bottom up? How to establish a common ground for semantic interoperability within e-government communities. Ralf Klischewski.

- A Semantic Web Service-based Architecture for the Interoperability of E-government Services. Alessio Gugliotta, Liliana Cabral, Jhon Domingue, Vito Roberto, Mary Rowlett, Rob Davies.
- Semantic Web Enables E-Government Services. Brahin Medjahed, Athman Bouguettaya, Mourad Ouzzani.
- The Semantic Electronic Government: Knowledge Management for citizen relationship and new assesment scenarios. Miltiadis D. Lytras.
- WSMO-PA: Formal Specification of Public Administration Service Model on Semantic Web Service Ontology. Xia Wang, Tomas Vitvar, Vassilios Peristeras, Adrian Mocan, Sotirios K. Goudos, Konstantinos Tarabanis.
- IRS-III: A Broker for Semantic Web Services based Applications. Liliana Cabral, John Domingue, Stefania Galizia, Alessio Gugliotta.
- E-government administrative and semantic cooperation: the intelligent document role. F. Corradini, A. Polzonetti, R. Pruno.
- Exploring e-Government Evolution. The influence of Systems of Rules on Organizational Action. J. Ramón Gil-Garcia, Ignacio J. Martinez-Moyano.
- E-Government Cross-Agency and Intergovernmental Initiatives Research Project: Web Survey Results. Jane E. Fountain, Robin McKinnon, Eunyun Park.
- Survey of Graph Database Models. R. Angles, C. Gutierrez.
- Queryng RDF Data from a Graph Database Perspective. R. Angles, C. Gutierrez.
- Foundations of Semantic Web Databases. C. Gutierrez, C. Hurtado, A. O. Mendelzon.
- Minimal Deductive Systems for RDF. S. Muñoz, J. Perez, C. Gutierrez.
- The Semantics of Complexity of SPARQL. J. Perez, M. Arenas, C. Gutierrez.
- An Evaluation of Semantic Web Portals. Ruben Lara, Sung-Kook Han, Holger Lausen, Michael Stollberg, Ying Ding, Dieter Fensel.
- Semantic Web Portals - State of the Art Survey. Holger Lausen, Michael Stollberg, Ruben Lara Hernández, Ying Ding, Sung-Kook Han, Dieter Fensel.

Computación Evolutiva y Aprendizaje Automático para la Inferencia, Modelado y Simulación de Redes Regulatorias de Genes

Carballido Jessica Andrea, Ponzoni Ignacio, Gallo Cristian Andrés
Laboratorio de Investigación y Desarrollo en Computación Científica (LIDeCC)
Dto. de Cs. e Ing. de la Computación (DCIC) - Planta Piloto de Ingeniería Química (PLAPIQUI)
Universidad Nacional del Sur

CONTEXTO

La línea de investigación aquí presentada constituye una de las tres grandes líneas de investigación abarcadas por el laboratorio de investigación LIDeCC (Laboratorio de Investigación y Desarrollo en Computación Científica), dirigido por la Dra. Nélidea Beatriz Brignole. Actualmente, los doctores Ignacio Ponzoni, Jessica Carballido y el becario doctoral Cristian Gallo están trabajando en estos temas. Es una línea de investigación que se encuadra en el área de Bioinformática, y que cuenta con el soporte de varios proyectos de investigación¹.

RESUMEN

Los alcances principales de esta línea de I/D consisten en diseñar técnicas computacionales que asistan a expertos en bioinformática en la obtención de nuevos conocimientos sobre el funcionamiento de los mecanismos de regulación existentes a nivel molecular en los organismos biológicos. Más específicamente, se busca desarrollar sistemas de software que asistan en la reconstrucción (o descubrimiento) de la estructura relacional presente en las redes regulatorias de genes.

Palabras clave: bioinformática, computación evolutiva, aprendizaje automático, redes regulatorias de genes

INTRODUCCION

El advenimiento de nuevas tecnologías en el área de genómica, tales como los *microarrays* [Sohler, 2006], han posibilitado la obtención de grandes volúmenes de información relativas al funcionamiento

molecular de seres vivos. Esto ha derivado en una verdadera revolución del conocimiento en el campo de la Biología Molecular, la cual ha dado origen a una nueva disciplina, la Bioinformática. En la actualidad, gran parte de los esfuerzos llevados adelante en esta área apuntan a generar modelos computacionales que permitan extraer conocimiento a partir de la gran cantidad de datos biológicos disponibles actualmente [Schlitt y Brazma, 2007]. En este contexto, el uso de técnicas de aprendizaje automático [Bishop, 2006] y computación evolutiva [De Jong, 2006] está teniendo gran impacto en el diseño de modelos predictivos orientados a facilitar el descubrimiento automatizado de nuevo conocimiento biológico [Cios *et al.*, 2007; Handl *et al.*, 2007].

En el caso particular de la línea de I/D aquí presentada, el principal objetivo consiste en el desarrollo de software para la inferencia, validación y simulación de redes regulatorias de genes, tema de vital y actual relevancia en el campo de la genómica funcional [Schlitt y Brazma, 2007]. Una red regulatoria de genes (GRN, por su sigla en inglés) representa las relaciones de alto nivel que gobiernan las tasas de transcripción (regulación) de los genes en ARNm (ácido ribonucleico mensajero). Esta interacción puede representarse mediante una red (grafo), donde los nodos identifican genes cuyos niveles de expresión son regulados por la acción (activadora o inhibitoria) ejercida por otros genes de la misma red. Dichas interacciones son denotadas mediante aristas que interconectan los nodos del grafo. Descubrir y estudiar la estructura de las GRNs de diferentes organismos resulta de fundamental importancia para avanzar en la comprensión del funcionamiento biológico de los seres vivos, y posee innumerables aplicaciones tanto en

¹ **PICT 2006 Categoría B.** Tema: Desarrollo teórico de técnicas de Computación Evolutiva basadas en dominancia de Pareto para Optimización Combinatoria Multiobjetivo. Código: 1652. Director: Dra. Jessica A. Carballido. **PGI-UNS** Tema: *Aplicaciones de Computación Científica.* **PGI-UNS** Tema: *Técnicas de Aprendizaje Automático y Computación Evolutiva aplicadas al Diseño de Modelos Predictivos en Bioinformática.*

medicina (por ejemplo, el desarrollo de nuevos medicamentos [Huang, 2002]), como en biotecnología (por ejemplo, el diseño de nuevas variantes de cultivos [Yamaguchi-Shinozaki y Shinozaki, 2006]). Esta última aplicación se encuadra dentro de las áreas de investigación prioritarias establecidas en el **Plan Estratégico Nacional de Ciencia, Tecnología e Innovación "Bicentenario" (2006-2010)**, más específicamente en el área temática denominada **"Competitividad y Diversificación Sustentable de la Producción Agropecuaria"**, que promueve el desarrollo de nuevas variantes de cultivos con resistencias genéticas a factores adversos (bióticos y abióticos) y mayor eficiencia en la captación de nutrientes, agua, radiación, etc [MinCYT, 2006].

Durante el último lustro, surgieron diferentes métodos para efectuar la ingeniería inversa de GRNs empleando técnicas de inteligencia artificial [Schlitt y Brazma, 2007]. Estos métodos utilizan datos de expresión genética, los cuales constituyen una medida de la abundancia de ARNm de un subconjunto (o totalidad) de los genes presentes en el genoma de un organismo. Dichas mediciones son usualmente obtenidas mediante experimentos en laboratorios con *microarrays*, los cuales permiten recolectar datos de expresión genética a gran escala. A los fines prácticos, la información de expresión genética obtenida experimentalmente puede ser organizada en una estructura matricial, E , la cual representa una serie de tiempo, donde las filas corresponden a los genes observados y las columnas a las distintas muestras relevadas mediante los *microarrays*. De este modo, el elemento e_{ij} de E contiene un número real que indica el valor de expresión del gen i -ésimo en la muestra j -ésima. Sobre la base de esta información, los métodos de inteligencia computacional propuestos para reconstruir GRNs infieren potenciales asociaciones de regulación entre genes, mediante la búsqueda sistemática de correlaciones entre los datos presentes en la matriz E . Este proceso usualmente requiere efectuar primero una discretización de los datos, obteniéndose una matriz E' que representa la evolución del valor de expresión de cada gen en términos de un conjunto finito

de estados. La idea básica detrás de esta metodología es que los diferentes estados de un gen constituyen una medida de su grado de actividad dentro de cada muestra. De este modo, la identificación de un patrón de comportamiento similar (o contrapuesto) entre diferentes genes puede estar indicando la presencia de un potencial vínculo de regulación entre los mismos, entendiendo por patrón a una secuencia consecutiva de estados contenida dentro de una o más filas de E' . Las asociaciones identificadas mediante estos métodos pueden luego ser expresadas como potenciales reglas de regulación, las cuales permitirán la posterior reconstrucción del grafo correspondiente a la GRN estudiada.

Varios métodos provenientes de la inteligencia artificial fueron propuestos para inferir GRNs a partir de datos obtenidos mediante *microarrays*. Las redes booleanas fueron una de las primeras técnicas empleadas [Akutsu *et al.*, 1999], y algunas variaciones de este enfoque han sido publicadas más recientemente [Mehra *et al.*, 2004]. Estos algoritmos utilizan variables booleanas para expresar el estado de un gen, activo (1) o inactivo (0), mientras que las interacciones son modeladas mediante funciones booleanas que calculan el estado de un gen a partir de la activación de otros genes. La sencillez de este enfoque constituye su principal ventaja pero también su mayor limitación. Las redes booleanas permiten analizar en forma eficiente redes regulatorias de gran dimensión, dado que se realizan fuertes suposiciones que simplifican la estructura y dinámica de los sistemas regulatorios genéticos. Por ejemplo, se asume que el estado del gen es siempre activo o inactivo, sin permitir el modelado de niveles de expresión intermedios. Además se asume que las transiciones entre estados ocurren sincrónicamente. Por ende, cuando las transiciones no suceden simultáneamente, lo cual es usual en la mayoría de los casos, ciertos comportamientos quedan fuera del modelado computacional. Las redes bayesianas proveen un enfoque probabilístico para abordar el modelado de GRNs. En este caso la representación es un grafo dirigido acíclico, donde cada nodo representa una variable

(usualmente genes) y las aristas representan dependencias. Ejemplos de la aplicación de redes bayesianas al modelado de GRNs pueden encontrarse en [Friedman, 2004; Pe'er, 2005]. En general el uso de esta metodología para estudiar GRNs posee interesantes ventajas, tales como la capacidad de lidiar con aspectos estocásticos y con la presencia de ruido en los datos. Sin embargo, hay situaciones, como las relaciones autoregulatorias, en que las reglas booleanas y las funciones lineales no son suficientes para expresar la lógica de control de una GRNs [Schlitt y Brazma 2007]. En estos casos, se requieren funciones de control más complejas que las brindadas por las redes bayesianas.

Un aspecto que en general ha sido escasamente tratado por los métodos antes mencionados es el caso de la **regulación diferida en el tiempo** (*time-lagged regulation*). Esta forma de regulación sucede cuando la relación causa-efecto de una acción regulatoria requiere varias unidades de tiempo para que efectivamente ocurra. En términos de la matriz E' , esto implica que un cambio en el estado de un gen i -ésimo en la muestra k puede implicar una modificación en el estado del gen j -ésimo recién en la muestra $k+m$, siendo m un número entero positivo. Este fenómeno empezó a ser tratado recientemente por **algoritmos de reconstrucción de GRNs basados en inferencia de potenciales reglas de regulación**. Dentro de esta categoría se encuentran los algoritmos propuestos por Soinov *et al.* [2003] y Li *et al.* [2006]. Ambos tratan la reingeniería de GRNs como un problema de clasificación mediante el uso de árboles de decisión, y aplican el algoritmo C4.5 [Quinlan, 1992] para inferir estas estructuras arbóreas que conducen a la identificación de **potenciales reglas de regulación**. Lamentablemente, ambos métodos heredan las limitaciones del algoritmo C4.5 que pierde precisión en problemas de clasificación que requieren la evaluación de atributos con valores comprendidos dentro de subrangos de los reales [Ruggieri, 2002].

RESULTADOS OBTENIDOS

En este contexto propusimos un nuevo método, denominado GRNCOP [Ponzoni *et al.*, 2007], que emplea un proceso de discretización adaptiva y optimización combinatoria que supera las limitaciones de los métodos anteriores. No obstante, GRNCOP sólo infiere relaciones con ciertos patrones específicos de regulación diferida en el tiempo, por lo cual su aplicabilidad sufre de restricciones en la práctica.

Otra cuestión importante que ha comenzado a ser estudiada recientemente es el desarrollo de métodos computacionales que faciliten la validación y simulación de las GRNs reconstruidas mediante técnicas inteligentes [Mendes *et al.*, 2003; Batt *et al.*, 2005]. Al respecto, la generación de **redes regulatorias artificiales** es de fundamental importancia para los procesos de validación de los algoritmos de inferencia de GRNs [Mendes *et al.*, 2003; Carballido y Ponzoni, 2008]. En particular, el modelo evolutivo basado en el Genoma Artificial de Reil [1999] propuesto por Quayle y Bullock [2006] constituye una alternativa prometedora. También la verificación automática de propiedades está empezando a ser utilizada en el estudio de GRNs [Antoniotti *et al.*, 2003; Batt *et al.*, 2005; Siebert *et al.*, 2008]. El objetivo aquí consiste en poder simular el comportamiento dinámico de una GRN, partiendo de modelos computacionales de GRNs. En este contexto, las **técnicas de Model-Checking**, provenientes del área de ingeniería de software, resultan muy atractivas para su empleo en este campo, dado que las GRNs pueden ser modeladas como autómatas finitos [Bérard *et al.*, 1998]. Luego, las técnicas de Model-Checking pueden ser utilizadas para razonar sobre el funcionamiento de una GRN mediante el uso de sistemas deductivos y lógica temporal proposicional. Por ejemplo, una herramienta de Model-Checking puede detectar inconsistencias entre las relaciones inferidas para una GRN, ayudando a descartar asociaciones entre genes obtenidas por casualidad (*by chance*). Más aún, estas herramientas pueden ser utilizadas para extraer nuevo conocimiento a partir de la corroboración de nuevas hipótesis biológicas,

las cuales pueden ser expresadas, y evaluadas por el sistema deductivo, mediante consultas lógicas (*logical queries*). De este modo, resulta factible elaborar una metodología integral para la inferencia y simulación de GRNs que facilite el descubrimiento de conocimiento en este tópico, lo cual constituye precisamente el objetivo general, a largo plazo, de estas investigaciones.

Por último, existe un tercer aspecto relacionado con la capacidad de obtener modelos de GRNs complejas mediante la integración de diferentes métodos de inferencia, e incluso de distintas fuentes de información. Una de las principales falencias, reconocida por la mayoría de los especialistas en el área, es la dificultad de lograr modelos completos de redes regulatorias a partir de un único proceso de inferencia orientado a procesar exclusivamente datos de expresión genética provenientes de series de *microarrays* [Pridgeon y Corne, 2004; Schlitt y Brazma, 2005; Schlitt y Brazma, 2006].

RESULTADOS ESPERADOS

Establecidos los antecedentes, es posible explicitar las metas perseguidas en esta línea de I/D. Principalmente, se busca a futuro lograr el desarrollo completo de una metodología integral de inferencia, validación y simulación de redes regulatorias de genes con las siguientes características específicas:

1. El proceso de reconstrucción de redes regulatorias estará basado en la **inferencia de reglas** de regulación. Se elige esta metodología porque promueve un enfoque constructivista del modelo de una red, favoreciendo el descubrimiento de conocimiento que pueda ser explicado en términos de relaciones *causa y efecto*.
2. El proceso de inferencia debe permitir el descubrimiento de **reglas de regulación diferida en el tiempo**. Esto resulta fundamental para modelar el comportamiento de una red como sistema de eventos discretos que facilite su posterior validación y simulación mediante técnicas de verificación formal de sistemas.

3. El proceso de inferencia central, basado en el descubrimiento de reglas de regulación, debería permitir la **incorporación de conocimiento biológico adicional** proveniente de otras fuentes, más allá de los datos de expresión genética obtenidos mediante *microarrays*. Esto permitirá enriquecer los modelos inferidos para redes regulatorias complejas.
4. Además del proceso de inferencia central, basado en el descubrimiento de reglas de regulación, se deberían desarrollar modelos de **inferencia más específicos orientados** a la detección grupos de genes co-expresados y sub-topologías (genes *in-hub* y *out-hub*). De este modo, mediante la **integración del conocimiento** aportado por cada método de inferencia es posible reconstruir redes regulatorias más complejas. El postulante ya ha efectuado contribuciones en este tema, a través del desarrollo de un algoritmo evolutivo multiobjetivo para detección de biclusters (Gallo *et al.*, 2008).
5. También es necesario estudiar como **adaptar y aplicar técnicas de Model-Checking** para simular y evaluar nuevas hipótesis biológicas sobre el comportamiento de estas redes, estableciendo los alcances y limitaciones de estas técnicas, y el tipo de propiedades lógicas que pueden ser evaluadas sobre estos sistemas biológicos.

En conclusión, con el trabajo realizado en el contexto de esta línea de investigación y desarrollo se pretende integrar todas las herramientas y conocimientos desarrollados en estas investigaciones en una arquitectura de software que asista a científicos en el descubrimiento, análisis y simulación de redes regulatorias de genes, con el fin de obtener un producto tangible y práctico.

BIBLIOGRAFIA

1. Akutsu, T.; Miyano, S.; Buhara, S. "Identification of genetic networks from a small number of gene expression patterns under the Boolean network

- model”, Pacific Symp. Biocomput. 4, pp. 17-28, 1999.
2. Antoniotti, M.; Policriti, A.; Ugel, N.; Mishra, B. “Model Building and Model Checking for Biochemical Processes”, *Cell Biochemistry and Biophysics*, 38:271-286, 2003.
 3. Batt, G.; Ropers, D.; de Jong, H.; Geiselman, J.; Mateescu, R.; Page, M.; Schneider, D. “Validation of qualitative models of genetic regulatory networks by model checking: analysis of the nutritional stress response in *Escherichia coli*”, *Bioinformatics*, 21 S.1:19-28. 2005.
 4. Bérard, B.; Bidoit, M.; Finkel, A.; Laroussinie, F.; Petit, A.; Petrucci, L.; Schnoebelen, P.; McKenzie P. “Systems and Software Verification. Model-Checking Techniques and Tools”, Springer. 1998.
 5. Bishop, C.M. “Pattern Recognition and Machine Learning”, Springer, 1st edition, 2006.
 6. Carballido, J., Ponzoni, I. “On Artificial Gene Regulatory Networks”, *Electronic Journal of SADIO*. 8(1):25-34. 2008.
 7. Cios, K.; Kurgan, L.; Reformat, M. “Machine learning in the life sciences-How it is used on a wide variety of medical problems and data”, *IEEE Engineering in Medicine and Biology Magazine*, 26(2):14-16. 2007.
 8. De Jong, K.A. “Evolutionary Computation. A Unified Approach”, MIT Press, 2006.
 9. Friedman, N. “Inferring cellular networks using probabilistic graphical models”, *Science*, 303(5659):799-805, 2004.
 10. Gallo, C.; Carballido, J.; Ponzoni, I. “A Hybridized Multiobjective Evolutionary Approach for Microarray Biclustering”, *CLEI 2008 (XXXIV Conferencia Latinoamericana de Informática)*, 8-12, Septiembre, 2008. En prensa.
 11. Handl, J.; Kell, D.B.; Knowles, J. “Multiobjective optimization in bioinformatics and computational biology”, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(2):279-292. 2007.
 12. Huang, S. “Rational drug discovery: what can we learn from regulatory networks?”, *Drug Discovery Today*, 7(20):S163-S169, 2002.
 13. Li, X.; Rao, S.; Jiang, W.; Li, C.; Xiao, Y.; Guo, Z.; Zhang, Q.; Wang, L.; Du, L.; Li, J.; Li, L.; Zhang, T.; Wang, Q.K. “Discovery of time-delayed gene regulatory networks based on temporal gene expression profiling”, *BMC Bioinformatics*, 7:26, 2006.
 14. Mehra, S.; Hu, W-S.; Karypis, G. “A Boolean algorithm for reconstructing the structure of regulatory networks”, *Metabolic Engineering*, 6:326-339, 2004.
 15. Mendes, P.; Sha, W.; Ye, K. “Artificial gene networks for objective comparison of analysis algorithms”, *Bioinformatics*, 19(Suppl. 2):ii122-ii129, 2003.
 16. MinCYT, Plan Estratégico Nacional de Ciencia, Tecnología e Innovación “Bicentenario” (2006–2010). Anexo - Documentos consensuados con las Secretarías de Estado responsables de las políticas sectoriales. Prioridades en Investigación, Desarrollo e Innovación para el Programa PROTIS. Página 35, 2006.
 17. Pe'er, D. “Bayesian network analysis of signaling networks: a primer,” *STKE* 2005, 281:pl4, 2005.
 18. Ponzoni, I.; Azuaje, F.; Augusto, J.; Glass, D. “Inferring association rules between genes using a combinatorial optimization learning process and adaptive regulation thresholds”, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(4): 624-634, 2007.
 19. Pridgeon, C., Corne, D. “Genetic network reverse-engineering and network size; can we identify large GRNs?”, *Proc. 2004 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB '04: 7-8 October 2004; La Jolla, California, USA; pp. 32–36, 2004.*
 20. Quayle, A., Bullock, S. “Modelling the evolution of genetic regulatory networks”, *J. Theor. Biol.* 238:737–753. 2006.
 21. Quinlan, J.R. “C4.5: Programs for Machine Learning,” Morgan Kaufmann, 1992.
 22. Reil, T. “Dynamics of gene expression in an artificial genome: Implications for biological and artificial ontogeny”. In: Floreano, D., Mondada, F. & Nicoud, J. D. (ed.): *Proc. 5th European Conference on Artificial Life. Lecture Notes in Computer Science*, Vol. 1674. Springer-Verlag, pp. 457–466, 1999.
 23. Ruggieri, S. “Efficient C4.5,” *IEEE Trans. on Knowledge and Data Engineering* 14:438-444, 2002.
 24. Schlitt, T.; Brazma, A. “Modelling gene networks at different organisational levels,” *FEBS Letters*, vol. 579, pp. 1859-1866, 2005.
 25. Schlitt, T.; Brazma, A. “Modelling in molecular biology: describing transcription regulatory networks at different scales,” *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 361(1467), pp. 483-494, 2006.
 26. Schlitt, T.; Brazma, A. “Current approaches to gene regulatory network modelling,” *BMC Bioinformatics*, 8 (Suppl. 6):S9, 2007.
 27. Siebert, H.; Bockmayra, A. “Temporal constraints in the logical analysis of regulatory networks”, *Theoretical Computer Science*, 391(3): 258-275, 2008.
 28. Sohler, F. “Contextual Analysis of Gene Expression Data”. Master Thesis, Germany, 2006.
 29. Soinov, L.A.; Krestyaninova, M.A.; Brazma, A. “Towards reconstruction of gene networks from expression data by supervised learning,” *Genome Biology*, 4:R6, 2003.
 30. Yamaguchi-Shinozaki, K.; Shinozaki, K. “Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses”, *Annual Review of Plant Biology*, 57:781–803, 2006.

Análisis para el Desarrollo Multiparadigma

Silvia Amaro, Pablo Quiroga

{samaro, pquiroga}@uncoma.edu.ar

Claudio Vaucheret, Lidia López, Daniel Dolz

{vaucheret, lidiamlopez, daniel.dolz}@gmail.com

Andrea Granados, Ingrid Godoy, Viviana Sánchez, Maximiliano Klemen

{agranados80, ingriduncoma, sánchez.viviana, maximilianoklemen}@gmail.com

Paola Pérez

ppercat@hotmail.com

Dpto. de Ciencias de la Computación

Facultad de Economía y Administración

Universidad Nacional del Comahue, Argentina

1. Contexto

La línea de investigación orientada al desarrollo multiparadigma se enmarca dentro del proyecto de investigación "Técnicas Avanzadas y Análisis para el Desarrollo Multiparadigma"

2. Resumen

En el proceso de desarrollo de software, es especialmente importante lograr los resultados esperados en el rango de tiempo esperado y con la calidad esperada, aplicando las técnicas y prácticas más apropiadas al problema planteado. Para ello es necesario ser capaces de identificar la naturaleza del problema a resolver para elegir las herramientas y técnicas que mejor se adapten a la situación, teniendo en cuenta que diferentes tipos de problemas pueden ser atacados utilizando los paradigmas apropiados. Debido a que usualmente los problemas complejos tienen una naturaleza multidimensional, en la mayoría de los casos es apropiado utilizar una aproximación multiparadigma. Nuestro objetivo es el estudio de técnicas y análisis globales de programas den-

tro del contexto del desarrollo multiparadigma, tendiendo a establecer modelos, procesos y técnicas para mejorar el desarrollo de sistemas en un ambiente multiparadigma.

Palabras clave: Programación multiparadigma, Orientación a Componentes, Orientación a Aspectos, Orientación a Servicios, Análisis estático, Análisis dinámico.

3. Introducción

En el proceso de desarrollo de software las abstracciones juegan un rol central. Una abstracción se enfoca en la esencia del problema y excluye los detalles especiales. El desarrollo de software es fundamentalmente una actividad humana de forma que, a pesar de que en la actualidad es ampliamente soportado por herramientas automáticas, aún es influenciado por experiencias personales, tradiciones, convenciones y costumbres. Estos patrones de comportamiento tienen un impacto importante sobre el proceso de desarrollo de software determinando la forma en que los desarrolladores

seleccionan y forman las abstracciones.

En tecnologías de software un paradigma representa las directivas para crear abstracciones. El paradigma es el principio por el cual un problema puede ser comprendido y descompuesto en componentes manejables. Un paradigma de programación es el modo de conceptualizar lo que significa realizar una computación y cómo las tareas a llevarse a cabo sobre una computadora deben ser estructuradas y organizadas. El lenguaje en el cual un programador piensa un problema influirá en modo básico fundamental la forma en la cual un algoritmo es desarrollado.

Algunos paradigmas ya establecidos son:

- Programación Imperativa o Procedural: Modelo tradicional de computación, está estrechamente relacionado con el modelo de ejecución en el hardware subyacente.
- Programación Orientada a Objetos: Representación de los problemas intuitivamente. La computación se realiza a través de objetos comunicándose entre ellos y requiriendo que otros objetos realicen acciones.
- Programación Funcional: Tratamiento de las variables y sus valores tratados como entidades únicas. Una vez creados, los valores no son modificados, sino que son transformados en nuevos valores independientes de los originales.
- Programación Lógica: Basado en el cálculo de predicados. La programación es declarativa, el programador simplemente provee una serie de aserciones de hechos, un conjunto de reglas de inferencia y consultas.
- Programación orientada al Acceso: Uso de efectos laterales o "demonios" que son vinculados a las variables de manera que ciertas funciones sean realizadas en el momento en que las variables son manipuladas.

- Programación de restricciones: Especificación de una secuencia de restricciones que deben ser mantenidas durante la ejecución del programa.
- Programación literal: Combinación de documentación y código fuente en un único archivo con el fin de mantener la correspondencia entre la documentación y el código mismo.

Además, se puede identificar: (a) programación visual, (b) programación paralela, (c) programación orientada a aspectos, (d) programación orientada a componentes, las que se basan en alguno o varios de los paradigmas anteriores, (e) programación orientada a servicios.

Ningún paradigma es capaz de resolver todos los problemas de forma sencilla y eficiente, por lo tanto es útil poder elegir entre distintas técnicas de programación dependiendo del tipo de problema. En este escenario el desarrollo multiparadigma tiene como principal objetivo proveer al programador con un conjunto variado de herramientas de manera que cuando enfrente un problema no se vea forzado a utilizar una única técnica de programación, sino que tenga la libertad de seleccionar la técnica de solución que mejor se adapte a las características del problema a resolver.

En particular, un lenguaje que se está tomando como caso de estudio y plataforma de desarrollo dada sus características únicas es *Ciao Prolog* [9, 10]. *Ciao* es un lenguaje multiparadigma basado en expansiones sintácticas que implementan distintos paradigmas mediante la técnica de compilación de control.

La aparición de la programación orientada a objetos permitió comenzar a desarrollar software con una visión diferente. En este escenario surge la Programación Orientada a Aspectos (AOP), que propone una mejor modularización

[11]. AOP trabaja con paradigmas y lenguajes ya existentes para mejorar su expresividad y funcionalidad, y por ser un desarrollo que continúa el paradigma de Programación Orientada a Objetos soporta la descomposición orientada a aspectos, orientada a objetos y funcional [12, 13]. Es una metodología de programación que crece y evoluciona rápidamente.

Las técnicas de orientación a aspectos, en conjunto con técnicas de metadatos pueden ser utilizadas para inspeccionar un sistema y extraer los datos necesarios para diseñar heurísticas que sean utilizadas por los mecanismos de reflexión y orientación a aspectos para manejar la evolución [6].

La programación orientada a componentes se propone producir componentes de software para un mercado y una futura composición. Los compositores son terceras partes, posiblemente usuarios finales, que no están habilitados a realizar cambios. Esto requiere que componentes creadas de forma independiente puedan interoperar y además requiere de especificaciones que ubiquen al compositor en posición de decidir qué puede ser compuesto y bajo qué condiciones. En este sentido la coordinación y adaptación son puntos clave, y se enfocan en las interacciones entre entidades computacionales y los problemas que surgen cuando las entidades no concuerdan, respectivamente [4]. Los protocolos de interacción pueden ser complementados con el uso de aspectos para soportar la evolución de los componentes [1, 2]. En el mismo sentido existen varias propuestas para utilizar las tecnologías de la orientación a aspectos en el contexto de los servicios web [5].

4. Líneas de Investigación y Desarrollo

Para lograr los objetivos planteados se están abordando distintas temáticas en el contexto del desarrollo multiparadigma, como:

- Análisis global estático de programas no dependiente del paradigma. Uno de los tipos de análisis realizado en Ciao Prolog que puede ser transparente al paradigma utilizado, es la inferencia de tipos por medio de análisis estático "top-down" de programas basados en interpretación abstracta que tiene gran aplicación en la optimización y depuración de programas [3, 8]. La interpretación abstracta de programas es una tecnología que permite el análisis estático de programas para obtener propiedades de la ejecución de los mismos sin ejecutar realmente los programas.
- Análisis y adecuación de técnicas orientadas a aspectos. En general, al trabajar con lenguajes de AOP, la lógica de los aspectos se entrelaza en la aplicación destino, y en esta situación múltiples aspectos, desarrollados de forma independiente pueden producir comportamiento inesperado en el sistema. Este problema ha sido identificado como el problema de la interacción de características (Feature Interaction Problem). La interacción entre distintos aspectos puede darse cuando especifican comportamiento en el mismo punto, pero también cuando los aspectos no comparten el punto de encuentro. Además, los aspectos a menudo ejercen amplias influencias sobre los intereses, por ejemplo modificando su semántica, estructura o comportamiento. Estas dependencias entre elementos aspectuales y no aspectuales pueden derivar en interacciones tanto deseables como no deseables [7].
- Análisis de propiedades estáticas y dinámicas en orientación a componentes. Es interesante determinar si la evolución basada en aspectos preserva las nociones fundamentales de co-rectitud, en particular la compatibilidad y sustitutabilidad de componentes. ¿Se preserva la naturaleza de caja negra de los componentes?, ¿cómo afecta a la naturaleza de caja negra de un componente y a su composición el uso de un paradigma de programación declarativo?.

- Extensiones de lenguajes para incorporar paradigmas no nativos. Se está realizando un análisis de los lenguajes que tienen características multiparadigma, para definir un entorno comparativo de los mismos. Interesa estudiar la posibilidad de extender algunos lenguajes, incorporando características multiparadigma.

5. Resultados esperados

El desarrollo multiparadigma permite la construcción de sistemas mediante la selección del paradigma apropiado en cada situación. El producto de la presente propuesta puede ser de utilidad para el desarrollo de aplicaciones en dominios particulares y que deban cumplir con propiedades particulares, dando la posibilidad de evaluar y seleccionar las mejores técnicas y paradigmas para la construcción de sistemas más eficientes, seguros y de mejor calidad.

6. Formación de Recursos Humanos

El mayor impacto del presente proyecto se centra en la formación de recursos humanos, consolidación de grupos de investigación e interacción entre grupos interdisciplinarios. Permitirá a parte de los autores dar sus primeros pasos en investigación. Se prevee continuar con trabajos conjuntos con docentes de la Universidad de Málaga y la Universidad de Huelva, España. Se aspira a comenzar un trabajo conjunto con docentes de la Universidad Libre de Bruselas (Vrije Universiteit Brussel), y con docentes de la Universidad de Madrid. En particular las líneas de investigación propuestas posibilitarán a sus integrantes la conclusión de estudios de grado, conclusión de estudios de posgrado en curso y la iniciación de nuevos estudios de posgrado.

Referencias

- [1] Guido Soldner and Rudiger Kapetza. *AOCI: An Aspect-Oriented Component Infrastructure*. Workshop on Component Oriented Programming WCOP, ECOOP 2007, Berlin, Alemania.
- [2] Angel Nunez, Jacques Nayé. *A Seamless Extension of Components with Aspects using Protocols*. Workshop on Component Oriented Programming WCOP, ECOOP 2007, Berlin, Alemania.
- [3] Debra J. Richardson. *Static Analysis: Analysis of Models, Data flow analices, Finite-state verification*. <http://www.ics.uci.edu/08-StaticAnalysis.pdf>, University Of California. 2000
- [4] S. Amaro, E. Pimentel and A.M. Roldán, *Reo Based Interaction Model*, International Workshop on Formal Aspects of Component Software, FACS'05, UNU-IIST, Macao, China, Octubre 2005.
- [5] Florian Irmert, Marcus Meyerhofer, Markus Weiten. *Towards Runtime Adaptation in SOA Environment*. Workshop on Reflection, AOP and Meta-Data for Software Evolution, RAM-SE, ECOOP 2007, Berlin, Alemania.
- [6] Dong Ha Nguyen and Mario Südholt. *Property-Preserving Evolution of Components Using VPA-Based Aspects*. Workshop on Reflection, AOP and Meta-Data for Software Evolution, RAM-SE, ECOOP 2007, Berlin, Alemania.
- [7] Pablo Quiroga. *Control-Flow Interaction in Aspect-Oriented Programming*. Tesis Master of Science in Computer Science, Nantes, Francia. Septiembre 2007.
- [8] R. Barbuti, R. Giacobazzi, and G. Levi. *A general framework for semantics-based bottom-up abstract interpretation of logic programs*. ACM Transactions on Programming Languages and Systems, 15(1):133-181, 1993.
- [9] P. Van Hentenryck, A. Cortesi, and B. Le Charlier. *Type analysis of prolog using type graphs*. Journal of Logic Programming, 22(3):179-209, 1995.

- [10] C. Vaucheret and F. Bueno. *More precise yet efficient type inference for logic programs*. In International Static Analysis Symposium, number 2477 in LNCS, pages 102-116. Springer-Verlag, September 2002.
- [11] Robert E. Filman, Tzilla Elrad, Siobhán Clarke, and Mehmet Akşit, editors. *Aspect-Oriented Software Development*. Addison-Wesley, Boston, 2005.
- [12] K. Mehner and A. Wagner. *An assessment of aspect language design*. Young Researchers Workshop, First International Symposium on Generative and Component-Based Software Engineering, September 1999.
- [13] Éric Tanter and Jacques Noyé. *A Versatile Kernel for Multi-Language AOP*. In Proceedings of the ACM SIGPLAN/SIGSOFT Conference on Generative Programming and Component Engineering (GPCE 2005), LNCS, Springer-Verlag, Tallin, Estonia, September, 2005.

APRENDIZAJE Y COMPARTICIÓN DE CONOCIMIENTOS ENTRE SISTEMAS INTELIGENTES AUTÓNOMOS

Jorge Salvador Ierache y Ramón García Martínez

Programa de Doctorado de la Facultad de Informática de la Universidad Nacional de la Plata
Instituto de Sistemas Inteligentes y Enseñanza Experimental de la Robótica, Facultad de Informática Ciencias
de la Comunicación y Técnicas Especiales. Universidad de Morón.
Laboratorio de Sistemas Inteligentes Facultad de Ingeniería Universidad de Buenos Aires

jierache@unimoron.edu.ar, rgarciamar@fi.uba.ar

CONTEXTO

La línea de investigación planteada se desarrolla en el marco del área de sistemas autónomos inteligentes en el contexto de robótica autónoma, ingeniería de software, e ingeniería del conocimiento

RESUMEN

La línea de investigación se orienta a definir un marco teórico e experimental que modelice la compartición de conocimientos entre sistemas autónomos con la incorporación de un ciclo de vida de aprendizaje compuesto de distintos layers que acompañan el aprendizaje y compartición de conocimiento entre Sistemas Inteligentes Autonomos (SIAs) de robots. Se enfoca en plantear un modelo que facilite a SIAs interoperar en ambientes distribuidos, compartiendo desde las teorías del creador, las teorías que se generan como producto de su entrenamiento y las que surgen de la interacción entre con el ambiente de su mundo de actuación y otros SIAs.

Se consideran de la Ingeniería del Software los conceptos de ciclo de vida, arquitecturas y modelos. Desde la perspectiva de la Ingeniería del Conocimiento, la modelización de conocimiento a través de una ontología preliminar que represente al SIA para su actuación aprendizaje y compartición de conocimientos.

Palabras clave: *Sistemas autónomos, robótica, compartición de conocimiento, interoperabilidad, aprendizaje.*

1. INTRODUCCIÓN

Un sistema inteligente autónomo puede definirse [Fritz *et al.*, 1989; García Martínez y Borrajo, 1996] como aquél capaz de descubrir y registrar si una acción efectuada sobre una situación dada fue beneficiosa para lograr su objetivo. Necesita construir planes, monitorizar la ejecución de esos planes para detectar expectativas violadas y diagnosticar y rectificar errores que los datos inconsistentes revelen [García Martínez y Borrajo, 2000].

El aprendizaje es necesario porque en un nuevo entorno, el sistema no puede saber “a priori” las consecuencias de sus acciones ni las Recientemente

en [García Martínez *et al.*, 2006], se abordara el problema de compartir el conocimiento entre distintos sistemas inteligentes autónomos (SIAs). Los sistemas inteligentes autónomos, son autónomos en el sentido de que tienen sus propias funciones de utilidad, y ninguna noción de utilidad global [ni siquiera una implícita] influye en su diseño.

Para compartir el entendimiento común de la estructura de información requiere desarrollar ontologías [Gruber, 1993]. Una ontología contiene definiciones, conceptos básicos y relaciones entre estos que pueden ser interpretados por un sistema. El desarrollo de ontologías permite compartir el entendimiento común de las estructuras de información entre personas o SIAs, la reutilización de conocimientos del dominio, explicitar suposiciones del dominio, separar el conocimiento del dominio del conocimiento operacional, analizar el conocimiento de un dominio.

La interoperabilidad semántica se a través de M2M (Machine to Machine) resulta de de interes para los SIAs, [Pulvermacher *et al.*, 2004] presenta la integración semántica de sistemas de Comando y Control (C2), Aplicaciones en ambientes de misión crítica, aplicados al control de trafico aéreo bajo el concepto de Free Flight, [Battista *et al.*, 2007], representan un claro ejemplo de de sistemas autónomos inteligentes en un contexto M2M.

Para la actuación de los SIAs se requiere de un marco que integre el aprendizaje y la compartición de conocimiento sobre la base de un modelo común para los SIAs, se presenta un problema crucial cuando se está tratando con sistemas que deben "autonómicamente" adaptarse a un ambiente dinámico y desconocido.

Uno de los principales objetivos de cada SIA, es cómo los operadores aprenden automáticamente (modelos de acción) que predice los efectos de acciones en el ambiente por observación de las consecuencias de estas acciones. Para poder aprender estas descripciones, éste es capaz de lograr metas auto-definidas, ejecutar los planes, encontrar los comportamientos correctos o incorrectos, y aprender de la interacción con el ambiente y otros SIAs.

Cada SIA recibe la percepción desde el ambiente, llamado situaciones y aprende desde su interacción

con el mundo externo (el ambiente y otros SIAs). Al principio, el SIA percibe la situación inicial, y selecciona una acción al azar para ejecutar en el ambiente. Entonces, entra en un *loop* para la ejecución de una acción, percibiendo las situaciones resultantes y la utilidad de la situación, aprendiendo de la observación los efectos de aplicar las acciones en el ambiente.

Cada uno de estos SIAs recibe un *input*: percepciones desde el ambiente (situaciones y utilidades); un conjunto de acciones que pueden ejecutarse, operadores aprendidos propios y de otros SIAs.

La salida de cada SIA es una secuencia de acciones sobre el tiempo (para el ambiente), regularmente ese conjunto de operadores que él aprende y comparte con otros AIS. El operador O es una tupla $\langle C, A, F, P, K, U \rangle$ [García Martínez, R. y Borrajo, D., 2000] donde: C es la situación inicial (condiciones), A acción a ser ejecutada, F situación final (post-condiciones), P cantidad de veces que el operador O_i fue aplicado satisfactoriamente (la citación final F fue obtenida), K cantidad de veces que la acción A fue aplicada a la condición C, U nivel de utilidad del operador, obtenido de aplicar la acción A, a la situación C, se determina sobre la relación P/K.

LOPE (Learning by Observation in Planning Enviroments) es una arquitectura de SIAs que integra planificación, aprendizaje y ejecución en un reiteración cerrada, mostrando cómo funciona el comportamiento de inteligencia autónoma [García Martínez *et al.*, 2006].

Trabajos anteriores basados en aprendizaje de máquinas eran aplicados a resolución de problemas que principalmente estaban enfocados en aprendizaje de conocimientos cuya meta era mejorar la eficiencia de la tarea de resolución de problemas [Borrajo y Veloso, 1997; García Martínez y Borrajo, 2000].

Nuestro trabajo se concentra en el desarrollo de un modelo para el aprendizaje y compartición de teorías entre múltiples SIAs sobre la base de la arquitectura LOPE-LLC (LOPE-Learning Life Cycle) [Ierache *et al.*, 2008b], en esta propuesta se considera que el SIA aprenda autónomamente sobre la base del ciclo de vida para el aprendizaje y compartición de conocimientos entre SIAs que considera estados de evolución del SIA como nacido, novato, entrenado, maduro los que se desarrollan a través de tres Layers:

[a] La layer de BI (Built-In Operators) que se desarrolla desde el nacimiento del SIA hasta que éste alcanza su estado de novato, el aprendizaje del SIA que se desarrolla en esta layer es a partir de los operadores implantados por el programador o creador. [b] La layer TB (Trained Base Operators) que se desarrolla desde el SIA novato hasta que el AIS alcanza el estado de entrenado, esta es la capa de aprendizaje en los que los operadores evolucionan en el contexto de entrenamiento del SIA

sobre la base inicial de los operadores del SIA novato.

Y [c] la layer WI (World Interaction Operators) que se desarrolla desde que el SIA es entrenado hasta que alcanza el estado de maduro, es la capa de aprendizaje en la que los operadores se aprenden por la interacción con el mundo y el intercambio con otros SIAs. La evolución de los estados (born, newbie, trained, mature) del SIA se desarrollan a través de las Layers (BIO, BTO, WIO), los SIAs pueden compartir sus operadores a través de sus layers (BIO, BTO, WIO), de acuerdo a su estado (born, newbie, trained, mature). Se muestra el ciclo de vida en Figura 1.

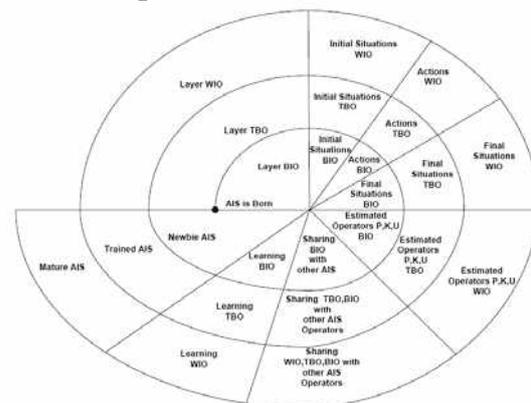


Figura 1. Modelo de Ciclo de vida del aprendizaje del SIA

En el contexto propuesto se imponen los siguientes axiomas de compartición entre SIAs, un SIA Mature puede compartir operadores con otros SIA entrenados, novatos y nacidos, un SIA entrenado comparte sus operadores con SIA entrenados y novatos, un SIA novato comparte sus operadores con SIAs novatos y nacidos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Nuestro enfoque en este trabajo apunta a aplicar de la ingeniería del software los conceptos de ciclo de vida, arquitecturas y modelos. Desde la perspectiva de la ingeniería del conocimiento, la modelización de conocimiento a través de una ontología preliminar que represente al SIA en su actuación, aprendizaje y compartición de conocimientos. Se desarrollan en la actualidad tres ejes contribuyentes con la investigación.

El primer eje se enfoca en el planteo de un ciclo de vida para el aprendizaje y compartición de conocimientos entre sistemas autónomos, que brinde las bases para el desarrollo de un framework para la actuación y compartición de conocimientos de sistemas autónomos de robots distribuidos.

El segundo eje apunta a la investigación de plataformas de robots y la experimentación sobre la base de prototipos, desarrollados sobre la base actuación de juguetes autónomos.

El tercer eje desarrolla las bases de un modelo de compartición de conocimientos entre sistemas autónomos de robots, orientado al desarrollo de una ontología.

3. RESULTADOS OBTENIDOS/ESPERADOS

En relación al primer eje se presentaron las bases de la arquitectura para el aprendizaje y compartición de conocimientos entre sistemas autónomos distribuidos [Ierache, J. *et al.*, 2008a], y se desarrolló la propuesta del ciclo de vida para el aprendizaje de sistemas autónomos [Ierache, J., García Martínez, R., De Giusti, A. 2008].

En relación al segundo eje en materia de experimentación y desarrollo de prototipos de plataformas autónomas de robots se presentaron resultados sobre la paliación de plataformas orientados al espacio de oportunidades en la educación, en este orden se abordaron tanto robots como juguetes autónomos [Ierache, J., Bruno, M., Dittler, M., Mazza N., 2008]. En materia de prototipo de presentaron resultados sobre la integración de un bípedo (Robosapiens V1), de programación abierta, [Ierache, J., Dittler, M., 2008].

Finalmente se presentaron resultados del tercer eje con la propuesta de una ontología para el aprendizaje y compartición de conocimientos entre sistemas autónomos [Ierache, J., Bruno, M., García Martínez, R., 2008].

Se llevarán a cabo experimentaciones sobre el framework en desarrollo basado en el ciclo de vida de aprendizaje y compartición propuesto que posibilita la interacción de varias arquitecturas Lope, considerando la compartición de operadores entre sistemas autónomos inteligentes y la valoración de estos como eje principal del trabajo en el contexto de actuación de SIAs simulados y reales.

Para estudiar la convergencia de aprendizaje y probar la generalidad del conocimiento compartido entre SIAs, se llevarán a cabo experimentos orientados a verificar que los SIAs tienen un mejor comportamiento utilizando conocimiento compartido que cuando no lo usan. Se explorará la velocidad de convergencia de este tipo de aprendizaje

La experimentación real se realizará sobre la base de las plataformas de robots estudiadas y el prototipo de bípedo de programación abierta. Los escenarios distribuidos se desarrollarán inicialmente sobre SIAs Simulados, interoperando en todos los casos sobre la base de la ontología propuesta.

Se trabajó a la fecha con plataformas NXT, RCX, robots de la línea wowwee Robosapiens V2 y Robosapiens V1 modificado con un microcontrolador BX-24P para generar una Bípedo de programación abierta, figura 2.

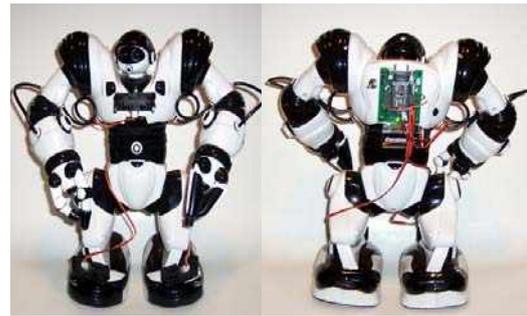


Figura 2. Robosapien V1 y BX-24

Se incorporará durante el segundo trimestre del año herramientas de simulación Webots 5 que facilitarán la simulación de robots khepera, lego Mindstorms, Aibo entre otros.

En materia de robots se incorporarán dos robots e-punk con ground sensors de cyber-robotics, un Bioloid comprehensive Kit de Crustcrawler y dos Lego Mindstorms NXT equipados con acelerómetros, sensores de color, compás y giróscopo. Este material enriquecerá la experimentación a través de la diversidad de robots que se aplicarán.

Para llevar adelante el proyecto se consolidará un framework basado en la propuesta para el aprendizaje y compartición de conocimiento entre sistemas inteligentes autónomos. Los AIS distribuidos compartirán conocimiento a través de una conexión a Internet. En futuras líneas se plantea la validación empírica en colaboración con el Instituto de Investigaciones en Informática LIDI de la UNLP.

4. FORMACIÓN DE RECURSOS HUMANOS

En el marco de este proyecto se está desarrollando una tesis doctoral y tres tesis de ingeniería informática (una de ellas defendida y dos en curso).

5. BIBLIOGRAFÍA

- Battista, V., Ierache, J., Britos, P., Rodríguez, D., García-Martínez, R., 2007. *Modelo Multiagente en Sistemas de Misión Crítica Aplicado al Control de Tráfico Aéreo Bajo el Concepto de Free Flight*. JIISIC-2007: 419-426.
- Fritz, W., García Martínez, R., Rama, A., Blanqué, J., Adobatti, R, y Sarno, M. 1989. *The Autonomous Intelligent System*. Robotics and Autonomous Systems, 5(2): 109-125.
- García Martínez, R. & Borrajo Millán, D. 1996. *Unsupervised Machine Learning Embedded in Autonomous Intelligent Systems*. Proceedings of the XIV International Conference on Applied Informatics. Páginas 71-73. Innsbruck. Austria.
- García Martínez, R. & Borrajo Millán, D. 1998. *Learning in Unknown Environments by Knowledge Sharing*. Proceedings of the Seventh European Workshop on Learning Robots. Páginas 22-32. Editado University of Edinburg Press.

- García Martínez, R. y Borrajo Millán, D. 1997. *Planning, Learning and Executing in Autonomous Systems*. Lecture Notes in Artificial Intelligence. 1348:208-210.
- García Martínez, R. y Borrajo Millán, D. 2000. *An Integrated Approach of Learning, Planning and Executing*. Journal of Intelligent and Robotic Systems 29(1):47-78.
- García-Martínez, R., Borrajo, D., Britos, P. y Maceri, P. 2006. *Learning by Knowledge Sharing in Autonomous Intelligent Systems*. Lecture Notes in Artificial Intelligence, 4140: 128-137.
- Gruber, T.1993. *A Translation Approach to Portable Ontology Specification*. Knowledge Acquisition 5:199-220.
- Ierache, J., Naiouf, M., Garcia Martinez, R., De Guiusti, A. 2008.a *Un modelo de Arquitectura para el aprendizaje y compartición de conocimiento entre sistemas inteligentes autónomos distribuidos*, IISIC-2008: 179-188.
- Ierache, J., Bruno M., Dittler, M.,Mazza N., *Robots y juguetes autónomos, una oportunidad en el contexto de las nuevas tecnologías de la educación*. IISIC 2008: 371-380.
- Ierache, J., M Dittler 2008. *Sistemas Autónomos de Programación Abierta a Partir de Juguetes Bipedos en el contexto de Fútbol de Robots*, Anales del V Workshop de Inteligencia Artificial aplicada a la robótica móvil, Universidad Nacional del Comahue, pp 105-111.
- Ierache, J., García Martínez., R., De Giusti, A.,2008b. *Learning Life Cycle in Autonomous Intelligent Systems*. En IFIP Volumen 276 , Artificial Intelligence and Practice II, Max Bramer,(Boston:Springer), pp 451-455.
- Ierache, J., Bruno, M García-Martínez, R. 2008. *Ontología para el aprendizaje y compartición de conocimientos entre sistemas autónomos*. XIV Congreso Argentino de Ciencias de la Computación, Universidad Nacional de Chilecito
- Pulvermacher M., Stoutenburg S., Semy, S. 2004. *Netcentric Semantic Linking: An Approach for Enterprise Semantic Interoperability*, MITRE Technical Report.

Gestión del conocimiento aplicada al peritaje informático

Leopoldo Sebastián M. Gómez ^(1,2), Hernán Horacio Herrera ⁽¹⁾

⁽¹⁾ Laboratorio Pericial Informático (LPI)
Poder Judicial del Neuquén

⁽²⁾ Programa de Magíster en Informática (dirección de tesis)
Universidad Nacional de La Matanza

sebastian.gomez@jusneuquen.gov.ar

CONTEXTO

Esta línea de investigación apoya la actualización profesional y capacitación permanente de recursos humanos especializados en informática forense dentro del Poder Judicial del Neuquén.

Se contribuye a nivel académico con la Escuela de Posgrado de la UNLaM, mediante la dirección externa de tesis de la Maestría en Informática.

RESUMEN

Desde el Laboratorio Pericial Informático del Poder Judicial del Neuquén se impulsan actividades de I+D orientadas a la gestión del conocimiento aplicada al peritaje informático. Se pretende generar un cuerpo de conocimientos que sea útil para la formación de profesionales auxiliares de la Justicia noveles, reduciendo los tiempos de aprendizaje de metodologías, técnicas y herramientas de informática forense. Se ha instrumentado la implementación de un repositorio de información digital que permitirá compartir el conocimiento especializado con otras instituciones judiciales y de las fuerzas de seguridad. Se espera integrar tecnologías de enseñanza virtual para contribuir con una futura formación a nivel de posgrado de especialistas en peritaje informático, haciendo énfasis en la investigación de delitos informáticos.

Palabras clave: *digital forensics, knowledge management*

1. INTRODUCCION

Durante la investigación de actividades delictivas, habitualmente se plantea un conjunto de objetivos o medidas tendientes a esclarecer un presunto ilícito. Esta modalidad de trabajo permite definir tres capas de abstracción a partir del caso judicial. El nivel superior, que mantiene la visión del caso en forma integral, es el marco de trabajo en el que se mueven los profesionales del derecho (jueces, fiscales, defensores, abogados, etc.). En el segundo nivel se establece la interacción entre los profesionales de diversas disciplinas. En este ámbito se plantean los objetivos - requerimientos periciales o informes técnicos- sobre los cuales actúan profesionales informáticos, contadores, médicos forenses, etc. Finalmente, el último nivel es el que pertenece en forma exclusiva a los expertos en la disciplina a la que pertenezca la actividad forense solicitada como objetivo para la investigación.

El avance constante de la tecnología obliga a los especialistas informáticos a mantener una actualización profesional permanente en desmedro de otros tiempos que son necesarios para la colaboración entre colegas. La posibilidad de capturar el conocimiento forense juega un rol crucial en el entrenamiento y la diseminación de mejores prácticas [1]. Es por ello que la gestión del conocimiento se presenta como un método adecuado para superar las limitaciones de tiempo y recursos humanos, contribuyendo a reducir la creación individual de información

redundante y permitiendo compartir en forma eficiente una mayor cantidad de recursos.

Aunque varias organizaciones han presentado definiciones alternativas del concepto de informática forense, se pueden encontrar algunas relaciones entre las metodologías [2]. Se necesita contar con un esquema de trabajo metódico y disciplinado para el desarrollo de labores periciales. La síntesis y ordenamiento sistemático de los componentes involucrados en tareas periciales permite reducir los tiempos del trabajo mediante la ejecución paralela de actividades forenses, siempre que se cuente con recursos humanos especializados y herramientas adecuadas [3].

En informática forense no existe una ventaja inmediata en el suministro de información a terceros. Los temores razonablemente fundados de que la información y experiencia documentada pueda ser utilizada con fines ilícitos para mejorar las actividades delictivas motivan a que los especialistas no deseen compartir el conocimiento aplicado al desarrollo de sus labores cotidianas [4].

2. LINEAS DE INVESTIGACION y DESARROLLO

La estructuración de contenidos periciales en un repositorio digital con los niveles de acceso apropiados permitiría al investigador menos experimentado contar con una valiosa fuente de información, accediendo a recursos compartidos ofrecidos por otros profesionales expertos. Contar con recursos humanos capacitados para el tratamiento de evidencia digital es uno de los principales puntos críticos para conducir la investigación a resultados exitosos [5]. Estas lecciones literalmente ahorran al perito o analista forense docenas de horas de investigación y métodos de prueba y error [6].

Por otra parte, las ventajas ofrecidas por un gestor de contenidos periciales minimizaría la necesidad de contar con personal especializado para la edición y publicación de información. Asimismo, es viable incorporar un mecanismo que permita el intercambio seguro de

información sensible o relevante entre analistas forenses y el personal policial. No menos importante es destacar la necesidad de utilizar tecnologías de código abierto, considerando los escasos recursos financieros con que cuentan los organismos judiciales y policiales.

La posibilidad de recopilar el conocimiento de investigación en informática forense, como así también las prácticas de laboratorio y la casuística juegan un rol crucial en el entrenamiento de nuevos peritos y la difusión de mejores prácticas en la comunidad forense.

Para motivar a los especialistas a compartir sus conocimientos se requiere implementar una solución tecnológica que integre en forma simple y clara las áreas clave del trabajo operativo pericial y facilite el acceso a los recursos en formato digital.

Se cuenta con un corpus importante de reseñas científicas de informática forense que han sido recopiladas en forma manual y son utilizadas como material de consulta del Laboratorio Pericial Informático del Poder Judicial del Neuquén. Dichos documentos digitales conformarán los contenidos mínimos del repositorio de información digital para la implementación de un sistema de gestión de contenidos periciales.

A la fecha, sólo se han detectado implementaciones de relevancia en esta línea de trabajo en instituciones extranjeras de gran porte como el Defense Cyber Crime Center (DC3) [7], quienes en forma conjunta con Oklahoma State University's Center for Telecommunications and Network Security (CTANS), auspician el desarrollo y operación del National Repository for Digital Forensics Intelligence (NRDFI) [8]. El material digital organizado en este sistema de gestión de conocimiento es accesible solamente a las agencias de seguridad y fuerzas de la ley de EEUU y países asociados al Five Eyes Agreement (Australia, Canadá, Inglaterra y Nueva Zelanda).

3. RESULTADOS OBTENIDOS/ESPERADOS

La esencia de generar un repositorio de información digital consiste en capturar y compartir las mejores prácticas de los peritos con aquellos que tengan que descubrir o desarrollar las mismas técnicas o similares. Un repositorio digital debe soportar virtualmente cualquier tipo de archivo de texto o binario, enlaces a páginas web, y permitir realizar búsquedas sobre todo el contenido que en él se almacene.

Se pretende obtener un cuerpo mínimo de conocimientos mediante la recolección de trabajos sobre informática forense y evidencia digital que sean de mayor aceptación en la comunidad forense, principalmente guías de mejores prácticas, guías de procedimientos y manuales de informática forense. Se han aplicado entrevistas estructuradas para comprender las necesidades de los peritos durante sus actividades operativas, con el objeto de refinar el repositorio hacia una taxonomía de tópicos forenses de uso frecuente.

Por otra parte, se ha seleccionado la tecnología apropiada para la gestión de contenidos periciales y se implementará un sistema informático capaz de integrar todas las áreas clave pertenecientes al dominio de trabajo de los especialistas en informática forense. Se definirán niveles de acceso con el objeto de garantizar el acceso restringido a información clasificada. Los usuarios del gestor de contenidos periciales tendrán acceso a la información en función a los permisos que le otorgue un administrador local.

Se espera contar con un repositorio organizado de contenidos educativos especializados que junto a la tecnología de e-learning adecuada permita ser compartido con otros profesionales del país y Latinoamérica, principalmente del ámbito judicial y de las fuerzas de seguridad.

4. FORMACION DE RECURSOS HUMANOS

En esta línea de I+D se está apoyando a un alumno avanzado para la elaboración del proyecto de tesina de grado de Licenciatura en

Ciencias de la Computación en la UNCo, y de un alumno de posgrado en la Maestría en Informática de la UNLaM. En el Poder Judicial del Neuquén se aplicarán los resultados obtenidos a la capacitación interna de profesionales que se incorporen al Laboratorio Pericial Informático para formarse como especialistas en peritaje informático.

5. BIBLIOGRAFIA

- [1]. Bruschi, D., Monga, M. and Martignoni, L., (2004), "How to Reuse Knowledge about Forensic Investigations", Proceedings of the Digital Forensic Research Workshop. Accedido el: 26/03/2009 de http://dfirws.org/2004/day3/D3-Martignoni_Knowledge_reuse.pdf
- [2]. Sansurooah, K., (2006), "Taxonomy of computer forensics methodologies and procedures for digital evidence seizure", Proceedings of the 4th Australian Digital Forensics Conference, ISBN 0-7298-0624-3.
- [3]. Gómez, L., (2006), "La investigación de actividades delictivas con alta tecnología", JAIIO, Simposio de Informática y Derecho.
- [4]. Biros, D., Weiser, M. and Whitfield, J., (2007), "Managing Digital Forensic Knowledge: An Applied Approach" Proceedings of the 5th Australian Digital Forensics Conference.
- [5]. Gómez, L., (2004), "El tratamiento de la Evidencia Digital", JAIIO, Simposio de Informática y Derecho.
- [6]. Harrison, W., Aucsmith, D., Heuston, G., Mocas, S., Morrissey, M., & Russelle, S., (2002), "A lessons learned repository for computer forensics", International Journal of Digital Evidence, 1(3). Accedido el: 26/03/2009 de <http://www.utica.edu/academic/institutes/ecii/publications/articles/A049D6C7-93E9-51F2-A468BF90038985DB.pdf>

Referencias citadas

- [7] Defense Cyber Crime Center
<http://www.dc3.mil>
- [8] National Repository for Digital Forensics Intelligence
<http://www.nrdfi.net>

GESTION DEL CONOCIMIENTO APLICADO A LA EDUCACION PROFESIONAL CONTINUA EN GESTION DE PROYECTOS INFORMATICOS

Ing. Inés Casanovas¹, Lic. R.Erife², Lic. C. Tomassino³

Dto. De Sistemas - Universidad Tecnológica Nacional FRBA

[1inescasanovas@gmail.com](mailto:inescasanovas@gmail.com), [2eri_10@yahoo.com.ar](mailto:eri_10@yahoo.com.ar), [3tomassino@fibertel.com.ar](mailto:tomassino@fibertel.com.ar)

CONTEXTO

Este trabajo se encuadra en la línea de investigación de Gestión de Proyectos Informáticos abierta en 2008 en la Cátedra de la materia Proyecto, carrera Ingeniería en Sistemas de Información, UTN-FRBA, en cumplimiento del Plan de Carrera Docente de los integrantes de la misma. Se prevén tres etapas de trabajo sucesivas con incorporación de investigadores noveles (ayudantes y alumnos) en el transcurso de las mismas.

RESUMEN

En esta etapa realizamos una revisión del concepto de competencias y Gestión de Proyectos Informáticos en la literatura y artículos académicos publicados, y en los estándares de organizaciones profesionales. En la siguiente etapa se explorarán las competencias establecidas en el ámbito profesional de proyectos informáticos para obtener el conjunto a incluir en el modelo conceptual a desarrollar aplicando Tecnologías de la Información colaborativas. En una tercera etapa se buscará determinar el modelo de Gestión del Conocimiento más adecuado para representar estas incumbencias de los directores de proyectos. Como resultado de esta primera etapa hemos establecido un marco conceptual actualizado en forma de compilación de las definiciones de competencia para la Gestión de Proyectos (PM) según las asociaciones profesionales de mayor peso internacional, comentando el nuevo rol de los Directores de Proyecto y la aparición de las Certificaciones profesionales y las “core competences” en el contexto. Según las últimas investigaciones, las empresas debieran capitalizar las habilidades y experiencias de la propia gestión convirtiéndolas en core competences a nivel organizacional mediante TIC’s, recurriendo a modelos de Gestión del Conocimiento. Las Tecnologías de la Información aplicadas al proceso de enseñanza-aprendizaje colaborativo mediante comunidades de aprendizaje y de práctica se presentan como alternativas eficaces.

Palabras clave: Educación profesional continua, Gestión de Proyectos, Gestión del Conocimiento, Tecnologías de la Información, competencias, comunidades de aprendizaje y de práctica.

1. INTRODUCCION

La gestión de proyectos informáticos se puede definir tradicionalmente como la aplicación de conocimientos, aptitudes, herramientas y técnicas a las actividades necesarias para llevar a cabo exitosamente un proyecto de desarrollo de sistemas de información, satisfaciendo las necesidades y expectativas de una organización. Para satisfacer estas necesidades hay que equilibrar las demandas entre: alcance, plazos, coste y calidad.

De acuerdo a la opinión de consultoras de primera línea, el contexto de negocios de hoy requiere que los Directores de Proyectos Informáticos, dominen conocimientos tan disímiles como lo básico en materia de contrataciones, finanzas empresariales, control integrado de presupuestos y cronogramas, medición del rendimiento en el trabajo, control de calidad y análisis de riesgo o negociar, manejar el cambio, ser astutos políticamente y entender las necesidades y los deseos de las personas con las que trabaja (incluyendo clientes, pares, personal y superiores), además de los conocimientos tecnológicos imprescindibles en un ambiente computacional.

CESSI¹ desarrolló la Guía de Perfiles y Competencias de los puestos informáticos especificando las responsabilidades de un Director de Proyecto: “Gerenciamiento del Proyecto, incluyendo los grupos de procesos y las nueve áreas del conocimiento definidas por PMI: Alcance, Costo, Tiempos, Calidad, Integración, Riesgos, Comunicaciones, RRHH y Contrataciones.”

Dentro de una larga lista de conocimientos requeridos, se menciona “conocimientos de lineamientos del PMI en Gerenciamiento de Proyectos y sus principales estándares (PMBOK) y preferentemente Certificaciones relacionadas (PMP – Project Management Professional de PMI)”. La Guía desarrollada por el PMI es una descripción general de los conocimientos y las prácticas aplicables a la mayoría de los proyectos. Es el estándar más reconocido en todo el mundo sobre Project Management. Está aprobada como estándar por el American National Standards Institute (ANSI) y su finalidad principal es identificar el subconjunto de fundamentos de la Dirección de Proyectos, y las buenas prácticas.

¹ Cámara de Empresas de Software y Servicios Informáticos-
<http://www.cessi.org.ar>

De acuerdo a estos lineamientos, la disciplina de Dirección de Proyectos puede ser estudiada desde dos puntos de vista: sus procesos y sus áreas de conocimiento.

Los procesos están agrupados por etapas del Ciclo de Vida del Proyecto y para una mejor comprensión, los subprocesos de este ciclo fueron agrupados en áreas de conocimiento que los Directores de Proyecto deben dominar: Integración de los procesos, Alcances, Tiempo, Costo, Calidad, Recursos Humanos, Riesgos, y Contrataciones.

Este organismo también recomienda una bibliografía base que un Director de Proyecto debe dominar para alcanzar la certificación. Comprende más de 20 títulos, algunos compuestos de varios tomos en competencias que son transversales a la formación típica de un profesional de sistemas. Como si esto fuera poco, entran en juego los estándares de aseguramiento de calidad del proyecto informático mismo.

Hasta ahora las empresas han venido desarrollando una serie de acciones al respecto, en general capacitar a sus directores de proyecto o incorporarlos con los conocimientos requeridos. Esta última opción suele ser costosa, pero también costear la capacitación lo es. En un mundo empresarial enormemente competitivo, las ventajas cuentan. Las organizaciones son cada vez más conscientes de la importancia del conocimiento, por eso es necesario abrir caminos para la gestión del conocimiento.

La Sociedad del Conocimiento (o de la Información como también suele llamarse) es una realidad de fin del siglo XX. La cuestión reside en desarrollar y fortalecer las habilidades y el conocimiento haciendo uso eficiente de Tecnologías de la Información y las Comunicaciones (Oxbrow, 2004) y aplicando modelos de Gestión de Conocimiento adecuados a partir de la revisión del marco teórico existente.

1.1 Competencias profesionales y certificaciones en la Gestión de Proyectos Informáticos

El cambio, el aprendizaje y el liderazgo son los tres conceptos claves en la Nueva Economía (Editorial, 2000). Tradicionalmente la gestión de proyectos implicaba trabajo hecho dentro del alcance, tiempo y presupuesto planificados, variables conocidas como “el triángulo de prioridades o de hierro” (Archibald, 2000). Es decir, una visión táctica. Varios factores influenciaron directa e indirectamente en el crecimiento de la Gestión de Proyectos:

- El aumento de la competitividad a nivel global
- El énfasis en la eficiencia
- El impacto de las TIC's
- La adopción a nivel corporativo de prácticas de aseguramiento de calidad, outsourcing, downsizing, reingeniería de procesos de negocios y trabajo basado en equipos.

Uno de los indicadores de este crecimiento es la cantidad de miembros del Project Management Institute²: 78.000 en 2001, a 240.000 en 2007. Pero con su crecimiento y el surgimiento de nuevas disciplinas se hicieron necesarios nuevos enfoques. El principal de ellos es la visión estratégica, que ha abierto en los últimos años nuevas líneas de investigación (Morris & Jones, 2001). Según varios investigadores, resulta difícil este proceso de aprendizaje: examinar las propias prácticas y desafiar axiomas y verdades fuertemente afincadas desde los 60's (Morris, 2001; Delisle, 2001). Esta visión más actual ve a la Gestión de Proyectos como:

- Una disciplina basada en la experiencia que permite alcanzar eficiencia, efectividad e innovación a nivel proyecto y a nivel organizacional
- El éxito medido en forma multidimensional de acuerdo a los siguientes parámetros:
 - Posicionamiento estratégico de la gestión recurriendo a prácticas innovadoras
 - Alineamiento de los objetivos del proyecto con las prioridades estratégicas del negocio y de la organización
 - Balance de las demandas de los actores del proyecto direccionando el objetivo a sus expectativas
 - Inclusión de aspectos culturales, estructurales, prácticos y humanos
 - Arte y ciencia de convertir una visión en realidad.

A pesar de que la Gestión de Proyectos se inicio dentro de áreas de Defensa durante la Segunda Guerra, a partir de la década del 70 empieza a ser vista como una profesión. En los 80 se convierte en una disciplina “académica”. La Sociología ha definido una serie de atributos que diferencian una profesión de una ocupación: cuerpo de conocimientos, una cultura y código ético sostenido por una asociación profesional de reconocida autoridad en la comunidad, y sanciones establecidas por esa comunidad (Greenwood, 1966; Millerson, 1964).

Project Management o Gestión de Proyectos se ha ajustado en los últimos tiempos a esos indicadores, ya que cuenta con:

- PM Body of Knowledge Guide (PMI, 3ra. edición)
- Cultura sostenida por iniciativas de desarrollo profesional en varios países, conferencias y seminarios locales e internacionales
- Reconocida autoridad de certificaciones profesionales internacionales

² Project Management Institute (PMI) es una organización profesional dedicada a la mejora del estado del arte de la Gestión de Proyectos. Es la asociación internacional líder en esa profesión. Ha desarrollado estándares y certificaciones profesionales, y ha publicado la Guía del cuerpo de conocimientos de la Gestión de Proyectos.

- Oferta de carreras de postgrado y doctorados en Project Management.

En cuanto a competencias, empezaremos diciendo que deriva del latín “competere” que significa ser adecuado. Es bastante común que se utilicen indistintamente los términos conocimientos, habilidades y competencias (Goff, 2003). El error que la mayoría de las empresas comete es asimilar el concepto de conocimiento a competencia. Según el Diccionario Webster, competencia significa calificado, capaz o adecuado para un propósito estipulado. La Norma ISO 17024 (2003) que fija estándares de competencias, la define como la capacidad demostrada para aplicar conocimiento y/o habilidades, y donde sea relevante, atributos personales demostrados tal como están definidos en el esquema de certificación respectivo.

En forma más completa, Parry (1998) la define como la unidad de conocimientos, actitudes, habilidades y otras características personales, relacionados, que afectan la mayor parte de una tarea individual, que se correlaciona con el desempeño en esa tarea, que puede ser medida mediante estándares aceptados, que puede ser mejorada por medio de capacitación y desarrollo, y que puede ser desglosada en dimensiones: lo que un Project Manager puede aportar al proyecto a través de su conocimiento individual, la habilidad que puede demostrar en su desempeño al aplicar su conocimiento, y las características de personalidad que muestran su comportamiento en la tarea.

Un modelo de competencias de Gestión de Proyectos debe ser desarrollado a partir del comportamiento observable de profesionales exitosos en Gestión de Proyectos de un área específica, en nuestro estudio, la de proyectos informáticos. Su objetivo es guiar la capacitación, identificando diferencias entre los requerimientos del puesto y los niveles de habilidad incumbentes, a la vez que proveer a la organización de un método que le asegure que los Project Managers más capacitados son asignados a los proyectos críticos más complejos.

Las competencias no tienen que ser sólo consideradas para los Project Managers. Cada integrante del equipo de proyecto informático debe ser competente en su rol. Un Project Manager extraordinariamente competente no podría compensar la falta de competencia en el desempeño de los roles de su equipo.

El proyecto de investigación liderado por la Universidad de Sidney, Developmental Assessment of Project Management Competence, llevado a cabo en 1996, financiado por el Australian Research Council, el Australian Institute of Project Management, varias empresas líderes, y las organizaciones IPMA (International Project Management Association), APM (Association for Project Management), y PMI, fue uno de los primeros proyectos a nivel internacional sobre

Gestión de Proyectos y es el que da base al objetivo de esta investigación de sistematizar la Gestión de Proyectos dentro del marco de la Gestión de Conocimientos (KM). Su objetivo era identificar perfiles de conocimiento subyacente, actitudes y comportamientos que condujeran a un alto desempeño de los roles de la Gestión de Proyectos y que a la vez proveyeran un marco para el logro de competencias como base para definir perfiles de un equipo de proyectos efectivo.

2. LINEAS DE INVESTIGACION y DESARROLLO

La investigación se enmarca en el propósito de integrar los ejes teóricos de la Gestión de Proyectos Informáticos y el desarrollo y actualización continua de los Project Managers, con los modelos de Gestión del Conocimiento, como aplicación en el ámbito del ingeniero en Sistemas de Información de las herramientas propuestas por las nuevas Tecnologías Educativas colaborativas.

3. RESULTADOS OBTENIDOS/ESPERADOS

Reflexiones preliminares

La redefinición de la que hablamos en párrafos anteriores se basa en la consideración del rol de la Gestión de Proyectos a nivel estratégico de la organización. Este nuevo enfoque fue llamado “core competences” por Prahalad y Gary (1990). Leonard-Barton (1992) especificó que ciertas capacidades son consideradas como “core competences” si diferencian estratégicamente a la organización. Se distingue de la visión clásica que considera solo el valor tangible de las competencias en que esta nueva perspectiva sostiene el criterio de los recursos internos de una organización como únicos y contribuyentes a su ventaja competitiva (Foss, 1997).

Core competences abarcan: prácticas de sus recursos humanos, especialización, aprendizaje individual y colaborativo, cultura, trabajo en equipo y confianza. El conocimiento organizacional y las prácticas colectivas, según Hanel (1994) es una acumulación de experiencias de aprendizaje y comprende el conocimiento tanto tácito como explícito. Es dinámico y variable en el tiempo (Bogner; Turner & Crawford, 1994).

Según las últimas investigaciones de Gestión de Proyectos, las empresas debieran capitalizar las habilidades y experiencias de la propia Gestión de Proyectos convirtiéndolos en core competences a nivel organizacional mediante TIC's, dentro de la filosofía de Gestión del Conocimiento.

Muchas empresas están sufriendo o han sufrido las consecuencias del aprendizaje “a golpes” de sus Project Managers, por lo que ya han empezado a captar que los proyectos exitosos no suceden porque sí. Se facilita el éxito mediante el aprendizaje

experiencial, y el refuerzo de prácticas organizacionales. La transferencia de habilidades es el vínculo entre capacitación y resultados en el negocio. Pero esta transferencia es una estrategia a largo plazo.

Como resultado de esta primera etapa de la investigación hemos establecido un marco conceptual actualizado en forma de compilación de las definiciones de competencia para la Gestión de Proyectos (PM) según las asociaciones profesionales de mayor peso internacional, comentando el nuevo rol de los Directores de Proyecto y la aparición en este contexto de las Certificaciones profesionales y las "core competences". Gestión de Proyectos en forma conjunta con Gestión de Calidad y Gestión del Conocimiento, envuelven la mejora de las prácticas del profesional informático al compartir conocimientos organizacionales tácitos y explícitos. De lo contrario se llega a un repositorio de conocimiento fracturado y no compartido. Las Tecnologías de la Información aplicadas al proceso de enseñanza-aprendizaje colaborativo mediante comunidades de aprendizaje y comunidades de práctica se presentan como alternativas eficaces. En la siguiente etapa de la investigación se explorarán las competencias establecidas en el ámbito profesional de proyectos informáticos para obtener el conjunto que será incluido en el modelo conceptual a desarrollar aplicando Tecnologías de la Información para educación profesional continua. En una tercera etapa se buscará determinar cuál es el modelo de Gestión del Conocimiento más adecuado para representar estas responsabilidades e incumbencias de los directores de proyectos.

4. FORMACION DE RECURSOS HUMANOS

Ya hemos mencionado en la sección I la problemática detectada respecto a los requerimientos que las consultoras de sistemas de información de primera línea exigen a los Directores de Proyectos Informáticos. Esto hace imprescindible desarrollar y fortalecer las habilidades y el conocimiento de los profesionales informáticos durante su formación universitaria y a lo largo de su vida profesional haciendo uso eficiente de Tecnologías de la Información y las Comunicaciones (TIC's) aplicando conjuntamente modelos de Gestión de Conocimiento en un entorno de educación continua.

El objetivo general del proyecto es contribuir a la formación de recursos humanos en el área informática proveyendo marcos tecnológicos que faciliten su educación y además mediante su concientización, desde varias perspectivas y en diferentes etapas de su vida: la necesidad de educación continua postgraduación inculcada durante la formación universitaria de nuestros ingenieros, el aprovechamiento de las ventajas del aprendizaje colaborativo durante su desempeño profesional, y la contribución al campo disciplinar mediante investigación académica. Se prevé además

la incorporación de investigadores noveles (ayudantes de cátedra, graduados de la carrera, y colaboradores de otras disciplinas) para el inicio de su formación como investigadores en un proyecto interdisciplinar.

5. BIBLIOGRAFIA

Archibald R. (2000). *What CEO's must demand to achieve effective Project Management*, Iberoamerican Project management Forum, México.

Borner W. (1994). *Core Competences and Competitive Advantages*, J.Wiley & Sons.

Delisle C. (2001). *Communication and success in virtual projects*, Doctoral Thesis, Univ. Calgary.

Editorial (2000). Change Time, *Fast Company*, Vol. 39.

Foss N. (1997). *Resources, firms and strategies: a reader in the resource-based perspective*, Oxford University Press.

Goff S. (2003). *Distinguishing PM Competence in Training and Development, organizational assessments and certifications*, <http://www.projectexperts.com>

Greenwood E. (1966). *Elements of professionalization*, Prentice Hall.

Hamel G. (1994). *The Concept of Core Competences*, J.Wiley & Sons.

Leonard-Barton, D. (1992). Core Capabilities and Core Rigidities: A Paradox in Managing New Product Development. *Strategic Management Journal*, 13.

Millerson G. (1964). *The Qualifying Association: a study of professionalization*, Routledge.

Morris P. & Jones I. (2001). Current Research Directories in the Management of Projects, *UMIST*, Vol. 2001, Univ. De Calgary, Canadá.

Morris P. (2001). The management of Projects: the new models, Center for the Management of Projects, *UMIST UK*, 2001.

Oxbrow N. (2004). *The Knowledge proposition*, <http://www.conspectus.org>

Parry, S. (1998). Just What Is a Competency? (And Why Should You Care?) *Training* (June).

Prahalad & Gary (1990). The Core Competence of the Corporation, *Harvard Business Review*, vol. 68, no. 3.

Turner D. & Crawford M. (1994). *Managing current and future competitive performance: the role of competences*, J.Wiley & Sons.

Uso del lenguaje unificado en la modelización de sistemas operativos

Inés Casanovas, Nicanor Casas, Graciela De Luca

inescasanovas@gmail.com, pepecasas2002@yahoo.com.ar, graciela.edl@gmail.com

Universidad Tecnológica Nacional - FRBA

Departamento de Ingeniería de Sistemas

CONTEXTO

Este proyecto está siendo desarrollado en forma interdisciplinaria por las cátedras de Sistemas Operativos y Análisis de Sistemas de la carrera de Ingeniería de Sistemas de la Universidad Tecnológica Nacional- Regional Bs.As. Están involucradas las dos cátedras ya que en la primera se aplican todos los conceptos y prácticas de Sistemas Operativos y en la segunda se desarrollan los conceptos teóricos de Orientación a Objetos, UML en particular, ya que preliminarmente se pensó en la utilización de Rational como herramienta de modelado.

RESUMEN

El abordaje de la enseñanza de sistemas operativos orientados a objetos choca contra un elemento de singular peso: el escaso material de estudios que apunte en forma directa al centro del problema. Es por ello que en la actualidad las diferentes cátedras abordan la enseñanza de sistemas operativos tradicionales y muy pocas complementan con sistemas operativos orientados a objetos. En los trabajos existentes la arquitectura de integración del sistema está basada en una máquina abstracta que ofrece una funcionalidad mínima, y un sistema operativo que debe extender la funcionalidad de la máquina, chocando con la funcionalidad de arranque del sistema y en la utilización de herramientas de aplicación parcial, pues las estructuras existentes se basan en conceptos tradicionales de proceso, posición de memoria y archivo.

El objetivo es constatar que el modelado de un sistema operativo mediante técnicas de Lenguaje Unificado presenta inconvenientes para definir los procedimientos, que requiere

un sistema operativo, en especial todo lo relacionado con la interacción de bajo nivel, siendo necesaria la creación o redefinición de herramientas que permitan la modelización de un sistema operativo general y que tenga la visión de un sistema integral orientado a objetos como parte fundante de un sistema operativo distribuido.

Palabras Claves: *UML (lenguaje unificado de modelado), Flexibilidad, Máquina abstracta, Encapsulamiento, Reusabilidad.*

1. INTRODUCCION

La característica principal del estudio a desarrollar es constatar que el modelado de un sistema operativo a través de técnicas de lenguaje unificado presenta inconvenientes para definir los procedimientos a bajo nivel que requiere un sistema operativo, en especial todo lo relacionado a la interacción entre lenguajes cercanos a la máquina, o lenguajes de bajo nivel.

Por otro lado todo software de base debe adaptarse a la modularidad de una máquina que en principio puede tener estructuras muy diferentes con otras que sirvan para la misma función.

Los modelos UML tienen significado para el análisis lógico y para la implementación física, facilitando con ello las construcciones de implementación, pero dificultando en muchos de los casos las estructuras de construcción.

Suzuki y Yamamoto [SUZ99], ya en 1999 proponían una extensión de UML para soportar aspectos, considerando a los aspectos como un clasificador en un meta-modelo.

Los objetos proporcionan modularidad y encapsulación, dado que separan claramente

la interfaz de construcción con la interfaz de implementación.

La comunicación cuenta con un mecanismo uniforme de alto nivel, la invocación de métodos o paso de mensajes entre objetos, que es bien conocido y para el que existen técnicas para conseguir implementaciones eficientes.

Los objetos proporcionan un medio adecuado para compartir información. La ventaja respecto a los ficheros compartidos es una semántica de más alto nivel y una mejor protección ante errores de programación.

Los sistemas estructurados en términos de objetos evolucionan fácilmente (más fáciles de extender y mantener), por reemplazo de sus partes. Además los subsistemas y objetos individuales podrán ser reutilizados con facilidad

Los sistemas de software de tiempo real, en particular los sistemas operativos, representan un desafío de modelización y diseño lógico porque tienen comportamientos orientados a eventos y contienen limitaciones temporales y de recursos. El análisis basado en escenarios (en todas sus modalidades) proporciona soluciones a las problemáticas de alto nivel, mientras que la programación orientada a objetos proporciona las adecuadas a las de bajo nivel. Queda pendiente de resolver el gap que se produce al tratar de modelizar estos sistemas en base a las técnicas propuestas por UML.

Por tanto entre los inconvenientes para definir los procedimientos a bajo nivel debemos tener en consideración especial todo lo relacionado con:

- la interacción entre lenguajes de bajo nivel
- las causas, efectos y responsabilidades a nivel de comportamiento de sistemas donde los escenarios pueden cambiar en tiempo de ejecución, y se requiere capacidad de refinamiento durante la variación de estos escenarios.

No importa cuan completo sea el conjunto de facilidades que proporcione un lenguaje de referencia, siempre se querrán realizar

extensiones al mismo por varias razones que son muy entendibles, por ejemplo:

- Para mejorar la visión de los diagramas que se desarrollen
- Porque existen áreas oscuras que no pueden ser representadas por el lenguaje.
- Porque la evolución de las necesidades operativas requieran un mejor interacción entre las diferentes estructuras.

Por otro lado los estereotipos representan una nueva clase de elemento de modelado con la misma estructura de un elemento existente, pero con restricciones adicionales, un interpretación diferente de algún elemento, como ser un ícono y un tratamiento diferente de los generadores de código.

En la mayor parte de los sistemas operativos actuales, la aplicación de las herramientas de modelado y en especial las correspondientes a las de orientación a objetos se realiza de una manera parcial, dado que se sustentan en los conceptos tradicionales de proceso, posición de memoria y archivo. Por otro lado los mecanismos y políticas definidos por el sistema operativo están codificados de manera rígida en el mismo, y esto reduce la posibilidad de adaptarlos al entorno en el que va a prestar servicios.

Consideramos que una aproximación interesante a la integración de todas las tecnologías, la del modelado, las políticas y los mecanismos, es la creación de un sistema que ofrezca un soporte que permita la generación de un sistema operativo general y que tenga la visión de un sistema integral orientado a objetos como nervio central para el desarrollo de un sistema operativo distribuido.

En este sistema, todos los elementos, interfaces de usuario, aplicaciones, lenguajes, etc., comparten un mismo paradigma. La integración de la máquina abstracta y el sistema operativo se consigue a través de la arquitectura reflectiva de la primera, que permite definir la máquina como una estructura de objetos perfectamente cuantificables.

Uno de los inconvenientes que se registran en los trabajos que preceden es que la arquitectura para realizar la integración del sistema está basada en una máquina abstracta que puede o no estar orientada a objetos, pero que fundamentalmente ofrece una funcionalidad mínima, y un sistema operativo que tiende a ser orientado a objetos, que debe extender la funcionalidad de la máquina chocando con la funcionalidad de arranque del sistema.

Desde un punto de vista general un sistema operativo puede verse como la interacción entre dos componentes bien definidos: una base y un meta-nivel.

La base intenta solucionar el problema de las aplicaciones, tomándolas como entidades que reflejan la problemática real a resolverse y el meta-nivel es el responsable de que ese sistema base funcione manteniendo la información y describiendo de que manera debe resolver los conflictos el sistema base.

Si se integran los lenguajes de programación, como se indicó anteriormente, el programa fuente correspondería al sistema base y el intérprete del lenguaje corresponde al meta-nivel.

Al ser los sistemas de tiempo real orientados a eventos, la determinación de las especificaciones basadas en lenguaje natural y el proceso de análisis y modelización los diferencia notablemente del modelado de software de gestión, pues las técnicas tradicionales de determinación de requerimientos proporcionan modelos de alto nivel.

Se requiere un entendimiento detallado de dependencias, concurrencias, tiempo de respuesta y uso de recursos, es decir, cobran destacada importancia los requerimientos no funcionales. En este tipo de software los requerimientos han sido clasificados como requerimientos de comportamiento y requerimientos temporales (Ekelin C. y Jonson J, 1999, Real Time Systems Constrains, Alexandria, VA). Los primeros, como se relacionan con la funcionalidad, se identifican rápidamente; los temporales son el

resultado de la interacción de los objetos componentes y mas difíciles de definir.

Una visión panorámica de los enfoques actualmente discutidos en los congresos y áreas de investigación del dominio de los Sistemas Operativos muestra los siguientes métodos:

a) UML Use Cases: es una aproximación basada en casos donde los actores definen requerimientos que son modelizados en casos de uso y expandidos en escenarios que esclarecen relaciones específicas entre actores y el sistema a nivel individual, usando lenguaje natural.

b) Diagramas de transición de estados: Glintz ha promovido el uso de estos diagramas como una mejora para la representación de los requerimientos de Sistemas Operativos a partir de un trabajo anterior de Harel. Este método propone la representación de escenarios sencillos, compuestos y abstractos, usando diagramas y texto formalizado, que permitiría vistas a nivel de detalle de eventos y sus transiciones. En su versión original no definía formas para manejar requerimientos temporales pero es posible de ser ampliado.

c) Notación jerárquica gráfica: fue propuesto por Regnell [REG96] y en este método, el escenario presenta un grafo de episodios conectados, donde se definieron notaciones de tiempo para mostrar excepciones e interrupciones, y cada evento a su vez es un diagrama de secuencia de mensajes. Se refleja también las dependencias entre escenarios y el entorno de ejecución.

d) Mapas de Casos de Uso (UCM): describen visualmente relaciones causales entre responsabilidades impuestas en estructuras de componentes abstractos. Estas responsabilidades podrían ser tareas, acciones, operaciones etc. Mientras que los componentes pueden ser entidades de software (objetos, procesos etc.) o no (hardware, por ejemplo). Las relaciones causales vinculan

precondiciones o eventos disparadores, con postcondiciones y resultantes, secuencialmente o concurrentemente. Básicamente, son caminos que muestran el desarrollo o evolución temporal de un escenario dentro de un Caso de Uso, es decir es un modelo dinámico para representar las variaciones de comportamiento y de estructura.

2. LINEAS DE INVESTIGACION y DESARROLLO

El grupo de investigación desarrolla en la actualidad un sistema operativo de características didácticas el cual basa su análisis en través del uso de metodologías ágiles. Se presenta así, debido al intercambio en diferentes congresos, el interés de realizar un intento de llevar el mismo a diseño a través de UML, si bien existen muchos informes sobre las falencias que el lenguaje unificado de modelización tiene, pero los mismos no se refieren a sistemas operativos, ni a software de base.

3.RESULTADOS OBTENIDOS / ESPERADOS

3.1 Resultados alcanzados

Dado que el proyecto está en sus comienzos, hasta el momento no se han visto resultados, sino que todo se basa sobre el supuesto de que las técnicas de modelado, en función de la hipótesis de trabajo de este estudio, resultarían inadecuadas para representar claramente todos los atributos presentes en tiempo real.

3.2 Objetivos

Permitir que los alumnos puedan definir un universo de discurso, aplicando conceptos que son claves de la aplicación, sus propiedades internas y las relaciones entre cada una de ellas, generando a posteriori nuevas relaciones y establecer la factibilidad de cada una de ellas.

Permitir que los alumnos obtengan y generen conocimientos que permitan visualizar un

sistema operativo en forma integrada a través de diagramas.

Permitir que los alumnos observen un sistema operativo bajo una herramienta de diseño de uso general ingresando diferentes modelos de máquinas y generando vistas estáticas de diagramas de clase.

Lograr que los alumnos puedan realizar estudios avanzados sobre sistemas operativos aprovechando los diferentes modelados básicos estructurados anticipadamente, para ser incrementados y corregidos.

Permitir implementar áreas estructurales, dinámicas y de gestión de modelos.

Permitir el análisis detallado de extensiones de UML que permitan salvar problemas, si los hubiera, y/o generar nuevos diagramas, no definidos, para facilitar la confección de software de base.

4. FORMACION DE RECURSOS HUMANOS

El grupo de investigación espera transferir los conocimientos alcanzados a los alumnos que cursan las materias de Análisis de Sistemas y Sistemas Operativos para proveerles de una estructura que enlace ambas materias.

Intercambiar información con Universidades Nacionales y Privadas del ámbito nacional e internacional publicando los avances que se realicen en diferentes congresos, teniendo como objetivo extender y realizar convenios de colaboración que permitan ampliar conocimientos de docentes e investigadores.

5. BIBLIOGRAFIA REFERENCIADA Y DE CONSULTA

[AG96] Allen Robert y David Garlan, "The Wright Architectural Description Language", Technical Report, Carnegie Mellon University. 1996.

[ALL97] Allen Robert. "A formal approach to Software Architecture". Technical Report, CMU-CS-97-144, 1997.

[BIN93] Binns Pam, Matt Englehart, Mike Jackson y Steve Vestal. "Domain- Specific Software Architectures for Guidance, Navigation, and Control". Technical report, Honeywell Technology Center, <http://www-ast.tds-gn.lmco.com/arch/arch-ref.html>, 1993.

[BUR98] Buhr R., 1998, Use Case Maps as Architectural Entities for Complex Systems, IEEE, Transactions on Software Engineering 24(12) pag. 1131-1155 - 1998.

[DH01] Eric Dashofy y André van der Hoek. "Representing Product Family Architectures in an Extensible Architecture Description Language". En Proceedings of the International Workshop on Product Family Engineering (PFE-4), Bilbao, España, 2001.

[HAR87] Harel D., State Charts: a visual formalism for complex systems. SCI Program, 1987

[JAC92] Jacobson J. et al., Object Oriented Software Engineering: a use case driven approach, Addison Wesley), 1992

[GLI00-a] Glinz Martin. "Problems and deficiencies of UML as a requirements specification language". En Proceedings of the 10th International Workshop of Software Specification and Design (IWSSD-10), pp. 11-22, San Diego, noviembre de 2000.

[GLI00-b] Glintz M., 2000, Improving the Quality of Requirements with Scenarios, Proceedings of the 2nd. World Congress for Software Quality, Japón

[Mon98] Monroe Robert. "Capturing software architecture design expertise with Armani". Technical Report CMU-CS-163, Carnegie Mellon University, Octubre de 1998.

[OOS99] OOSPLA'99, Workshop #23: "Rigorous modeling and analysis with UML: Challenges and Limitations". Denver, Noviembre de 1999.

[PIN89] Pinkert J. & L.Wear – Operating Systems: Concepts, Polices and Mechanisms,

Englewood Cliffts NJ, Prentice Hall – First Edition – 1989

[REG96] Regnell et al., A Hierarchical Use Case Model with Graphical Representation, Proceedings of IEEE Internacional Symposium on Engineering of Computer Based Systems, 1996

[RUM99] Rumbaugh James, Ivar Jacobson, Grady Booch – El lenguaje unificado de modelado, Manual de Referencia – Addison Wesley - 1999

[SMC00] Standard Microsystems Corporation *Application note 6.12*, 2000

[SUZ99] Suzuki J. y Y. Yamamoto - Extending UML with Aspects: Aspect Support in the Design Phase - ECOOP Workshop on Aspect-Oriented Programming - 1999.

Revistas y Publicaciones

Diomidis Spinellis - Computer Standars & Interfaces 20:1-8 November 1998 "A Critique of the Windows Application Programming Interface" University of the Aegean

Internet

Akehurst David y Gill Waters. "UML deficiencies from the perspective of automatic performance model generation". OOSPLA'99 Workshop on Rigorous Modeling and Analysis with the UML: Challenges and Limitations, Denver, Noviembre de 1999.

<http://www.cs.kent.ac.uk/pubs/1999/899/content.pdf>,

Douglas Bruce Powel. "UML – The new language for real-timed embedded systems".

<http://wooddes.intranet.gr/papers/Douglass.pdf>, 2000.

BLEND-LEARNING COMO HERRAMIENTA PARA DISMINUIR LA DESERCIÓN DE ALUMNOS

Ing. Inés Casanovas¹, Ing. Miguel Benegas², Lic. Gladys Fernandez³

(1) Universidad Tecnológica Nacional –FRGPacheco
inescasanovas@gmail.com

(2) Universidad Tecnológica Nacional –FRGPacheco
mbenegas@frgp.utn.edu.ar

(3) Universidad Tecnológica Nacional –FRGPacheco
gladyaf@yahoo.com.ar

Universidad Tecnológica Nacional

Facultad Regional Gral. Pacheco

Departamento de Ciencias Básicas. Cátedra Informática I

Proyecto de Investigación Homologado EDINGP806 - (G016) 2008 - 2009

CONTEXTO

Desde su nacimiento la UTN (Universidad Tecnológica Nacional) se ha caracterizado por contar con una matrícula con un alto porcentaje de alumnos que trabajan. La Facultad Regional Gral. Pacheco por hallarse inmersa en un polo industrial, posee un 90% de alumnos con una inserción laboral superior a las 40 hs semanales. Teniendo en cuenta los horarios de dictado, el tiempo que los educandos le pueden dedicar al estudio es escaso. En 1999 se inicia en la cátedra de Informática I del departamento de Ciencias Básicas de la Universidad Tecnológica Nacional Facultad Regional Gral. Pacheco con un trabajo metódico y sistemático de exploración y análisis de las estadísticas, a esa fecha como experiencia piloto, que permitió detectar una tasa de abandonadores superior al 40 % de la matrícula en los alumnos del primer año así como un promedio de duración de la carrera de más de 10 años. Para poder identificar las causas de ese alto índice de abandono se realizaron durante seis meses entrevistas informales con los alumnos y de ellas se pudo determinar que la causa más mencionada era la incompatibilidad de las obligaciones laborales con las obligaciones de cursada. Así, surge la necesidad de ofrecerles estrategias pedagógicas y herramientas que les permitan estudiar, reduciendo los tiempos presenciales, mejorando y ampliando los materiales que requiere para el autoaprendizaje. Para estos casos, la modalidad b-learning sería una alternativa altamente viable.

RESUMEN

La problemática de la deserción en los primeros años de las carreras de grado es sin duda uno de los mayores retos con que se enfrenta el docente universitario. Si bien se han realizado innumerables estudios sobre las variables intervinientes en el aprendizaje también se deben considerar las condiciones socioeconómicas y laborales de los alumnos ya que este incide significativamente en la concreción de los objetivos propuestos.

La búsqueda de recursos o estrategias por parte del docente para disminuir el grado de deserción se ve facilitada por la incorporación de las Nuevas Tecnologías de la Información y la Comunicación. El

aprendizaje a distancia bajo la modalidad B-Learning es sin duda una de las herramientas más valiosas con las que cuenta la enseñanza y el docente así como las instituciones educativas.

El presente proyecto fue planteado con el fin implementar el uso de un aula de educación virtual así como un programa que sustente su uso para el dictado de la cátedra Informática I de la Licenciatura en Organización Industrial en la Facultad Regional Gral. Pacheco.

Palabras claves: abandonadores, B-Learning, estudio, trabajo, metodologías

1. INTRODUCCIÓN

El uso de las TICs se ha incrementado en los últimos años y la Regional no puede quedar ajena a los beneficios que esta modalidad le traerá, implementando programas y contenidos educativos virtuales, desarrollando técnicas educativas a través de la telemática, orientando los esfuerzos con vistas en perfil del profesional capacitado acorde con las nuevas tecnologías e incursionando en nuevos mercados de oferta educativa.

Los destinatarios fueron alumnos cursantes que están en constante interacción con las materias que cursan, alumnos abandonadores saturados en sus horarios de cursada y trabajo, y quienes, por falta de tiempo, ven frustradas sus posibilidades de crecimiento profesional. Esto redundara en posibilidad de ver disminuido el índice de abandono en la cátedra.

El proyecto se sustentó sobre bases teóricas sólidas que refuerzan la creación de este tipo de cursos y la utilización de las nuevas tecnologías aplicadas a la educación superior utilizando metodologías constructivistas basadas en fuentes expertas, tanto nacionales como internacionales.

La plataforma seleccionada, para el desarrollo de la experiencia, tendiendo a sus características, prestaciones específicas y siendo un Sistema de gestión de aprendizaje de alta performance ha sido Moodle (Modular Object-Oriented Dynamic Learning Environment - Entorno de Aprendizaje Dinámico Orientado a Objetos y Modular). Ésta, por ser de código abierto, basada en una pedagogía social

constructivista, donde la comunicación tiene un espacio relevante (Weller 2002) para facilitar la construcción del conocimiento, se adecuó a las necesidades de nuestro proyecto educativo.

El proyecto se desarrolla a lo largo del periodo febrero del 2008 y Febrero del 2010, con base institucional el Departamento de Ciencias Básicas. El equipo principal que lo integra de carácter multidisciplinario, abarcando los siguientes perfiles técnicos: comunicación educativa y pedagogía y producción audiovisual y gráfica, redes informáticas.

Las primeras etapa del proyecto ya concluidas, consistieron en recabar información y analizarla con enfoque descriptivo analítico cuantitativo, redefinir los marcos pedagógicos, y la implementación y documentación de la experiencia piloto. Los resultados obtenidos integrarán la piedra angular para la implementación tecnológica-educativa, la asistencia a los docentes en el diseño de los materiales didácticos y el seguimiento de la participación de los alumnos. Para iniciar actualmente el proceso de análisis y estudio de la información administrativa producida por la plataforma y el análisis interpretativo de la experiencia y sus resultados académicos.

1.1 Objetivos

- Revisar y reconsiderar los marcos pedagógicos del blended-learning, con énfasis en los factores motivantes y las barreras para su adopción por docentes y alumnos
- Analizar los aspectos tecnológicos, económicos y culturales que enmarquen la implementación de la experiencia
- Incorporar las Nuevas Tecnologías de la Información y Comunicación al ámbito educativo de la Universidad Tecnológica Nacional Facultad Regional Gral. Pacheco mediante la implementación de un entorno de Educación Virtual en plataforma Moodle (Open Source) para la materia Informática I de la carrera de Licenciatura en Organización Industrial, y la generación de los contenidos correspondientes.
- Sistematizar, evaluar y documentar la experiencia de docentes, técnicos y alumnos, y los resultados.
- Contribuir con nuestro aporte didáctico a la mejora de la problemática detectada por el Departamento de Ciencias Básicas de la Facultad, y de la comunidad educativa de la Universidad.

1.2 Metodología

La puesta en marcha del modelo de procesos de B-Learning consistió en dos etapas:

Etapas 1: La definición del modelo conceptual y funcional a implementar.

Etapas 2: La implantación efectiva del nuevo modelo. Un modelo global de B-Learning presenta los siguientes componentes cuyo papel en la institución y funcionamiento fueron necesarios concretar ya que condicionaban el resto de decisiones y actuaciones que se llevaron a cabo:

1- Definición del modelo: definir el modelo de aprendizaje apropiado y la estructura de gestión asociada.

2- Adecuación de los contenidos: redefinir los contenidos que los docentes utilizaban en su cátedra

3- Gestión de contenidos: Determinar el modo y forma en el que se producen u obtienen los contenidos que van a ser utilizados en el proceso educativo. Este componente constituye un pilar especialmente importante del proceso desde el punto de vista tecnológico.

4- Configuración del aula virtual: decidir la estructura y distribución de los temas dentro del aula virtual en función de las características y herramientas que ofrece la plataforma seleccionada

El uso de una estructura por módulos fue acordada por el equipo multidisciplinario por considerarla de fundamental importancia por su posible impacto en la integración de contenidos y la gestión de los alumnos dentro del aula virtual. Los temas a desarrollar encapsulados en una estructura modular permitieron una integración de los contenidos en forma vertical y horizontal.

5- Implementación:

Se realizó la presentación de la propuesta a los alumnos, la cual consintió en la descripción de los siguientes aspectos:

- Modalidad de cursada
- Presentación del aula virtual tal como se pudo observar en las figuras 1 y 2, forma y datos de acceso.
- Presentación del equipo docente y sus roles (Profesor, Tutores y soporte técnico)
- Presentación del cronograma de cursada, donde se indicó el temario, su fecha de inicio y de fin así como las actividades a realizar. En el mismo estaban establecidas las clases presenciales teóricas prácticas en el laboratorio en la cual se desarrollaron los contenidos de cada unidad temática y se asesoró en la resolución de las actividades propuestas en el aula virtual. Esta combinación de recursos tecnológicos (no presenciales) con encuentros presenciales entre el docente y el alumno, posibilitaron un proceso guiado de aprendizaje que fue disminuyendo a medida que el alumno ganó confianza y experiencia (Kirschner et al., 2006; Heinze et al., 2004).

Las actividades fueron de tres tipos bien definidas:

- Interacción en el aula a través de las herramientas que ofrece la plataforma: foros, sala de chats y mensajes
- Resolución en forma individual un conjunto de problemas que lo ayudaron a comprender cada unidad temática.
- Resolución en forma grupal, diferentes problemas integradores diseñando una solución adecuada. Esta actividad tuvo carácter obligatorio y debió ser entregada para su calificación dentro de los plazos establecidos

Como ninguna de las clases antes mencionadas tuvo carácter de obligatorio, el alumno pudo seguir el cronograma de temas dictados y prácticos presentados

a través del aula virtual cuando sus actividades laborales le impedían la asistencia presencial.

El equipo docente fue responsable de mantener actualizada el aula virtual del curso, de modo que los estudiantes tuvieron permanente acceso a:

- Información general de la materia: contenido, bibliografía, equipo de cátedra, etc.;
- Información administrativa: horarios, cronograma, metodología de trabajo y evaluación
- Recursos: transparencias de las clases teóricas, apuntes, publicaciones de lectura obligatoria, trabajos prácticos, acceso a los recursos de la web.

El curso se evaluó a través de:

- La interacción en el aula:
 - Participación en los foros: de dudas, Cafetería y temático.
 - Uso de la sala de Chat
- Aprobación de los trabajos prácticos
- Así, el alumno estuvo habilitado a rendir los exámenes presenciales teórico/prácticos correspondientes.

Durante toda la experiencia, se ha podido realizar un seguimiento minucioso de los alumnos, utilizando los registros de rastreo de actividades incluidos en la plataforma Moodle, esto le permitió recabar información sobre la situación de los alumnos.

2- LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

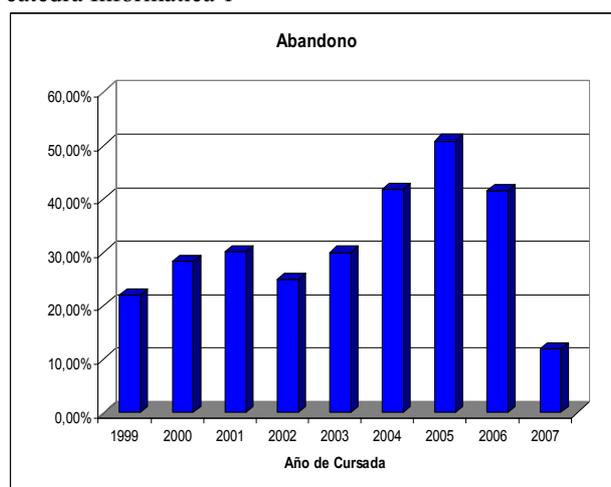
Con esta problemática detectada de estudiantes de nivel universitario que abandonan su carrera por incompatibilidad laboral se abrió un proyecto de investigación que se radicó en el Departamento de Cs. Básicas de UTN – FRGP con Resolución de Consejo Académico –FRGP N° 501/07 y posteriormente fue aprobado en el Programa de Incentivos mediante Resolución de Rectorado EDINGP806 - (G016). Se encuadra en la línea de Investigación Aplicada y en la disciplina Tecnología Educativa de acuerdo al esquema definido por Secretaría de Ciencia y Tecnología de la universidad. El resultado final de este trabajo bianual nos permitirá incorporar las Nuevas Tecnologías de la Información y Comunicación (NTIC) al ámbito educativo de la Universidad Tecnológica Nacional-Facultad Regional Gral. Pacheco mediante la implementación de un entorno de Educación Virtual en plataforma Moodle (Open Source) para la materia Informática I de la carrera de Licenciatura en Organización Industrial, y la generación de los contenidos correspondientes, sistematizando y documentando la experiencia de docentes, técnicos y alumnos, como aporte al cuerpo de conocimiento de las metodologías educativas.

3- RESULTADOS OBTENIDOS/ESPERADOS

Si bien la implementación de una asignatura a ser impartida bajo la modalidad B-Learning con soporte telemático requiere, por parte del equipo docente, un proceso sistemático y estructurado que abarca tanto la plataforma como la gestión, comunicación y evaluación de los alumnos pasando por la elaboración de un material adecuado al medio utilizado, redundando en

beneficios altamente positivos. Los resultados observados demuestran que se ha logrado disminuir la deserción cuando las causas son mayormente, la falta de disponibilidad de horarios de cursada más flexibles. Del estudio y análisis de los registros existentes en la universidad se observó claramente el aumento de estudiantes que abandonan sus estudios en la cátedra (Figura 1). Se detectó además que la tendencia es creciente, notándose una leve disminución en último año 2006, pero se observa un índice notablemente bajo para la cursada del año 2007, donde se implemento la experiencia de la cursada bajo la modalidad B-Learning.

Figura 1: Tasa de abandono: Años 1999 a 2007 para la cátedra Informática 1



Fuente: Sysacad, Sistema Académico de la UTN-FRGP

El uso del B-learning, como apoyo a la formación presencial, se adapta perfectamente a los modelos pedagógicos propuestos, cuya finalidad no es otra que la construcción del propio conocimiento por parte del alumno. Por ello podemos considerar que el proceso de aprendizaje llevado a cabo en los entornos virtuales, el conocimiento y la formación que se posea de la metodología a utilizar resulta una variable clave para el éxito docente.

La metodología aquí expuesta, aplicada a las cursadas virtuales plantean una modificación de los roles de los actores que intervienen en el proceso educativo (Lopez de Parra, en línea), por un lado el docente, que asume el rol de tutor y ejerce sus funciones como orientador, mediador y facilitador de los procesos de aprendizaje; por otro lado y como eje central del proceso de aprendizaje, el alumno llevando a cabo procesos de análisis, búsqueda y organización de la información tendientes a resolver las propuestas planteadas con el fin de desarrollar sus competencias.

Si bien el proceso es complicado y laborioso al inicio, redundando en ventajas no disponibles de otra manera tales como facilidad para incorporar materiales y recursos interactivos, mayor libertad horaria para los alumnos, y favorece el carácter multidisciplinar de los estudios, la actitud crítica de los educando (Johnson et al, 1984,

Palloff & Pratt, 2001, Weller, 2002) y el trabajo colaborativo.

En este momento de la investigación podemos resumir los logros alcanzados:

- Hemos revisado y reconsiderados los marcos pedagógicos del blended-learning, con énfasis en los factores motivantes y las barreras para su adopción por docentes y alumnos.
- Hemos analizado los aspectos tecnológicos, económicos y culturales que enmarcan la implementación de la experiencia
- Hemos incorporado las Tecnologías de la Información y Comunicación (NTIC) al ámbito educativo de la Universidad Tecnológica Nacional - Facultad Regional Gral. Pacheco mediante la implementación de un entorno de Educación Virtual en plataforma Moodle (Open Source) para la materia Informática I de la carrera de Licenciatura en Organización Industrial
- Hemos implementado un curso en línea y generado los contenidos correspondientes.

Nos encontramos en pleno proceso de:

- Sistematizar, evaluar y documentar la experiencia de docentes, técnicos y alumnos, así como los resultados obtenidos hasta la etapa en curso.
- Contribuir con nuestro aporte didáctico a la mejora de la problemática detectada por el Departamento de Cs. Básicas de la Facultad, y de la comunidad educativa de la Universidad.

4- FORMACIÓN DE RECURSOS HUMANOS

- Contribución a la formación de recursos en el ámbito de la investigación ya que el equipo está integrado por una minoría de investigadores formados cuyo objetivo es iniciar y capacitar al resto de los integrantes sin experiencia en la actividad de investigación
- Transferencia de conocimientos y experiencias entre los integrantes de un grupo de investigación de carácter interdisciplinario (integrantes con formación en informática, en docencia universitaria, en diseño de materiales para educación a distancia y en Ciencias Básicas)
- Formación de docentes en el uso de plataformas e-learning, en actividades de tutoría, en la generación de contenidos de didáctica específica para educación no presencial y en la gestión administrativa y organizacional en una modalidad novedosa para ellos. La capacitación brindada hizo énfasis en el aspecto comunicacional desglosando este en dos áreas:
1- La interacción escrita que el docente deberá tener con sus alumnos ya sea de tipo uno a uno o de uno a muchos. El volumen y la diversidad aumentan en esta modalidad de aprendizaje y es de suma importancia que el docente adquiera nuevas estrategias para manejarlas.
2- El docente deberá integrar equipos de trabajos con especialistas de distintas disciplinas (técnicos y pedagogos), por lo cual aprenderá a interactuar con ellos y a trabajar como un todo apuntando a un objetivo común.

5- BIBLIOGRAFÍA

- Ajzen, I. – (1991) The Theory of Planned Behavior. *Organizational behavior and human decision processes*, 50
- Ajzen, I. & Fishbein, M. (1997) Attitude-Behavior relations: A theoretical analysis and review of empirical research. *Psychological bulletin*, 84
- Ausubel, D. et all, (1987) *Psicología Educativa: Un Punto de Vista Cognitivo*, México: Trillas.
- Bautista Pérez, Guillermo; Borges Sáiz, Federico; Miravalles, Anna Forés i. *Didáctica universitaria en entornos virtuales de enseñanza-aprendizaje*. Madrid: Narcea, 2006
- Brodsky M., (2003) E-learning trends, today and beyond, *Learning Innovations Magazine*, <http://www.elearningmag.com>
- Casanovas, I., Benegas, M., & Fernandez, G. (2008) Marcos pedagógicos para un diseño Blended-learning como herramienta para disminuir la deserción de alumnos de nivel universitario por razones laborales.
- Davis, F.- (1999) Perceived usefulness, Perceived Easy of Use, And User Acceptance of Information Technology. *MIS Quarterly*
- Díaz Barriga Frida, otros, *Estrategias Docentes para un aprendizaje significativo*, Ed. MacGrawHill 1997
- Driscoll M. – (2002) *Blended Learning: Let's Get Beyond the Hype*, IBM Global Services. www-8.ibm.com/services.
- Eklund J.(2004) *Driving the Future of e-learning, Computational and Informational Technologies for Research, Engineering and Education CTMM*
- Gadner, Howard (2006) *Inteligencias Múltiples*. 2da reimpresión, Buenos Aires: Paidós
- Harrison C.- (2001) *Technologies for Learning Spaces*, European Schoolnet Eminent II conference, Portugal
- Haywood J. et al.- (2000) *Learning Technology in Scottish Higher Education*, Vol 8
- Heinze A. et al., *Reflections on the use of Blended Learning*. In *proceedings of Education in a Changing Environment Conference*, Universidad de Salford, 2004
- Johnson D. W., Johnson, R & Holubec, E. (1984). *Circles of Learning: Cooperation in the classroom*. Interaction Book Company, Minnesota.
- Keller C.- (2006) *Technology Acceptance in Academic Organizations: Implementation of Virtual Learning Environments*, 14th European Conference on Information Systems, Gothenburg
- Khan B.- (2002) *Dimensions of E-learning*, *Educational Technology* N° 42
- Kirschner, P. et al., *Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure*

- of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching, *Educational Psychologist* 41(2):75-86; (en línea) <http://www.cogtech.usc.edu/publications>.
- López de Parra L. (en línea) <http://www.oei.es/oeivirt/salacredi/ConferenciaElias.pdf>
- Littleton, K., & Häkkinen, P. (1999). Learning Together: Understanding the Processes of Computer-Based Collaborative Learning. In P. Dillenbourg (Ed.), *Collaborative Learning: Cognitive and Computational Approaches*: 20-30. Oxford: Elsevier Science Ltd.
- Manual de Buenas Prácticas de Alfabetización Tecnológica de Extremadura (en línea) <http://www.nccextremadura.org/descargas/manual/>
- Naidu S.- (2004) Trends in Faculty Use and Perceptions of e-learning
- Pallof, R., & Pratt, K. (2001). *Lessons from the Cyberspace Classroom: The Realities of Online Teaching*. San Francisco: Jossey-Bass.
- Piaget, Jean. (1975) *Introducción a la Epistemología Genética*. Tomo 3. El pensamiento Biológico, Psicológico y Sociológico Ed. Paidós. Argentina.
- Reding V.(2003) Is e-learning going mainstream? *Proceedings of LearnTec Forum*
- Ristola A.- (2005) Understanding consumer perceptions of technology intensive services, (en línea) <http://www.elasm.be/documents>
- Salgado García, E. 2006. Procesos de aprendizaje en e-learning. DGPEP: Seminario 3, Módulo 2. Pag. 27
- Smith H. and Oliver M.- (2002) University teachers' attitudes towards the impact of innovations in information and communication technology on their practice. 9th Improving Student Learning Symposium
- Study of Environments to Support e-learning in UK Further and Higher Education (2005) Joint Information Systems Committee (JISC): Education for Change Ltd, The Research Partnership Social Informatics Research Unit, University of Brighton
- Underwood, J., & Underwood, G. (1999). Task effects on co-operative and collaborative learning with computers. In K. Littleton & P. Light (Eds.), *Learning with Computers: Analysing productive interaction*: 10-23. London: Routledge.
- Venkatesh, V., et al. (2003) User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27
- Weller, M. (2002). *Delivering Learning on the Net: the why, what & how of online education*. Kogan Page.
- Whitelock, D. & Jelfs, A. (2003). Editorial: Journal of Educational Media Special Issue on Blended Learning, *Journal of Educational Media*, 28

PERSONALIZACIÓN EN SISTEMAS DE ENSEÑANZA VIRTUAL

**Elena Durán, Rosanna Costaguta, Roberto Farías, Melina Trejo, Francisco Torales,
Victor Ozán y Pablo Martínez**
Departamento de Informática - Facultad de Ciencias Exactas y Tecnologías (FCEyT)
Universidad Nacional de Santiago del Estero (UNSE)

{ eduran, rosanna } @unse.edu.ar, { fariasrobertoa, melina.trejo, toralesfrancisco } @gmail.com,
{ victorozan, pablo_m8181 } @hotmail.com

CONTEXTO

Este proyecto fue aprobado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Santiago del Estero (SICYT - UNSE), en el marco de la convocatoria 2008 para Proyectos de Investigación Interna de Promoción (PIIP). El mismo se encuentra en ejecución desde el mes de enero del corriente año, y es una continuación de la línea de investigación “Sistemas Adaptativos Inteligentes”, iniciada en 2005 por el proyecto “Herramientas conceptuales, metodológicas y técnicas de la Informática Teórica y Aplicada” (COD. 23/C062), aprobado y financiado por SICYT – UNSE.

RESUMEN

Las nuevas tecnologías aportan a la educación aspectos innovadores que permiten mejorar las formas de enseñar y aprender. Una de las principales innovaciones son los sistemas de educación a distancia, que ofrecen formación continua a estudiantes que por diversas razones no pueden asistir a clases. Como en la enseñanza presencial, en la enseñanza virtual es necesario que los alumnos logren una asimilación efectiva del conocimiento. Sin embargo, frecuentemente tal asimilación no ocurre porque los cursos de e-learning se diseñan sin considerar las características particulares de cada estudiante. Actualmente, los procesos de aprendizaje centrados en el alumno requieren que estos sistemas sean capaces de personalizar la enseñanza a las

características y necesidades individuales de cada estudiante.

En este artículo se presenta el proyecto de investigación denominado “Personalización en Sistemas de Enseñanza Virtual”, cuya finalidad es diseñar, desarrollar, y evaluar módulos para personalizar contenidos en entornos de e-learning. En el trabajo se detallan objetivos, líneas de investigación y desarrollo, resultados esperados, y formación de recursos humanos.

***Palabras clave:** Sistemas de Enseñanza Virtual, Modelo de Estudiante, Personalización, E-learning.*

1. INTRODUCCION

Las nuevas tecnologías han aportado al campo de la educación innovaciones que suponen una mejora cualitativa en las formas de enseñar y aprender. La enseñanza virtual o e-learning es una de las innovaciones más importantes, y tiene como ventaja la capacidad de ofrecer a cada estudiante la libertad de aprender a su propio ritmo, de acuerdo con sus propias capacidades, y sin la necesidad de presenciar las clases. Sin embargo, estas facilidades no aseguran una efectiva apropiación de conocimientos. Un porcentaje considerable de alumnos que actualmente inicia cursos virtuales no los disfruta, pierde motivación, y no logra obtener nuevos conocimientos. El problema radica en que estos cursos no tienen en cuenta las diferentes características y

habilidades que poseen los estudiantes, sus diferentes formas de aprender, de interactuar, en suma, las particularidades que los hacen diferentes aunque tomen el mismo curso.

Atendiendo a esta problemática han surgido los llamados Sistemas de Aprendizaje Adaptativos (SAA) que, a diferencia de los programas tradicionales, muestran la capacidad de adaptarse a cada uno de los alumnos que los usan para aprender (Brusilovsky, 2001). Esta capacidad de adaptación hace que estos programas puedan contribuir en los nuevos procesos de enseñanza; ya que al estar basados en modelos centrados en el alumno, favorecen un aprendizaje significativo y activo. Por esto han tenido gran difusión en el campo de la educación asistida por computación (Kinshuk y Patel, 2001; Gaudioso Vázquez, 2002; Gaudioso y Boticario, 2002; Bull y McEvoy, 2003; Colace y De Santo, 2007).

Existen diferentes alternativas para lograr la capacidad de adaptación o personalización que poseen estos sistemas. En el caso de los Sistemas Tutores Inteligentes (STI), esto se logra mediante la construcción de un Modelo del Alumno, donde se almacena la información que representa el estado actual de un estudiante (Van Lhen *et al.*, 1988). Dicha información se deriva directamente desde los datos que se tengan disponibles: datos previos, respuestas a preguntas que se le vayan formulando, comportamiento que el alumno muestra durante la interacción con el sistema, etc.; o se infieren después de un procesamiento e interpretación de los datos de uso, como por ejemplo, patrones de comportamiento durante el proceso de aprendizaje (Duran y Amandi, 2009; Hartley *et al.*, 1995). El modelo del alumno es utilizado para mantener información actualizada sobre cada estudiante, con el objetivo de proveer una enseñanza adaptada a sus preferencias, estilo de aprendizaje, nivel de conocimiento del dominio, etc. Especialmente, el estilo de aprendizaje es una de las características más relevantes; puesto que diferentes estudiantes aprenden

de distinta manera. Por lo que, sería importante que los sistemas de e-learning se adapten en función del estilo de cada estudiante.

En otros tipos de sistemas de e-learning la adaptación se logra por medio de la tecnología de agentes inteligentes (Johnson, 1998; Beck *et al.*, 2000; Peña *et al.*, 2002; Limoanco and Sison, 2003; Sklar and Richards, 2006; Bica *et al.*, 2006). Los agentes inteligentes son programas de computación que aprenden los intereses, preferencias, y características de los estudiantes; y les brindan asistencia proactiva personalizada.

El proyecto de investigación aquí expuesto se centra en el análisis de la incidencia en el aprendizaje a distancia de las características manifestadas por los estudiantes, tales como: estilo de aprendizaje, roles y habilidades de colaboración, para luego poder diseñar, desarrollar y evaluar módulos que permitan a los sistemas e-learning personalizar sus contenidos y la estrategia de enseñanza a las características y necesidades de cada estudiante.

2. LINEAS DE INVESTIGACION y DESARROLLO

La principal motivación de este proyecto se centra en la necesidad de personalizar los contenidos de los sistemas de enseñanza virtual de acuerdo con las características de los estudiantes (estilos de aprendizaje, nivel de conocimientos previos, edad, estrato social, hábitat, léxico, etc.).

Este proyecto tiene por finalidad diseñar, desarrollar y evaluar, nuevos métodos para la personalización de sistemas de e-learning para mejorar el rendimiento de los estudiantes en este tipo de entornos. Como consecuencia, las preguntas centrales que guiarán su desarrollo son:

- ¿Es factible mejorar el aprendizaje de los estudiantes en entornos de e-learning, por medio de la personalización en

función del estilo de aprendizaje de los estudiantes?

- ¿Qué nivel de incidencia presentan otras características, como los roles y las habilidades de los estudiantes, en el aprendizaje individual y grupal en entornos de e-learning?

Considerando estos interrogantes se originaron dos líneas de investigación para este proyecto. La primera abocada a descubrir la incidencia de los roles y las habilidades manifestadas por los estudiantes en la personalización de sistemas de e-learning. Bajo esta línea se prevé aplicar técnicas de aprendizaje de máquina para descubrir la vinculación entre roles y habilidades de colaboración; y para identificar los roles dominantes en grupos “bien balanceados”. La segunda línea de investigación está dedicada a diseñar, desarrollar y evaluar módulos de personalización considerando los estilos de aprendizaje de los estudiantes aplicados a una plataforma de educación a distancia. Así, el trabajo en esta línea se centrará en la creación de módulos de personalización, basados en la tecnología de agentes inteligentes, y su prueba en algún curso on-line.

3. RESULTADOS OBTENIDOS/ESPERADOS

Con la concreción de este proyecto se espera:

- Alcanzar una mayor comprensión de la incidencia de características como el estilo de aprendizaje, los roles, y las habilidades de los estudiantes, en el aprendizaje en entornos de e-learning.
- Descubrir mediante técnicas de aprendizaje de máquina la incidencia de los roles y las habilidades de colaboración en la personalización de sistemas de e-learning.
- Diseñar, desarrollar y evaluar módulos para personalizar contenidos en entornos

de e-learning sobre la base del estilo de aprendizaje de los estudiantes.

- Realizar propuestas metodológicas para la personalización de sistemas de e-learning.
- Mejorar el nivel de aceptación y el aprendizaje de los estudiantes en los cursos de e-learning.

4. FORMACION DE RECURSOS HUMANOS

Este proyecto contribuirá a la capacitación de los investigadores involucrados, y también a la formación en investigación de estudiantes avanzados de informática, por cuanto los mismos desarrollarán sus Tesis Finales de Grado en el marco de este proyecto. También se espera contribuir a la capacitación de dichos estudiantes en teorías, métodos y técnicas vinculadas con la personalización de sistemas en el ámbito de la educación.

5. BIBLIOGRAFIA

Beck J., Woolf B.P. y Beal C.R. (2000). Advisor: A machine learning architecture for intelligent tutor construction. *Proceedings of the 17th National Conference on Artificial Intelligence*, Austin, TX, USA, Jul-Aug 2000.

Bica F., Verdin R. y Vicari R.M. (2006). Towards Cognitive Modeling of Students' Self-Efficacy. *Proceedings of Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, July 2006, Kerkrade, The Netherlands.

Bull S. y McEvoy A. (2003). An Intelligent Learning Environment with an Open Learner Model for the Desktop PC and Pocket PC, en U. Hoppe, F. Verdejo & J. Kay (eds), *Artificial Intelligence in Education*, IOS Press, Amsterdam, 389-391.

Brusilovsky P. (2001). Adaptive Hypermedia. *User Modeling and User-*

Adapted Interaction, Vol. 11 (1/2), pp. 87-110.

Colace F. y De Santo M. (2007). Adaptive Hypermedia System in Education: A User Model and Tracking Strategy Proposal, *Proceedings of 37th ASEE/IEEE Frontiers in Education Conference*. Session T2D. October 10-13, 2007, Milwaukee, WI, 18-22.

Durán E. y Amandi A. (2009). Personalised collaborative skills for student models. *Interactive Learning Environments*, 1-20, iFirst Article, Taylor & Francis Group.

Gaudioso E. y Boticario J. (2002). Supporting Personalization in virtual communities in distance education, en L.C. Jainy R.J. Howlett (eds.), *Virtual Environments for Teaching and Learning*. World Scientific Publishing Company Pte Ltd.

Hartley R., Paiva A., Self J. (1995). Externaliznig Learning Models. *Proceedings of International Conference on Artificial Intelligence in Education*, en J. Greer (ed.), pp. 509-516, Washington: AACE.

Johnson W. L. (1998). Pedagogical Agents. *Proceedings of International Conference on Computers in Education (ICCE'98)*, China Higher Educational Press and Springer-Verlag, Beijing, 13-22.

Kinshuk B. y Patel A. (2001). A student model for web-based intelligent educational system, en Montgomerie C. & Jarmo Viteli (eds), *Proceedings of EdMedia*, 2001. Norkfolk, VA, USA.

Limoanco T. y Sison R. (2003). Learner Agents as Student Modeling: Design and Analysis. *Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'03)*, Athenas, Greece.

Peña C.I., Marzo J.L. y De La Rosa J.L. (2002). Intelligent Agents in a Teaching Learning Environment on the Web. *Proceedings of the 2nd IEEE International Conference on Advanced Learning Technologies (ICALT'02)*. IEEE.

Sklar E. y Richards D. (2006). The Use of Agent in Human Learning Systems. *Agent-based Systems for Human Learning, AAMAS Workshop*, Mayo 2006, Hakodate, Hokkaido, Japan.

VanLehn K., Polson M. y Richardson J., (eds). (1988). *Student modeling in Foundations of Intelligent Tutoring Systems*, cap. 3, pp. 55-78, Lawrence Erlbaum Associates Publishers.

SISTEMAS DE GESTIÓN DE APRENDIZAJE DE CÓDIGO ABIERTO

Pedro A. WILLGING¹, Norma I. SCAGNOLI², Gustavo J. ASTUDILLO³, Marcela C. CHIARANI⁴, Pablo GARCIA³, Silvia BAST³, Adriana FIORUCCI³, Iván TRUSKALO³

¹ CONICET, Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, UNLPam

² College of Business, Universidad de Illinois en Urbana-Champaign

³ Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, UNLPam

⁴ Departamento de Informática, Facultad de Ciencias Físico-Matemáticas y Naturales, UNSL

RESUMEN

Este proyecto es la continuación de las investigaciones desarrolladas en un proyecto previo que se completó exitosamente y sentó las bases para la búsqueda de resultados más ambiciosos. Habiendo evaluado y seleccionado un sistema de gestión de aprendizaje (SGA) para enseñanza virtual, ahora se va a avanzar en líneas de trabajo que han aparecido como resultado de esta experimentación previa. Tres de esas líneas de trabajo son: (a) *red de cliente delgado con software libre*, (b) *compatibilidad de software de código abierto*, y (c) *objetos de aprendizaje (OA)*.

Se implementará una red de cliente delgado utilizando soporte de código abierto (Linux) reciclando computadoras de bajo rendimiento, para testear el funcionamiento del SGA en una red de esas características. Además del software de código abierto del sistema operativo y del SGA, se hará una búsqueda y testeo de compatibilidad y funcionalidad de otro software de código abierto, en especial para uso educativo.

Los SGA permiten en general la utilización de diferentes módulos para la incorporación de OA. Se investigará la funcionalidad, compatibilidad y flexibilidad de los objetos de aprendizaje en los SGA.

Palabras clave: cliente delgado, código abierto, software libre, objetos de aprendizaje

1. INTRODUCCION

Habiendo completado con éxito un proyecto que permitió a nuestra Universidad contar con una nueva herramienta pedagógica (un sistema de aprendizaje virtual), con la cual docentes, alumnos e investigadores han estado

experimentando y trabajando por algún tiempo, consideramos que es necesario ampliar la investigación y extenderla a (a) la exploración del funcionamiento del sistema de gestión de aprendizaje en una configuración de red de bajo costo; (b) lograr un conocimiento más acabado sobre los objetos de aprendizaje y como se implementan en diferentes sistemas virtuales e (c) investigar la compatibilidad de los sistemas de aprendizaje con software de código abierto externos a tales sistemas.

Código Abierto

Este término *código abierto* (open source en inglés) empezó a utilizarse en 1998 por algunos usuarios de la comunidad del software libre, tratando de usarlo como reemplazo al ambiguo nombre original del software libre (free software). Debido a que este último se asociaba directamente con "software por el que no hay que pagar", y no a la idea de *software que posee ciertas libertades*.

El significado obvio del término "código abierto" es "se puede mirar el código fuente", lo cual es un criterio más débil y flexible que el del software libre; un programa de código abierto puede ser software libre, pero también puede serlo un programa semi-libre o incluso uno completamente de propietario.

El software de código abierto es software para el que su código fuente está disponible públicamente, aunque los términos de licenciamiento específicos varían respecto a lo que se puede hacer con ese código fuente (Wikipedia, 2007).

Afirma Linus Torvalds (2001), "La teoría detrás de código abierto es simple. En el caso de un sistema operativo, el código fuente – las instrucciones de programación subyacentes al

sistema – es libre. Cualquiera puede mejorarlo, cambiarlo, usarlo. Pero esas mejoras, cambios y usos tienen que ser puestos a disposición libremente.”... “El proyecto pertenece a nadie y a todos. Cuando un proyecto de este tipo se presenta, hay un mejoramiento rápido y continuo”.

La metáfora de “la catedral y el bazar”, popularizada por Eric Raymond (1999), subraya dos posturas opuestas en el desarrollo de software. La metáfora del desarrollo de software como construir catedrales evoca un proceso tedioso y lento para crear una estructura meticulosamente, mientras que la metáfora del bazar sugiere un proceso más espontáneo y abierto, de una tribu de programadores conectados construyendo colectivamente sobre el trabajo de otros. Los programas de código abierto son un valor apreciado y difundido en las comunidades de conocimiento compartido (Young, 2001; Pernees, 1999). Cuando el código fuente es abierto, la generación de conocimiento evoluciona rápidamente, porque una comunidad entera trabaja sobre ello. En este sentido, la filosofía de código abierto es una extensión del método científico: la colaboración facilita el descubrimiento y la creatividad (DiBona, Ockman, & Stone, 1999).

Además, el código abierto les da a los usuarios control sobre las tecnologías que usan, en lugar de dejar que el proveedor del sistema le restrinja el acceso al código subyacente a la tecnología. Esto propicia la investigación y facilita la validación de los resultados obtenidos. Además de dar a los docentes una gran oportunidad de probar aplicaciones e implementar proyectos pedagógicos con herramientas disponibles en la Web y con libertad de ser utilizadas y distribuidas entre sus alumnos (Willging *et al.*, 2008).

El modelo de código abierto aplicado a objetos de aprendizaje y SGA tiene tres importantes implicancias. Primero, la base de conocimiento se desarrollará rápida y espontáneamente como lo hace en la ciencia. Segundo, los desarrollos en una institución que son creados de manera individual pueden ser adoptados y modificados para adecuarse al

contexto particular de docentes de otras instituciones. Tercero, el reconocimiento y evolución de una comunidad de contribuyentes de conocimiento y usuarios de material de código abierto provee las bases para criterios de evaluación de pares en excelencia académica.

En la actualidad, ya existe una cantidad considerable de software de código abierto, tanto herramientas administrativas como aplicaciones, lo cual hace necesario efectuar un relevamiento de las potencialidades de estos programas, en especial en lo que a este proyecto de investigación compete, que es la aplicación para fines educativos. Se pretende involucrar no solo a los docentes, sino también a los estudiantes en la búsqueda y análisis crítico del software de código libre disponible, sus características y sus aplicaciones las actividades áulicas.

Sistemas de Gestión de Aprendizaje

El SGA Moodle, es una plataforma de instalación muy sencilla. Debe contar con un conjunto de aplicaciones corriendo de antemano en el servidor. Si bien, puede haber varias configuraciones sobre las que Moodle puede funcionar, un servidor configurado con lo que denomina LAMP (Linux, Apache, MySQL y PHP) es una de las configuraciones más aconsejables (Moodle.org, 2007). En nuestro caso, nos interesa porque tanto Linux, como Apache, MySQL y PHP son aplicaciones de licencia GNU (código abierto). Actualmente, la Facultad de Ciencias Exactas y Naturales cuenta con dos servidores donde se encuentra instalado el SGA Moodle, ambos sobre una plataforma LAMP. Y esta en proceso de armado una red con máquinas con bajos recursos que funcione utilizando la tecnología de clientes livianos.

Red de Cliente Liviano

Una red de clientes livianos es una red basada en servidores donde la mayor parte del procesamiento, o bien todo el procesamiento, se efectúa en el servidor y no en las máquinas cliente. Las aplicaciones residen en el servidor, se ejecutan en el mismo, y sus resultados se visualizan en la máquina cliente.

Una red de clientes livianos típicamente consiste en:

- Uno o más servidores.
- Un conjunto de clientes livianos.
- Un sistema operativo que soporte los clientes livianos.

La tecnología de clientes livianos es factible de ser implementada utilizando escasos recursos de hardware y software de código abierto. Como uno de los objetivos de analizar software de código abierto es lograr introducirlo dentro de la comunidad educativa, sería deseable incorporar el software libre dentro del ámbito escolar a través de clientes livianos. Esto sería especialmente adecuado para las limitaciones de equipamiento a las que, generalmente, se encuentran sometidas las instituciones escolares nacionales.

Objetos de Aprendizaje

Uno de los principales desafíos de la educación mediada por la tecnología es la reutilización y/o readaptación del material utilizado en un curso a otro grupo-clase. Una solución al problema vino de la mano de una idea que no es nueva en el ámbito de la informática, particularmente en programación orientada a objetos, la reutilización. Esta idea fue retomada para el diseño de material educativo digital con el nombre de objetos de aprendizaje (OA). Los OA tienen su origen en la educación a distancia (EaD), y más precisamente como parte del denominado “diseño instruccional” (Navarro Cendejas & Ramírez Anaya, 2005).

Veamos algunas definiciones del término objeto de aprendizaje. Navarro Cendejas & Ramírez Anaya (2005) definen los objetos de aprendizaje como herramientas educativas de diseño instruccional basadas en tecnología de la información, que pueden ser utilizadas en programas curriculares y metodologías de diverso orden.

David Wiley (2000) habla de los objetos de aprendizaje como componentes instruccionales digitales, distribuidas a través de Internet, que pueden ser reutilizadas, varias veces, en diferentes contextos de aprendizaje.

En enciclopedia digital, Wikipedia nos encontramos con que un objetos de aprendizaje es una entidad informativa digital desarrollada para la generación de conocimiento, habilidades y actitudes que

tiene sentido en función de las necesidades del sujeto y que se corresponde con la realidad.

Según la comunidad APROA (Aprendiendo con Objetos de Aprendizaje: un repositorio de objetos de aprendizaje) un objeto de aprendizaje corresponde a la mínima estructura independiente que contiene un objetivo, una actividad de aprendizaje, un metadato y un mecanismo de evaluación, el cual puede ser desarrollado con tecnologías de infocomunicación (TIC) de manera de posibilitar su reutilización, interoperabilidad, accesibilidad y duración en el tiempo.

Las definiciones presentadas nos pueden dar una idea de la función que se desea que tengan los OA, pero también nos hablan (en la mayoría) de la tecnología informática, por tanto es necesario que existan implementaciones concretas (informáticas) de estas entidades. De las definiciones también se desprende el carácter de reusabilidad que requiere un objeto. Para que un OA pueda ser reusable es necesario que esté accesible y que pueda ser utilizado en una variedad de plataformas. Como un aporte a esto último, ADL (Advanced Distributed Learning), crea SCORM (del inglés Sharable Content Object Reference Model), una especificación que permite crear OA que puedan ser utilizados en aquellos SGA que se apeguen al estándar (2006). Otro estándar que permite, a aquellos que diseñan contenidos para e-learning, compartir materiales instruccionales es IMS-Learning Design (Instructional Management System), una especificación que intenta dar un marco de referencia para la distribución de OA que pueden ser utilizados en educación a distancia o presencial (Berggren et al., 2005).

Existe, como era de esperar por estos días, una gran cantidad de herramientas de autor que permiten la creación de objetos de aprendizaje y la mayoría cumplen con el estándar SCORM. Dentro de estas herramientas, podemos nombrar a: Reload, eXe, o CourseLab como algunas de código abierto.

Otra forma de trabajar con OA (no apegada a SCORM) es utilizar LAMS (Learning Activity Management System). LAMS es una aplicación, distribuida como software libre por la Fundación LAMS de la Universidad

australiana de Macquarie. Esta institución, define LAMS como una revolucionaria herramienta para el diseño, administración y distribución de actividades colaborativas en línea.

Las herramientas descritas hasta aquí son de importancia para este grupo de investigación por dos motivos: son de código abierto y pueden ser incluidas como recurso didáctico en el SGA Moodle (también de código abierto).

2. LINEAS DE INVESTIGACION y DESARROLLO

Este proyecto apunta a profundizar la investigación iniciada, siguiendo varias líneas de trabajo que han surgido y que prometen aprovechar aún más las potencialidades de la red como herramienta para enseñanza y aprendizaje. El efecto de difusión que se pretende lograr, con la experimentación, puesta en marcha y desarrollo de este trabajo, es el de mejorar las actuales condiciones para hacer posible que la UNLPam incorpore de manera efectiva la educación digital/virtual como parte de sus recursos tanto para docentes como alumnos.

Algunas de las preguntas que se abordarán e intentarán responder son:

1. ¿Qué software de código abierto existe que pueden ser adoptado por la UNLPam como complemento de los SGA y para producir diferentes elementos para la integración de tecnología en el aula?
2. ¿Cómo funciona el SGA de código abierto en una red de cliente delgado?
3. ¿Cuán compatibles son los SGA y los sistemas de software investigados?
4. ¿Cómo puede la enseñanza en la UNLPam beneficiarse con el uso de objetos de aprendizaje?

El objetivo general del proyecto es el desarrollo de un sistema de enseñanza en línea (que constituya un recurso de importancia pedagógica) propio para la UNLPam. Los objetivos puntuales son:

- Explorar el uso de diferentes tipos de software de código abierto que pueden ser usados como complemento del SGA.
- Exponer a los estudiantes avanzados de los profesorado a los procesos de análisis de

sistemas de aprendizaje y software de código abierto.

- Testear el software de código abierto y la plataforma de educación virtual en redes de cliente delgado y bajo costo.
- Agregar un recurso pedagógico más al repertorio disponible para los docentes de la UNLPam.
- Facilitar la producción de contenidos multimediales mediante el uso de software de código abierto.
- Poner al alcance de los docentes materiales para la producción y enseñanza que se adapten a las condiciones y recursos de la UNLPam, que los motive a ampliar sus técnicas de aprendizaje, y que los incentive a crear nuevas aplicaciones que beneficien a los alumnos y a la institución.
- Mejorar la oferta educativa de la UNLPam proveyendo la posibilidad de cursos en línea para la comunidad.

La hipótesis de trabajo es que los sistemas para el aprendizaje a través de Internet (de código abierto) han madurado lo suficiente como para que su adaptación a las necesidades de la UNLPam sea posible a través del trabajo de desarrollo de un equipo de investigadores/docentes. Que existe disponibilidad de una gran variedad de software libre de aplicación educativa. Que las redes de cliente delgado podrían soportar de manera económica las aplicaciones de software de código abierto. Y que el desarrollo y posterior incorporación de estas herramientas al repertorio de recursos disponibles constituye un aporte valioso para la enseñanza e investigación.

3. RESULTADOS ESPERADOS

- La exploración de sistemas de enseñanza en línea generará una actualización en las áreas de e-learning y aprendizaje a través de tecnologías de la comunicación. Esto se espera tenga un efecto sinérgico no solo entre los integrantes del equipo de investigación, sino también, y tal vez más importante, en el resto de la comunidad educativa de la UNLPam.
- Continuación en el fortalecimiento de una nueva área de investigación.

- Formación de RRHH: se capacitarán recursos humanos en la aplicación de un modelo pedagógico basado en el uso de las Tecnologías de la Información y Comunicación (TIC)
- Experiencia con sistemas / plataformas para aprendizaje en línea
- Establecer las condiciones para que la UNLPam pueda ser un proveedor de instrucción en línea
- Producción de informes de investigación y publicaciones que sirvan para diseminar la experiencia/conocimiento adquirido.
- Posibilidad de implementación de este formato para EaD, suplantando o complementando los formatos que se usan actualmente, abaratando costos y ampliando su alcance.

5. BIBLIOGRAFIA

- ADL** (2006). Overview - SCORM® 2004 - Sharable Content Object Reference Model - 3rd EDITION. Fecha de consulta: marzo 18, 2007 desde www.adl.gov/scorm/20043ed/index.aspx
- Código abierto.** (2007). Wikipedia, La enciclopedia libre. Fecha de consulta: marzo 17, 2007 desde http://es.wikipedia.org/wiki/Código_abierto.
- DiBona, C., Ockman, S., & Stone, M.** (1999). Introduction. In C. DiBona, S. Ockman & M. Stone (Eds.), *Open sources: Voices from the open source revolution* (pp. 1-17). London, O'Reilly.
- Navarro Cendejas J., Ramirez Anaya L.** (2005). *Objetos de aprendizaje. Formación de autores con el modelo de redes de objetos.* México: UEDG Virtual.
- Raymond, E.** (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an accidental revolutionary.* Sebastopol, CA: O'Reilly.
- Torvalds, L., & Diamond, D.** (2001). *Just for fun: The story of an accidental revolutionary.* New York, HarperCollins.
- Wiley, D.** (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In D. A. Wiley (Ed.), *The Instructional Use of Learning Objects: Online Version.*

Consultado 27 de Agosto de 2007, de: <http://reusability.org/read/chapters/wiley.doc>

Willging, P., Astudillo, G. (2008). Software Libre para Matemática: en Búsqueda de Alternativas. Memorias de la II REPEM, (p. 135-144).

Young, J. R. (2001, May 4). Universities begin creating a free, 'open source' course-management system. *The Chronicle of Higher Education*, p. A36.

Recursos en la Web

APROA Comunidad - ¿Qué es un Objeto de Aprendizaje?.

<http://www.aproa.cl/1116/propertyvalue-5538.html>

CourseLab <http://www.courselab.com>.

eLearning XHTML editor. <http://exelearning.org/>.

eXe Authoring project. Online manual. What is eXe?. http://www.wikieducator.org/Online_manual.

Foxton L., Page J. AMS Documents. Unix Installer Help. <http://wiki.lamsfoundation.org/display/lamsdocs/Unix+Installer+Help>.

LAMS v1.0.2 Integration Setup Step-by-Step Guide. <http://lamsfoundation.org>.

RELOAD. Reusable eLearning Object Authoring & Deliberation. Project Summary. <http://www.reload.ac.uk>.

Capacitando Comunidades Marginales a través de un Medio Masivo de Comunicación

Universidad Nacional de La Matanza
Departamento de Ingeniería e Investigaciones Tecnológicas.
Florencio Varela 1903, San Justo, Buenos Aires, Argentina

Daniel A. Giulianelli
dgiulian@unlam.edu.ar

Graciela S. Cruzado
graciela@unlam.edu.ar

Rocío A. Rodríguez
rrodri@unlam.edu.ar

Edgardo J. Moreno
ej_moreno@unlam.edu.ar

Marcelo Garay
mgaray@unlam.edu.ar

Clara L. Rojas
crojas@unlam.edu.ar

CONTEXTO

Este artículo se basa en el trabajo realizado por el equipo de investigación de la Universidad Nacional de la Matanza que trabaja en el análisis de mecanismos para difundir TICs (Tecnologías de la Información y Comunicación) a las sociedades marginales a fin de insertar a las mismas en la sociedad del conocimiento y por consiguiente disminuir la brecha tecnológica existente entre las comunidades.

Palabras clave: Medios de Comunicación, TICs, Brecha Tecnológica, Educación a Distancia

RESUMEN

Se toma como muestra al partido de La Matanza situado en la provincia de Buenos Aires, el cual cuenta con más de 1.500.000 habitantes. Esta población que supera en número a la de 19 provincias tomadas separadamente, es solo inferior a la Ciudad Autónoma de la Provincia de Buenos Aires y las provincias de Buenos Aires, Córdoba, Santa Fe y Mendoza. Dentro de este Partido conviven comunidades residenciales junto a marginales, por ello se hace un estudio a fin de cuantificar la brecha tecnológica y luego aplicar una metodología que permita acercar el conocimiento a las zonas más marginales a fin de disminuir dicha brecha e incluir a las mismas en la sociedad del conocimiento.

1. INTRODUCCION

A fin de determinar cuán grande es la brecha tecnológica se realiza un extenso estudio en un

sector de la sociedad Argentina. Para ello el equipo de investigación realiza una encuesta entre la población de la provincia de Buenos Aires, más precisamente en el partido de La Matanza. Se elige éste partido por varios motivos:

1) A fin conocer las necesidades reales de la población que dio origen a la universidad pública (UNLaM) de la cual formamos parte.
2) La convivencia de barrios marginales y residenciales ha permitido determinar las cotas que permitieron establecer la magnitud de la brecha tecnológica.

3) El partido con una superficie de 323 Kilómetros cuadrados (Kms²) y una densidad poblacional superior a los 4.644 habitantes por Km² está formado por 15 localidades. Dichas localidades pueden clasificarse en tres cordones poblacionales, cada cordón contiene comunidades con características socio-culturales distintas. Cabe destacar que el nivel socio económico va decreciendo desde el primer cordón hacia el tercero. El tercer cordón posee las localidades más marginales.

A fin de relevar todas las localidades, se realizó un formulario de encuesta, el cual fue distribuido entre las escuelas medias que colaboraron en el proyecto a fin que las mismas lo distribuyan. Se le dio a cada escuela precisos rangos de edades y sexos, para que la muestra sea representativa. Por otra parte en las localidades del tercer cordón, pasantes de la universidad de las carreras del Departamento de Humanidades, se acercaron a los barrios más humildes a fin de encuestar a quienes estuvieran transitando por la zona. Se han definido dos universos y dos medios

distintos de relevar información a efecto de minimizar errores y cubrir la totalidad de localidades del partido:

a) Con una muestra de 4 habitantes cada 10.000 pobladores del partido, (0,04%) se llegó a través de las escuelas medias participantes.

b) Para el caso puntual de las localidades del tercer cordón poblacional y a efecto de disminuir el posible error, se materializó un trabajo de campo que alcanzó a 14 habitantes por cada 10.000 pobladores, (lo que representa un 0,14 %).

A fin de dar confiabilidad a la muestra, en el caso de las encuestas realizadas por pasantes se rechazaban aquellas en las que se podía inferir que no contestaba con veracidad el encuestado y se procedió a realizar dicha encuesta a otra persona del mismo sexo y rango de edad para su reemplazo. El total de encuestas obtenidas tanto las realizadas por pasantes en el tercer cordón como las obtenidas de las escuelas, fueron sometidas a un análisis a fin de corroborar la veracidad de las mismas. Esto se pudo realizar ya que, en la encuesta, a priori se incluyeron para este fin preguntas autochequeables que permitieron detectar respuestas contradictorias. Se obtuvieron en total 1029 formularios de encuesta válidos. El formulario de encuesta permite realizar muchas comparaciones entre los cordones, las cuales contemplan cuestiones de nivel de conocimiento general, en informática, cuestiones económicas, etc. a fin de poder comprender la situación en la que se halla inmersa cada comunidad. Con el objeto de citar los resultados más significativos se muestra en este ítem lo relacionado con capacitación y tecnología. Para la medición de la brecha tecnológica se seleccionaron algunos indicadores de distintas categorías. A saber:

I) Nivel de conocimiento en informática: Es posible destacar que en los tres cordones, gran parte de la población declara no poseer ningún conocimiento en Informática, creciendo las cifras a medida que las distancias se incrementan, desde el primer al tercer cordón. En la figura 1, se muestra el porcentaje de población sin conocimiento en informática. Añadiendo información a lo presentado en la

figura 1, el mayor porcentaje de pobladores que declaran tener un nivel excelente de conocimientos en informática, se presenta en el primer cordón siendo tan solo del 3%.

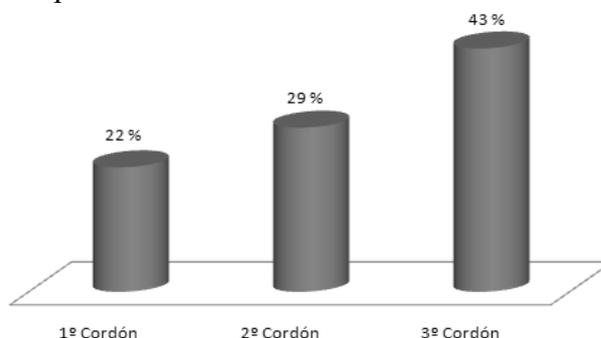


Figura 1. Porcentaje de población sin conocimiento en Informática.

II) Le interesa aprender informática: En aquellos casos en los cuales el encuestado declaraba no tener ningún tipo de conocimiento de informática, se les preguntó si les interesaba aprender. En los tres cordones los porcentajes arrojados superan el 50%. En el caso particular del tercer cordón el 75% declara su interés en aprender.

III) Conocimiento del paquete Office: Solo contestan cual es el nivel de conocimiento del paquete office aquellos que han declarado tener conocimientos en informática. En los tres cordones el nivel de conocimiento declarado es pobre. El tercer cordón poblacional, es el que presenta menor conocimiento en los programas del paquete office. En la figura 2 se muestra para cada cordón el porcentaje que declara no tener ningún conocimiento o un conocimiento escaso del paquete office.

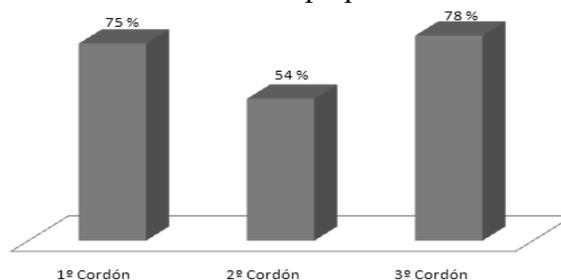


Figura 2. Porcentaje de población, con muy escasos o sin conocimientos del paquete office.

Los resultados obtenidos (publicados por el equipo de investigación en [4]) demuestran que: mientras algunas sociedades hacen uso de las TICs como parte de su vida cotidiana otras están tecnológicamente excluidas ya sea por falta de recursos y/o conocimientos. Esta

brecha tecnológica [1][2] se traduce en desigualdad de oportunidades, no permitiendo en determinados casos alcanzar ciertos puestos de trabajo e incluso mejorar su calidad de vida a través del uso de las TICs. Se mencionan a continuación tan solo cuatro de las múltiples medidas que se realizan para intentar minimizar la exclusión de diversas comunidades:

- Trabajar con autoridades educativas del partido a fin de igualar los planes de estudio de los colegios secundarios y primarios a fin que las comunidades puedan alcanzar el mismo grado de conocimiento.
- Realizar conferencias sobre temas tecnológicos en los barrios más alejados de los centros de educación superior. En estos barrios, sus habitantes, generalmente carecen de recursos como para trasladarse a los lugares donde se realizan habitualmente las mismas.
- Capacitar gratuitamente en temas vinculados con TICs, con la colaboración de los institutos educativos de nivel secundario, a los habitantes de las comunidades más alejadas, a través de los medios de comunicación masiva. Complementar esta capacitación con prácticas en laboratorios de informática de los Centros de Educación Superior (CES).
- Propiciar a través de los órganos de gobierno del partido el aumento de la inversión estatal en equipamiento informático y comunicaciones en colegios del tercer cordón a fin de retener a la población en su hábitat.

Respecto a la tercera y segunda medidas propuestas, la Universidad Nacional de La Matanza se suma a estos esfuerzos ofreciendo capacitar gratuitamente en computación a la población más marginal (tercer cordón poblacional). También se prevé, invitar a alumnos de colegios de zonas marginales al campus universitario, para participar de videoconferencias a través de la utilización de Redes de Altas Capacidades. En estas actividades, distintos organismos que cuentan con el recurso, muestran los distintos usos que le dan a la tecnología, (educación interactiva,

realización de intervenciones quirúrgicas a distancia, operación remota de telescopios, manipulación remota de instrumentos, etc.) a fin de incentivar a los participantes en temas vinculados con las TICs.

2. LINEAS DE INVESTIGACION y DESARROLLO

Consientes de la existencia de la brecha tecnológica interna y de su implicancia, se decide capacitar a las comunidades más marginales. El equipo de investigación ha estudiado diversas alternativas para acercar el conocimiento a dichas comunidades por medio de una capacitación gratuita. Si bien el interés por la propuesta ha sido masivo, han surgido algunos inconvenientes:

- Debido a la falta de recursos no hay posibilidad que los interesados se movilicen hacia la universidad a fin de capacitarse.
- No es posible disponer de equipos informáticos en las zonas marginales dado que las muchas instituciones educativas no cuentan con ellos, algunas ni siquiera tienen electricidad. Por otra parte en zonas aledañas a estos barrios por cuestiones de seguridad quienes disponen de equipamiento no están dispuestos a conceder el uso de los mismos a ciudadanos de otras comunidades.

Debido a esto la Universidad pone a disposición sus laboratorios para realizar clases prácticas y se ocupa del traslado de las personas. A medida que se avanza con la propuesta se observa que es difícil encontrar un horario en el cual la mayor parte de los interesados pueda trasladarse por lo que se decide realizar un curso semipresencial. La propuesta implica clases presenciales para la realización de prácticas en laboratorio y clases a distancia para proporcionar los conocimientos necesarios previos a las prácticas. Se elige como medio masivo de comunicación la radio y a través de la emisora de la universidad se acuerda la realización de microprogramas radiales de 20 minutos de duración dos veces por semana, con repeticiones en distintas franjas horarias. Cuando los inconvenientes de equipamiento y movilidad estaban solucionados aparece un nuevo inconveniente: el alcance de la radio.

Cabe destacar que en algunas zonas en una región de cuatro manzanas se sintonizaba perfectamente la emisora universitaria y en la manzana aledaña era tapada por radios locales. Se seleccionan dos colegios secundarios de zonas marginales explicando que la capacitación esta planificada para la comunidad y no para alumnos del colegio. Proponiendo como edad mínima 22 años, los directivos realizan una posible lista de candidatos y reparten los formularios de inscripción a vecinos del establecimiento, familias de los alumnos, etcétera. Seleccionando a la población con bajos recursos e imposibilitada de realizar una capacitación no gratuita. En el formulario de inscripción se realizan preguntas a fin de analizar el grado de conocimiento en informática y además si receptionan correctamente la señal de la emisora. De todos los que han completado el formulario de inscripción se descarta a aquellos que ya tienen conocimientos en informática ó quienes no receptionan la señal de la emisora. Muchos interesados aclaran que no la reciben pero que pueden movilizarse a casas de vecinos o familiares en donde han verificado que se receptiona correctamente. Las autoridades de los colegios ofrecen un aula con una radio para que quienes no pudieran sintonizar la emisora, escuchen allí el programa. Tres miembros del equipo de investigación se capacitan para el dictado de clases radiales. En las clases a distancia a diferencia de las presenciales el docente no puede observar la expresión de los alumnos, no puede hacer una pregunta para ver en que tiempo responden los alumnos, en el caso que lo hagan [3]. Por otro lado, hablar de una computadora por medio de la radio a personas que nunca la han visto y/o usado dificulta la tarea. Por ello se realizan láminas ilustrativas que se entregan a los inscriptos a través de colegios con la finalidad de visualizarlas mientras que se explican los contenidos (a modo de ejemplo la figura 3 presenta una lámina).

Los microprogramas se desarrollan dos veces por semana durante dos meses, cada uno de los ocho microprogramas se realiza en vivo con la siguiente estructura:

- 1) Breve repaso de la clase anterior.
- 2) Título y desarrollo de la clase actual.
- 3) Conclusiones de la presente clase.

La clase se desarrolla en un ámbito ameno en donde el docente establece ciertos diálogos con el conductor del programa quien hace más interactiva la explicación. Se menciona un número telefónico al cual los oyentes pueden llamar y otro número al cual es posible enviar mensajes de texto. Todas las consultas que llegan en el transcurso del programa se contestan en vivo. Se desarrollan tres módulos: Hardware, Software e Internet. Los cuales además de tener sus contenidos teóricos impartidos por medio de la radio, son complementados con cuatro clases prácticas de tres horas de duración cada una. A lo que se le adiciona un examen final teórico-práctico.

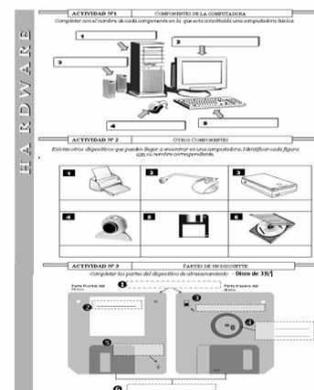


Figura 3. Vista general de una de las láminas

3. RESULTADOS OBTENIDOS/ESPERADOS

Se ha decidido acotar esta primer experiencia a 84 inscriptos. A fin de reducir la cantidad de alumnos por práctica y poder realizar un mejor seguimiento por alumno se dividieron a los inscriptos en tres grupos. La asignación de los inscriptos a los grupos fue basada exclusivamente en las posibilidades horarias de los mismos para la asistencia a las clases prácticas (quedando tres grupos de 34, 27 y 22 alumnos). En la figura 4, se muestra los rangos de edades de los inscriptos.

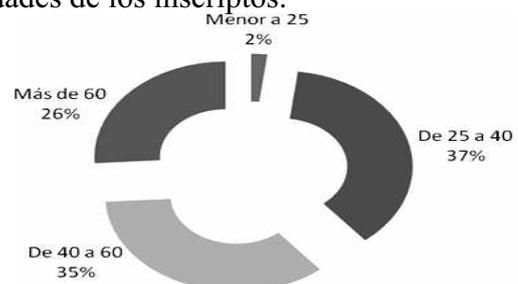


Figura 4. Rangos de edades

Como resultado, del plan de capacitación bajo la metodología propuesta por el equipo de investigación, se detalla lo siguiente:

- Aprobados: Cumplieron con el 75% de asistencia a las prácticas y aprobaron el examen final teórico-práctico.
 - En primera instancia.
 - A través de recuperatorio.
- Asistieron: cumplieron con el 75% de asistencia a las prácticas pero no se presentaron a rendir el examen final.
- Desaprobaron: “Asistieron” pero desaprobaron el examen final
- Ausentes: Su asistencia a prácticas no alcanzo al 75%. Este porcentaje indica el nivel de deserción al curso. En la figura 5, se muestran los porcentajes de: Aprobados (en primera instancia y con recuperatorio), Asistentes, Desaprobados y Ausentes.

De la población que participó de la experiencia, sólo el 16% tenía acceso a una computadora ya sea en su casa, en la de un familiar, vecino ó en su lugar de trabajo. Dado que las prácticas eran una vez por semana en los laboratorios de la universidad, se les aconseja que practiquen (cybercafes, locutorios, etc.) y en su mayoría traían dudas a la siguiente clase práctica de aquello que habían intentado realizar y no lo habían conseguido. Esto hizo que el curso tenga éxito y que el examen final teórico-práctico haya sido aprobado por el 66% (ver figura 5). La deserción ha sido baja, solo el 8% no ha concluido el curso, en algún caso el motivo ha sido no poder asistir en los horarios de práctica. El examen ha sido desaprobado por el 7% quienes no obstante han completado la encuesta final luego de conocer la nota del examen y han escrito muy buenos comentarios sobre esta experiencia. La brecha tecnológica que separa a las distintas comunidades del partido de La Matanza es grande pero este proyecto demuestra que se pueden tomar acciones que permitan disminuir la misma. En razón de la importancia de las TICs en la sociedad, se intento a través del proyecto encontrar un canal a través del cual brindar a los habitantes de la comunidad una herramienta que les posibilite perder el miedo y posibilite a cada quien decidir su mejor uso.

Queda demostrado que es posible implementar mecanismos que conlleven a la reducción de la brecha tecnológica interna siendo los medios masivos de comunicación un muy buen canal para poder difundir conocimiento.

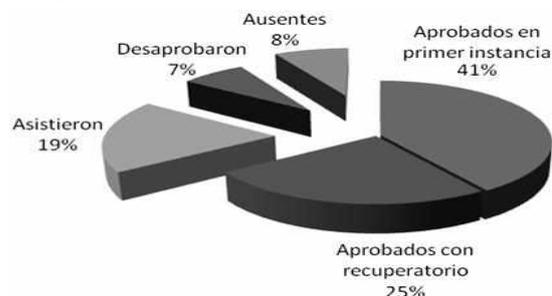


Figura 5. Porcentajes

4. FORMACION DE RECURSOS HUMANOS

El equipo de investigación está compuesto por docentes investigadores, tres alumnos becados e incluyó en la etapa de encuestas y posterior análisis a seis alumnos pasantes. Durante el proyecto se ha mantenido contacto con autoridades escolares definiendo los colegios que formaron parte de la propuesta. Por otra parte es necesario destacar la integración del equipo de Radio Universidad (FM 89.1), que gracias a su apoyo continuo permitió a integrantes del equipo de investigación realizar una tarea prácticamente profesional.

5. BIBLIOGRAFIA

- [1] A. Serrano Santoyo y E. Matinez Martinez “La brecha digital: Mitos y Realidades” Universidad Autónoma de Baja California, México, 2003. pp. 4 -10. ISBN: 978-970-9051-89-6
- [2] I. Munster. “La brecha en Latinoamérica: un caso de estudio” World Library and Information Congress: 70th IFLA General Conference and Council, 2004
- [3] Ramírez Ramírez. “La Educación a Distancia como instrumento de lucha contra la pobreza y de fortalecimiento democrático en América Latina”. Institución: Universidad Estatal a Distancia de Costa Rica.
- [4] A. Vázquez, D. Giulianelli, O. Dominguez Soler, M. Pérez Guntín, G. Blanco, G. Cruzado, R. Rodríguez, C. Farkas, C. Tenisi y Marcelo Garay. Comunicando Comunidades: Redes Informáticas y el Partido de La Matanza. Universidad Nacional de La Matanza. 2008. ISBN 978-987-9495-69-8

SIMULACIÓN Y MÉTODOS NUMÉRICOS EN CIENCIAS DE LA COMPUTACIÓN

Adair Martins Claudia C.Fracchia Claudia Allan
amartins@uncoma.edu.ar cfracchi@uncoma.edu.ar allanclau@gmail.com

Departamento Ciencias de la Computación
Facultad de Economía y Administración
Universidad Nacional del Comahue

CONTEXTO

En el año 2008 se presentó el proyecto “Simulación y Métodos Numéricos en Ciencias de la Computación”. El mismo integra el programa “Simulación y Otros Métodos Computacionales en Ciencias y Educación”. Es una colaboración entre el Departamento de Ciencias de la Computación de la Facultad de Economía y Administración y el Departamento de Electrotecnia de la Facultad de Ingeniería, de la Universidad Nacional del Comahue (UNCo). Participan también integrantes de los Departamentos de Estadística y de Matemática de la UNCo, además de las Universidades Brasileñas del Estado de São Paulo (UNESP), de Brasilia (UCB) y de Itajubá (UNIFEI).

RESUMEN

Se presenta el proyecto “Simulación y Métodos Numéricos en Ciencias de la Computación”, perteneciente a la Facultad de Economía y Administración de la UNCo. Está siendo llevado adelante por un grupo de trabajo interdisciplinario, lo que enriquecerá y potenciará los objetivos que se proponen realizar. Se abarcan tres líneas de investigación: Simulación de Modelos, Algoritmos Numéricos y Uso de Tecnologías de la Información y Comunicación (TICs). Mediante la técnica de modelado y simulación de sistemas se pretende trabajar en la creación de ambientes virtuales, tanto en el medio académico como en el empresarial. Profundizar en los fundamentos y aplicación de la simulación discreta a redes de

computadoras, confiabilidad de sistemas, etc., como así también en el estudio y aplicación de las técnicas de simulación continua utilizando la metodología de Bond Graphs. También se desarrollarán e implementarán algoritmos numéricos basados en los métodos iterativos, y las variantes que no requieren derivada utilizando el concepto novedoso de deflación de funciones para obtención las raíces de funciones no lineales y su multiplicidad. Mediante el potencial que ofrecen las TICs se continuará diseñando herramientas para la evaluación del aprendizaje y materiales educativos, que permitan fortalecer los procesos de enseñanza-aprendizaje mejorando la calidad educativa.

Palabras clave:

Simulación, Bond Graphs, Algoritmos numéricos, TICs

1. INTRODUCCION

El crecimiento formidable de la utilización de herramientas computacionales se debe a la facilidad de uso y sofisticación de los ambientes de desarrollo de modelos, sumada al creciente poder de procesamiento de las computadoras personales y a las interfaces gráficas cada vez más amigables. Dentro de esta temática se abarcan tres subtemas que formarán parte de tesis de maestría y de grado: Simulación de modelos discretos y continuos, Generalización del concepto de deflación para resolución de funciones no lineales y Estudio y aplicación de TICs en educación.

Simulación de modelos discretos y continuos.

En el ejercicio de las funciones típicas de cualquier actividad humana, el hombre tiene que tomar decisiones de uno y otro tipo en forma permanente. Esta situación trae consigo, por lo general, riesgo e incertidumbre, lo que compromete la calidad y el logro de la decisión; esto lo dificulta el hecho de que el responsable debe enfrentar la presión que implica la alta responsabilidad involucrada y, en ciertos casos, su inexperiencia o incompetencia. Para contrarrestar esta situación, el hombre ha desarrollado a través del tiempo una diversidad de herramientas que permiten minimizar el riesgo y la incertidumbre en la toma de decisiones.

La simulación es una de estas herramientas; con su aplicación no sólo se logra el anterior cometido, sino que se minimizan los costos involucrados en la decisión mediante un mejor uso de los recursos, la disminución del tiempo utilizado y la minimización de las probabilidades de riesgo. A través del proceso de diseño de un modelo de un sistema real, y dirigiendo el experimento con él, se puede entender el comportamiento del sistema, lo cual permite tomar mejores decisiones.

La simulación de Monte Carlo es una técnica que combina conceptos estadísticos (muestreo aleatorio) con la capacidad que tienen las computadoras personales para generar números pseudoaleatorios y automatizar cálculos. Así, en la actualidad es posible encontrar modelos que hacen uso de simulación de Monte Carlo en las áreas de informática, ingeniería, empresarial, económica, industrial e incluso social. Existen en el mercado varios programas de aplicación a la simulación de sistemas discretos, entre ellos puede mencionarse: ARENA, PROMODEL, SIMUL8, AUTOMOD, SIMPROCESS, etc. Inclusive, las últimas versiones de planillas de cálculo, particularmente Excel, pueden incorporar complementos para realizar simulación de Monte Carlo, siendo los más conocidos: @Risk y Crystall Ball. Las últimas versiones de algunos de estos programas han comenzado a incorporar la posibilidad de combinar la

simulación discreta con la continua en forma combinada [1-8].

Un método de modelado y simulación continua y que viene siendo desarrollado y utilizado en forma creciente en los últimos años es la técnica denominada en inglés Bond Graphs. Esta técnica de modelado y simulación de carácter universal se basa en el concepto de analogías entre sistemas de distinta naturaleza física con los sistemas eléctricos. Proporciona una visualización gráfica de las interacciones entre los distintos componentes del sistema y suministra implícita y sistemáticamente el modelado matemático del sistema en forma de variables de estado, facilitando el uso de recursos computacionales para la simulación, que se realiza directamente de esta representación gráfica [9-11]. Existe actualmente un gran número de programas comerciales disponibles en el mercado que permiten usar este método. Se destacan 20-SIM, SYMBOLS2000, entre otros.

Generalización del concepto de deflación para resolución de funciones no lineales.

En ciencias e ingeniería muchos problemas prácticos involucran la resolución de funciones no lineales que no son polinómicas mediante métodos numéricos. De los cursos de Álgebra de los primeros años de las carreras de ciencias e ingenierías es bien conocido el concepto de deflación de polinomios, la obtención de una nueva raíz permite disminuir el grado del polinomio simplificando progresivamente la obtención de las restantes. El conocimiento de los métodos numéricos iterativos para la obtención de raíces de ecuaciones no lineales, particularmente el método de Newton Raphson, permiten generalizar el concepto de deflación en la obtención computacional de todas las raíces de una ecuación no lineal [12]. Se pretende investigar la factibilidad de estimar numéricamente el orden de convergencia, consecuentemente la multiplicidad de una raíz, simultáneamente con el cálculo de la misma, para mitigar el efecto de las discontinuidades y optimizar el algoritmo.

Estudio y Aplicaciones de TICs en Educación

El problema de la enseñanza y aprendizaje de las Ciencias de la Computación no puede despegarse del empleo de las TICs. Estas nuevas tecnologías suponen, para la sociedad moderna, cambios profundos e inevitables. La tarea del docente requiere el diseño de metodologías, contenidos y actividades de aprendizaje, para que los estudiantes puedan desarrollar valores, habilidades y conocimientos significativos en una sociedad tecnológica. Esto implica la creación de escenarios educativos que permitan seleccionar información, administrar, procesar, modelizar, simular, y acceder a fuentes de información, para las cuales es necesario poner en marcha procesos que permitan la construcción de saberes.

También se posibilita definir nuevos marcos de interacción: relaciones a distancia (aulas virtuales), interacción continua y activa entre los estudiantes y sus tutores (redes interactivas), consideración del aprendizaje como la herramienta para la selección y organización de la ilimitada información disponible (recursos multimedia abiertos y flexibles), etc.

Las herramientas tecnológicas permiten complementar la educación presencial y superar las barreras espacio-tiempo al introducir las modalidades de educación semipresencial y a distancia. También facilitan la obtención de información sobre desempeños, usos de materiales, interacción entre los distintos participantes (alumnos, grupo de alumnos, profesores), que sirve para retroalimentar y modificar los procesos de enseñanza-aprendizaje empleados.

Se pretende continuar profundizando la línea de investigación de diseño de recursos tecnológicos, ya sean materiales educativos o herramientas computacionales, que sirvan para potenciar y fortalecer los procesos educativos, que estos permitan a los alumnos construir aprendizajes significativos, a los docentes realizar evaluaciones continuas de sus alumnos, diseñar material que resulte motivador, todo en busca de optimizar la calidad de la educación [14-19].

2. LINEAS DE INVESTIGACION y DESARROLLO

Se desarrollarán tres líneas de investigación:

Simulación de modelos: Se profundizará en el estudio de los distintos componentes para el modelado y simulación discreta y continua de sistemas. Se realizarán aplicaciones avanzadas a problemas de distintos campos del conocimiento. También se efectuarán relevamientos de los distintos softwares existentes para determinar cuales son los más adecuados para uso académico y profesional.

Algoritmos numéricos: Se desarrollarán e implementarán algoritmos interactivos para la obtención de raíces de ecuaciones no lineales utilizando el concepto de deflación de funciones. Adicionalmente se investigará la factibilidad de estimar numéricamente el orden de convergencia, consecuentemente la multiplicidad de una raíz, simultáneamente con el cálculo de la misma, para disminuir el efecto de las discontinuidades propias del método y optimizar el algoritmo.

Uso de TICs: Se continuará profundizando la línea de investigación de diseño de materiales educativos, priorizando el diseño de herramientas automatizadas para la generación y evaluación de contenidos que potencien aprendizajes significativos, especialmente las colaborativas, para asegurar la calidad del proceso de enseñanza-aprendizaje. Se seguirá trabajando con *software* libre, utilizando lenguajes de programación y base de datos que permitan seguir desarrollando módulos para integrar a la plataforma vigente (*software moodle*), o para la implementación de otras herramientas a ser usadas en cualquier instancia, de cualquiera de las modalidades de educación.

3. RESULTADOS OBTENIDOS/ESPERADOS

Conforme mencionado anteriormente este proyecto fue presentado en el año 2008, pero varios integrantes han participado en otros proyectos interdisciplinarios de investigación.

Se trabajaron técnicas de simulación discreta y continua aplicadas a simulación de centrales hidroeléctricas [9-11], se diseñaron recursos tecnológicos (Plataforma de Educación a Distancia Comahue - PEDCO), se realizaron experiencias educativas de dictado en la modalidad a distancia, se incorporaron técnicas de dinámica grupal en materias de programación [17-20].

Se pretende profundizar el estudio de las herramientas computacionales para simulación discreta existentes en el mercado direccionando la búsqueda y evaluación a las herramientas de *softwares* libre y de bajo costo. Realizar aplicaciones prácticas de modelado y simulación de sistemas de computación, procesos de piscicultura, etc., e identificar otras aplicaciones del medio que pueden estudiarse con esta metodología [1-8]. Continuar con el desarrollo de nuevos recursos hipermediales con el objetivo de complementar los dictados teóricos-prácticos en las distintas modalidades de educación. Implementar herramientas automatizadas para la evaluación del aprendizaje, generación de contenidos, seguimiento de las actividades de docentes y alumnos, etc., a fin de generar una base de conocimiento que permita mejorar la calidad educativa [14-19].

Diseñar e implementar algoritmos basados en métodos iterativos, particularmente Newton Raphson y las variantes que no utilizan derivada como el método de la Secante y de Muller, utilizando el concepto novedoso de deflación de funciones para obtener las raíces de funciones no lineales y su multiplicidad [12-13].

Las líneas de investigación referentes a simulación y desarrollo de TICs buscan principalmente profundizar el conocimiento sobre los fundamentos, las metodologías, los avances y las aplicaciones de tecnologías de última generación; pueden eventualmente dar origen a nuevos conocimientos científicos y tecnológicos. El desarrollo de algoritmos basados en deflación de funciones no lineales posee gran potencial para originar nuevos conocimientos científicos de gran aplicación práctica.

4. FORMACION DE RECURSOS HUMANOS

Se pretende contribuir a la formación de recursos humanos mediante:

- la participación en la ejecución del proyecto.
- la presentación de trabajos en revistas y congresos.
- el desarrollo de Tesis de Maestría y de Grado de varios integrantes del grupo. Actualmente dos integrantes cursan el segundo año de la maestría de Enseñanza de las Ciencias Exactas y Naturales y desarrollarán las respectivas tesis durante la ejecución del proyecto.
- continuar con el dictado de cursos en carreras de grado y posgrado (Simulación Discreta para la maestría de Estadística Aplicada y Simulación de Modelos para la maestría de Informática en proceso de implementación).
- el dictado y asistencia a cursos y seminarios.
- la interacción entre los integrantes de los proyectos que participan en el programa, con los de otros Departamentos, Facultades, Universidades y Empresas del país, y del Exterior.

5. BIBLIOGRAFIA

- [1] Kelton D., Sadowski R. P., Sturrock D. T., "Simulación con Arena", Mc Graw Hill, 2008.
- [2] Banks J., Carson J. S., Nelson B. L. y Nicol D. M., "Discrete System Simulation in Management Science", Prentice Hall, 2001.
- [3] Banks J., "Handbook of Simulation: Principles, Methodology, Advances, Applications and Practice", Wiley, 1998.
- [4] Chwif L., Medina A., "Modelagem e Simulação de Eventos Discretos – Teoria & Aplicações", Palas Athenas, Brasil, 2006.
- [5] Pidd M., "Computer Simulation in Management Science", Wiley, 1998.
- [6] Farrington, P. A., J. J. Swain, "Design of Simulation Experiments with Manufacturing

- Applications”, Proceedings of the Winter Simulation Conference, G. W. Evans et al., pp 69-75,1993.
- [7] Nance R. E. y R.G. Sargent, “Perspectives on Evolution of Simulation”, Operations Research, pp. 161-172, 2003
- [8] Sadowski, R. P., “Selling Simulation and Simulation Results”, Proceedings of the Winter Simulation Conference, G. W. Evans et al., pp. 65-68, 1993.
- [9] Karnopp Dean C., Donald L. Margolis y Ronald C. Rosenberg. “*System Dynamics: Modeling and Simulation of Mechatronic System*”.USA:Wiley-Interscience publication, 2000.
- [10] Martins A.,Tiago Filho G. L., Huerga M.y Laurent R., “Modeling and Simulation of hydraulic turbines using the bond graphs methodology”. IAHR, 24 Symposium on Hydraulic Machinery and Systems, Foz do Iguassu, Brasil, 2008.
- [11] Martins A, Tiago Filho G. L. y Laurent R, “La Sintonización de la Capacitancia en la Simulación del Golpe de Ariete con Bond Graphs”, , Revista PCH Noticias & SHP News, Centro Nacional de Referencias em Pequenos Aproveitamentos Hidroenergéticos, ISSN 1676-0220, Ano 9 N° 32, pp. 10-13, Brasil, 2007
- [12] Burden, R. y Faires J.D. *Numerical Analysis*, Brooks-Cole Publishing, 2004.
- [13] Gerard, Curtis F. y Patrick O. Wheatley. “*Applied Numerical Analysis*”. Massachussets: Addison Wesley Longman Inc, 2003.
- [14] Granda, J. ,”La Universidad en la Sociedad de la Información y el Conocimiento”. Revista Cognición N° 13 ISSN 1850-1974 Edición Especial II Congreso CREAD Andes y II Encuentro Virtual Educa UTPLoja, Ecuador ,2008.
- [15] Leal Ortiz, N. ,”Líneas de investigación y Calidad de la Educación Superior a Distancia”, Revista Cognición N° 13 ISSN 1850-1974 Edición Especial II Congreso CREAD Andes y II Encuentro Virtual Educa UTPLoja, Ecuador ,2008.
- [16] Omaña Cervantes, O., “Calidad a través del Seguimiento de Experiencias en Educación en Línea. La especialidad en Tecnología educativa de la Universidad Autónoma del Estado de Hidalgo”,. Revista Cognición N° 13 ISSN 1850-1974. Edición Especial II Congreso CREAD Andes y II Encuentro Virtual Educa UTPLoja, Ecuador, 2008.
- [17] C. Fracchia, A. Martins, C. Allan “Una Primera Experiencia del Dictado a Distancia de una Materia Introductoria a la Programación”, Primeras Jornadas de Ingreso y Permanencia en Carreras Científico-Tecnológicas, Universidad Nacional de Quilmes, Octubre, 2008.
- [18] C. Allan, A.Martins “Incorporación de recurso Metodologicos para la Enseñanza de Matemática en Carreras de Informática”., XX Jornadas de Matemática de la zona sur, Puyehue, Chile, 2006.
- [19] C.Fracchia, A.Martins “Uso de Recursos Tecnológicos y de Técnicas de Dinámica Grupal en Materias de Programación”,. Encuentro de Académicos e Investigadores Patagónicos de Argentina y Chile, Punta Arenas, Chile, 2005.
- [20] Fracchia C, Alonso de Armiño A, “PEDCO (*Plataforma de Educación a Distancia* Universidad Nacional del Comahue)”. Workshop de Tecnología Informática aplicada en Educación (WTIE). Congreso Argentino de Ciencias de la Computación. Universidad Nacional de La Matanza , San Justo, Buenos Aires, 2004.

RECUPERACIÓN PERSONALIZADA DE RECURSOS EDUCATIVOS

Ana Casali, Claudia Deco, Cristina Bender

Departamento de Sistemas e Informática
Facultad de Ciencias Exactas, Ingeniería y Agrimensura
Universidad Nacional de Rosario
(2000) Rosario, Argentina
Tel (+ 54 341) 4802649 int. 141
{acasali, deco, bender}@fceia.unr.edu.ar

Regina Motz

Instituto de Computación,
Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay
rmoz@fing.edu.uy

CONTEXTO

Esta línea de I+D se está llevando a cabo a través de proyectos de la Universidad Nacional de Rosario, de la Secretaría de Estado de Ciencia, Tecnología e Innovación de Santa Fe, y de un proyecto latinoamericano.

Los PIDs de la UNR involucrados son:

- [PID UNR ING125] *Búsqueda en Bases de Datos de Texto*, dirigido por Claudia Deco. (2007-2010)
- [PID UNR ING245] *Sistemas de agentes de software, para actuar en ambientes dinámicos e inciertos*, dirigido por Ana Casali. (2008-2009)

El proyecto de la provincia de Santa Fe es

- [PID SF 219308] *Sistema de apoyo al docente en la búsqueda y preparación de material didáctico para la enseñanza de las ciencias en las escuelas santafesinas*, dirigido por Ana Casali (2009)

El proyecto latinoamericano es:

- [LACCIR RFP2008] *JARDIN: Just an Assistant foR instructional DesIgN*, financiado por The Latin American and Caribbean Collaborative ICT Research (LACCIR), Microsoft, compuesto por integrantes de universidades de Uruguay, Brasil, Argentina, Ecuador y México, dirigido por Regina Motz (2009).

Resumen

Esta línea de investigación trata el problema de la recuperación personalizada inteligente de recursos de aprendizaje. Tiene como objetivo diseñar e implementar un sistema de recomendación que ayude a un usuario a encontrar los recursos educativos electrónicos que le sean más apropiados de acuerdo a sus necesidades y preferencias. Como hipótesis de trabajo se considera que se tienen diferentes repositorios de objetos de aprendizaje, donde cada objeto tiene metadatos descriptivos. Se propone utilizar estos metadatos para recuperar aquellos objetos que satisfagan no sólo el tema de la consulta, sino también el perfil de usuario, teniendo en cuenta sus características y

preferencias. Esto engloba el establecimiento de una estrategia de búsqueda adecuada y la definición de metadatos educacionales adecuados. También abarca el diseño de una arquitectura multiagente acorde a las diferentes funcionalidades del sistema y las arquitecturas de los agentes que la componen. Mediante la implementación de un prototipo se espera poder experimentar la arquitectura del sistema y la metodología de la recuperación propuesta.

1. INTRODUCCIÓN

En el dominio de la educación existe gran cantidad y diversidad de material que puede ser utilizado en la enseñanza y que constituye una importante contribución al proceso enseñanza-aprendizaje. Se conoce como Objeto de Aprendizaje a todo recurso digital que apoya a la educación y que puede ser reutilizado. El concepto de Objeto de Aprendizaje (Learning Object) abarca principalmente a un conjunto de materiales digitales los que como unidad o agrupación permiten o facilitan alcanzar un objetivo educacional. Algunos ejemplos de recursos digitales incluyen a textos, imágenes, cortos de video o audio, pequeñas aplicaciones Web, páginas Web completas que combinen texto, imágenes y otros medios de comunicación, entre otros.

En particular, con el desarrollo de la Web y su utilización masiva, se tiene una amplia gama de posibilidades de acceso a material útil e interesante para ser empleado tanto por un alumno que desea aprender un tema, como por un docente que desea preparar material didáctico. Sin embargo, se advierte una sobrecarga de información que obliga a estos usuarios a explorar espacios excesivamente densos, convirtiendo la selección de la información que les interesa en una tarea tediosa, que insume mucho tiempo y que es difícil de realizar sin la asistencia de herramientas de búsqueda intuitiva y eficiente. La recuperación de este tipo de material, por lo general se realiza en la Web a través de buscadores. Pero no siempre el resultado es el esperado por el usuario si la búsqueda se realiza sólo considerando palabras claves, porque un material recuperado no es el adecuado para todos los usuarios. Esto se debe a que los usuarios poseen distintas características y preferencias personales, que deberían también ser consideradas en el momento de la búsqueda.

En los últimos años, los sistemas recomendadores surgen para ayudar a resolver este tipo de problema puesto que son capaces de seleccionar, de forma automática y personalizada, el material que mejor se adapte a las preferencias o necesidades de un usuario. El uso de los sistemas multiagentes se ha incrementado para abordar el desarrollo de sistemas distribuidos complejos y en particular, se ha utilizado para el desarrollo de sistemas recomendadores. Esta tecnología de agentes es importante a la hora de modelar diferentes características que se espera de estos sistemas como por ejemplo: generar y considerar el perfil del usuario, inferir y agregar información proveniente de fuentes heterogéneas y distribuidas, obtener sistemas escalables, abiertos y seguros, y realizar la tarea requiriendo la menor intervención de las personas.

El enfoque de esta línea de investigación es desarrollar un sistema recomendador de recursos educativos que personalice la recuperación de los resultados, utilizando datos del usuario modelados en perfiles personales, y metadatos con las descripciones semánticas de cada recurso. El perfil del usuario se compone de sus preferencias y de la importancia relativa de cada una al momento de elegir un material dado. Esto en conjunto con los metadatos de cada recurso educativo constituye la base para el razonamiento del sistema recomendador.

2. LINEAS DE INVESTIGACION Y DESARROLLO

Para alcanzar los objetivos que proporcionen una ayuda a los problemas mencionados en la introducción, se han planteado distintas líneas de investigación las cuales se interrelacionan.

Una primera dirección está vinculada a las arquitecturas de agentes que los capaciten para actuar de forma flexible y eficiente, en entornos dinámicos e inciertos. En esta dirección se trabaja dentro del proyecto [PID UNR ING245], en el diseño de agentes BDI graduados (g-BDI) en un entorno social integrado por otros agentes. Resultados preliminares de la arquitectura g-BDI pueden verse en [11]. Se ha utilizado este modelo de agente para diseñar e implementar agentes recomendadores en el dominio turístico [12] y se ha realizado un diseño preliminar en el dominio de la educación [13]. En segundo lugar se trabaja en la expansión semántica de la búsqueda en el contexto del proyecto [PID UNR ING125]. El objetivo de esta línea es producir la estrategia de búsqueda temática. Cuando el usuario hace una consulta, ingresa un conjunto de términos que describen el tema de su interés. Luego es necesario un proceso que desambigüe estos términos y los expanda semánticamente incorporando sinónimos y conceptos relacionados. La salida del agente que se ocupe de esta tarea es una estrategia de búsqueda que consiste de la disyunción de las expansiones de cada concepto y luego se considera la conjunción de esas expansiones. Resultados en este sentido pueden verse en [3] [4].

Por último se trabaja en como utilizar las características y preferencias del usuario (docentes o alumno) para que un sistema recomendador no sólo recupere los recursos que respondan temáticamente a la consulta del usuario, sino que presente aquellos que sean los más adecuados de acuerdo a su perfil [3], [5] y [6]. Esta línea se trabaja en el marco de un proyecto de la Secretaría de Estado en Ciencia y Tecnología de la Provincia de Santa Fe [PID SF 219308] orientado en dar apoyo al docente en la búsqueda y preparación de material didáctico para la enseñanza de las ciencias en las escuelas santafesina. También en el marco del proyecto latinoamericano [LACCIR RFP2008] para asistir a la búsqueda de objetos de aprendizajes a partir de repositorios de esta región. Primeras propuestas en este sentido pueden verse en [1] y [9].

El objetivo global de esta línea de I+D es desarrollar un sistema recomendador de objetos de aprendizaje que asista al docente o alumno en la selección de material didáctico acorde a: la temática a abordar, el tipo de material buscado, el nivel académico del curso y las características y/o preferencias de cada alumno o grupo de alumnos.

Los objetivos específicos son:

- Estudiar la viabilidad de utilizar sistemas recomendadores en apoyo a la selección de material educativo.
- Diseñar la arquitectura del sistema recomendador como un sistema multiagente y diseñar los agentes que lo componen.
- Diseño de un repositorio de recursos de aprendizaje a utilizar en la experimentación.
- Desarrollar un prototipo de un sistema recomendador para la búsqueda de objetos de aprendizaje para la enseñanza de las ciencias.
- Experimentar y evaluar.
- Transferir los resultados a la comunidad científico-educativa.

3. RESULTADOS ESPERADOS

Los resultados esperados de esta línea de investigación son:

- Estudiar la viabilidad de utilizar sistemas recomendadores en apoyo a la selección de material educativo.
- Diseñar un repositorio que contendrá los objetos de aprendizaje obtenidos de la Web.
- Implementar un prototipo de un sistema recomendador para la búsqueda de objetos de aprendizaje para la enseñanza de las ciencias.
- Transferir los resultados a la comunidad científico-educativa.
- difundir los resultados mediante la publicación en revistas especializadas y la presentación de trabajos en congresos nacionales e internacionales.

4. FORMACIÓN DE RECURSOS HUMANOS

Tesis de grado (UNR – Licenciatura en Ciencias de la Computación)

- Evaluación de la Efectividad de un Motor de Búsqueda Web Basado en Ontologías y Técnicas de IR. A. Ponce.2008.
- Extensión de plataformas de desarrollo de agentes PRS/BDI para incluir incertidumbre y dinamismo. Alumno Adrián Biga, Directora: Ana Casali. , en curso.
- Agente recomendador de Objetos de Aprendizaje. Alumna Valeria Gerling, Directoras: Claudia Deco y Ana Casali, en curso.
- Presentación e implementación sobre plataforma e-learning de una técnica de recuperación y ordenamiento de documentos de acuerdo a características del usuario. Alumno: Matías Asás, Directoras: Claudia Deco y Cristina Bender, en curso.

Tesis de maestría

- ✓ Recuperación personalizada de e-cursos. C. Bender. (Universidad de la República, Uruguay) 2008.

Tesis de doctorado

- ✓ On Intentional and Social Agents with graded Attitudes. A. Casali, (Universidad de Girona, España), 2008.
- ✓ Mejora de la recuperación de información utilizando recursos lingüísticos. C. Deco. (UNR), en curso

5. BIBLIOGRAFIA

- [1].Deco Claudia, Bender Cristina, Casali Ana, Motz Regina. DESIGN OF A RECOMMENDER EDUCATIONAL SYSTEM.. 3ra. Conferencia Latinoamericana de Objetos de Aprendizaje LACLO 2008. Aguascalientes, México. Octubre 2008. pp 63-70 ISBN 978-970-728-067-0.
- [2].Motz Regina, Viola Alfredo, Palazzo de Oliveira José, Valdení de Lima José, Ochoa Xavier, Deco Claudia, Casali Ana, Bender Cristina, Pérez Alvarez Miguel Angel, Brunetto Angélica, Proença Mario. PROJECT JARDIN: JUST AN ASSISTANT FOR INSTRUCTIONAL DESIGN. 3ra. Conferencia Latinoamericana de Tecnología de Objetos de Aprendizaje LACLO 2008. Aguascalientes, México. Octubre 2008. pp 223-228. ISBN 978-970-728-067-0.
- [3].Ponce Adrián, Deco Claudia, Bender Cristina. PROPOSAL OF AN ONTOLOGY BASED WEB SEARCH ENGINE. En Proceedings del Workshop de Bases de Datos, CACIC 2008. Chilcito, Argentina, octubre 2008.
- [4].Deco Claudia, Bender Cristina, Severino Guimpel Federico, Reyes Nora. RECUPERACIÓN DE INFORMACIÓN EN BASES DE DATOS DE TEXTO. En Proceedings WICC 2008. General Pico, La Pampa, Argentina. Mayo 2008.
- [5].Bender, C., Motz, R., Deco, C., Saer, J. RECUPERACIÓN PERSONALIZADA DE e-CURSOS. En Proceedings del IX Congreso Iberoamericano de Informática Educativa, RIBIE 2008. Caracas, Venezuela, marzo 2008.
- [6].Bender, C., Deco, C., Saer, J., Fornari, J., Grieco, S., Motz, R. APLICACION DE METADATOS PARA LA BÚSQUEDA PERSONALIZADA DE INFORMACIÓN TECNICA. En Energeia Cuaderno de Investigación. Año 5 Nro 5. Publicación del Departamento de Investigación Institucional. Facultad de Química e Ingeniería "Fray Rogelio Bacon". pp 1-12 . ISSN 1668-1622. 2007.
- [7].Bender C., Deco C., Casali A., Motz R. UNA PLATAFORMA MULTIAGENTE PARA LA BUSQUEDA DE RECURSOS EDUCACIONALES CONSIDERANDO ASPECTOS CULTURALES. En Revista Iberoamericana de Tecnología en Educación y Educación en

- Tecnología. Diciembre 2006. Vol. 1 Nro. 1. (RedUNCI). ISSN 1850-9959. pp 20-29. 2006.
- [8]. Bender Cristina, Deco Claudia, Bernini María Belén, Asás Matías, Motz Regina. ORDENACIÓN DE TUPLAS PARA LA BÚSQUEDA DE MÚLTIPLES E-CURSOS SIMILARES. En Proceedings WTIAE, CACIC 2006. San Luis, Argentina, octubre de 2006.
- [9]. Casali Ana, Deco Claudia, Bender Cristina and Motz Regina. A MULTIAGENT APPROACH TO EDUCATIONAL RESOURCES RETRIEVAL. En Proceedings del Workshop on Artificial Intelligence for Education, en el marco del 35° JAIIO, ISSN 1850 2784. pp 35-41. Mendoza, Argentina. Septiembre 2006.
- [10]. Bender C., Deco C., Casali A., Motz R., Guzmán J. A MULTIAGENT PLATFORM FOR EDUCATIONAL RESOURCES RETRIEVAL. DRIVEN BY CULTURAL ASPECTS. En Proceedings del I Congreso de TE&ET. Aceptado con prioridad. ISBN 950-340373-1. pp 1-9. La Plata, Argentina. Agosto 2006.
- [11]. Casali A., Godo L. and Sierra C. GRADED BDI MODELS FOR AGENT ARCHITECTURES. Leite J. and Torroni P. (Eds.) CLIMA V, LNAI 3487, pp. 126-143, Springer-Verlag, Berling Heidelberg, 2005.
- [12]. Casali A., Godo L. and Sierra C. MODELLING TRAVEL ASSISTANT AGENTS: A GRADED BDI APPROACH. En Proceedings of the IFIP-AI, WCC, Volume 217, Artificial Intelligence in Theory and Practice, Max Bramer (ed.), Springer, pp. 415-424, 2006.
- [13]. Montaner M., Lopez B., de la Rosa J.L. A TAXONOMY OF RECOMMENDER AGENTS ON THE INTERNET, Artificial Intelligence Review, Kluwer Academic Publishers. Volume 19, Issue 4, pp. 285-330. June, 2003.
- [14]. Motz R., Guzmán J., Deco C. and Bender C. APPLYING ONTOLOGIES TO EDUCATIONAL RESOURCES RETRIEVAL DRIVEN BY CULTURAL ASPECTS. Journal of Computer Science & Technology. ISSN 1666-6038. JCS&T Vol 5, N° 4, pp 279-284, December 2005.
- [15]. Niinivaara O. AGENT-BASED RECOMMENDER SYSTEMS. Technical Report, University of Helsinki, Dept. of CS, 2004.
- [16]. Rao A. and Georgeff M. BDI AGENTS FROM THEORY TO PRACTICE, Technical Note 56, AAIL, April 1995.
- [17]. Romero C., Ventura S., Delgado J. and de Bra P. PERSONALIZED LINKS RECOMMENDATION BASED ON DATA MINING IN ADAPTIVE EDUCATIONAL HYPERMEDIA SYSTEMS. Second European Conference on Technology Enhanced Learning (EC-TEL 2007). Crete, Greece, 2007.
- [18]. Soonthornphisaj N., Rojsattarat E. and Yimngam S. SMART E-LEARNING USING RECOMMENDER SYSTEM. Lecture Notes in Computer Science. Springer Berlin, Heidelberg, Volume 4114. Computational Intelligence pp 518-523. 2006.
- [19]. Terveen L. G. and Hill W., BEYOND RECOMMENDER SYSTEMS: HELPING PEOPLE HELP EACH OTHER. In Carroll, J. (Ed.), HCI in the New Millennium. Addison Wesley, 2001.
- [20]. Wiley, D. CONNECTING LEARNING OBJECTS TO INSTRUCTIONAL DESIGN THEORY: A DEFINITION, A METAPHOR, AND A TAXONOMY. In D. A. Wiley (ed.) Instructional Use of Learning Objects. Editorial Association for Instructional Technology, 2002.
- [21]. Zaiane O.R. BUILDING A RECOMMENDER AGENT FOR E-LEARNING SYSTEMS. Proceedings of International Conference on Computers in Education, pp: 55-59, 2002.

Definición de los principales actores y sus roles para los cursos de EaD en la Facultad de Ciencias Económicas de la UNPSJB

Dans, Marta Isabel, Gomez Ana Maria, Gonzalez Maria Ester (Universidad Politécnica de Madrid) / Departamento Administración / Facultad de Ciencias Económicas / UNPSJB

CONTEXTO

La mayoría de los miembros que integran la unidad ejecutora del proyecto “Definición de los principales actores y sus roles para los cursos de EaD en la Facultad de Ciencias Económicas de la UNPSJB”, tienen más de diez años desarrollando investigación en la línea de la incorporación de TIC en educación superior. Por ser la unidad ejecutora en su mayoría de la Facultad de Ciencias Económicas de la UNPSJB, se han realizado aplicaciones en este ámbito. Recientemente se han incorporado docentes de la Facultad de Ingeniería, por lo que se espera comenzar a desarrollar aplicaciones en nuevas áreas.

RESUMEN

Actualmente se está estudiando a los principales actores que desempeñan un papel preponderante en los cursos de EaD en relación a las habilidades que deben desarrollar para interactuar a través de la plataforma MOODLE (Modular Object Oriented Distance Learning Environment, Plataforma modular para aprendizaje a distancia orientada a objetos), que fue adoptada por la Facultad de Ciencias Económicas como resultado del proyecto que antecede a este.

Los actores considerados en educación a distancia son: el profesor redactor de contenidos, el tutor, el diseñador gráfico, el docente especialista en didáctica y el alumno (MENA 1996).

Estos actores están presentes en los cursos sobre la plataforma adoptada por la Facultad. Tiene roles específicos y es necesario que reciban capacitación para lograr desempeñarse en sus funciones y aprovechar todas las ventajas que ofrece la plataforma.

Palabras clave: *e-learning, plataforma virtual, roles en MOODLE.*

1. INTRODUCCION

El proyecto que antecede a la presente proposición, “Plataforma virtual educativa para cursos de Educación a distancia”, nos permitió definir una nueva plataforma para los cursos de EaD, dictados en la Facultad de Ciencias Económicas de la UNPSJB.

Esta propuesta permite completar el proyecto anterior. Dado que percibimos la necesidad de precisar los actores y sus roles. Puesto que dicha plataforma brinda una variada gama de posibilidades que hasta el momento no habíamos experimentado, en cuanto a recursos a disposición de los cursos y actividades a implementar durante el desarrollo de los mismos. Consideramos que los actuales actores nos permitirán determinar las habilidades que deberán adquirir, si es que aún no las poseen, necesarias para la interacción con la nueva plataforma virtual aprovechando al máximo sus posibilidades.

La adaptación de los roles a las necesidades de nuestro entorno será especialmente enriquecedor para los cursos de Educación a Distancia.

La trayectoria en investigación en el área de tecnologías informáticas y de comunicación nos permitió corroborar lo expresado en los párrafos anteriores y nos exige continuar mejorando las distintas estrategias tecnológicas ya incorporadas en nuestro ambiente.

En los anteriores proyectos se han desarrollado: un modelo de diseño para desarrollo de software multimedial, se implementó el sitio Web de la Facultad de Ciencias Económicas de la UNP, se incorporaron los servicios de Internet de WWW y foros de discusión para materias con profesores viajeros y profesores locales. Oportunamente se comprobó que tales innovaciones mejoraron la calidad en cuanto a la comunicación docente - alumno durante el proceso de enseñanza aprendizaje.

Es el cuarto proyecto que continuamos en la línea de investigación en informática educativa. Diseñamos e implementamos la plataforma virtual para los cursos de educación a distancia, diseñamos el método MeCoVED (DANS 2003) para la construcción de videos didácticos digitales y seleccionamos la plataforma virtual para cursos de educación a distancia.

Además de mantener el crecimiento en las áreas ya experimentadas, aquí se propone investigar a fondo los roles necesarios para potenciar el entorno virtual de los cursos de educación a distancia, al servicio del aprendizaje de los alumnos.

En los años 2003 – 2004 la Facultad de Ciencias Económicas inició un proyecto de Educación a Distancia. Durante ese proyecto se definieron los actores: Profesores redactores de contenido, Tutores, coordinador de tutores y alumnos. Un actor fundamental fue el profesional

informático que administraba el medio maestro (la Web) y habilitaba los recursos de comunicación que requería cada curso editaba y subía las páginas Web. Se desarrollaron seis cursos y se realizaron dos pruebas piloto. En la primera experiencia fue un cuello de botella la dependencia que había con la carga de trabajo del especialista en informática.

De la primera experiencia surgió la necesidad de buscar una plataforma virtual para crear cursos que satisficiera las necesidades de la Facultad. Se realizaron estudios minuciosos y se seleccionó a MOODLE.

Los docentes redactores de contenidos

En nuestra primera experiencia, los redactores de contenido tuvieron la principal preocupación en producir, en medio digital, los materiales que soportaban el desarrollo de los contenidos. En nuestro caso se convino que además del material propiamente dicho, se produjeran ejercicios de autoevaluación, lecturas y actividades complementarias.

Al tratarse de materiales digitalizados se convino que debían tener cierta navegación. Por ejemplo, se debía poder saltar de un lugar del texto a otro que estuviera relacionado, clickeando en un vínculo. En algunos casos se presentó la necesidad de bifurcar a un documento diferente.

Esto requiere que el autor adquiera la habilidad de construirlo. Por ello se dieron charlas de capacitación a quienes lo solicitaron.

También se convino que los materiales disponibles en la Web debían estar en formato “pdf” (Portable Document Format) de manera que solo puedan ser visualizados o impresos con facilidad y no permitir otros fines.

Se acordó que como introducción a cada unidad, sería conveniente que hubiera un mapa conceptual con la relación entre los conceptos de la misma. Respecto a la construcción del mismo, no fue necesario capacitación alguna, ya que los redactores de contenidos contaban con la habilidad de construirlo por sí mismos. En este caso nos estamos refiriendo a los profesores de primer año que construyeron los materiales (matemática, filosofía, contabilidad y procesamiento de datos), no se extiende a todos los profesores en condiciones de construir los materiales de las distintas materias de los planes de estudios vigentes en la Facultad. Conocemos, que en algunos casos sería necesaria capacitación para adquirir esta habilidad.

La capacitación que recibieron estos actores fue acotada en temas y duración. En la mayoría de los casos requirieron de muy poco apoyo.

En nuestra segunda experiencia, sobre la plataforma Moodle, la diferencia radica en que la plataforma brinda una serie de recursos adicionales, y para utilizarlos es necesario conocer la interfaz que permite el acceso a los mismos. El uso de los recursos didácticos requiere de capacitación dado

que existe, como en todo software, una secuencia de pasos para lograr cada acción: colgar materiales, colocar etiquetas, utilizar el calendario, y muchas otras facilidades.

Los contenidos pueden presentarse en forma de página Web o enlazarse como archivos o construir lecciones, en las cuales el alumno puede ir avanzando después de contestar correctamente preguntas al finalizar la misma (este módulo presenta varias formas de configuración). El docente puede crear una estructura de directorios para organizar los distintos archivos que pone a disposición de sus alumnos.

Moodle soporta distintos tipos de autoevaluación pudiendo hacerse interactivos, on line, brinda la facilidad del análisis estadístico de los resultados de estas evaluaciones.

El docente puede cambiarlos, en el momento que desee, con absoluta independencia desde su computadora conectándose al sitio.

Se requiere del docente redactor de contenidos más habilidades aún que en la plataforma anterior, respecto del manejo de las herramientas informáticas. Por ejemplo sería conveniente que aprendiera a diseñar un documento tipo html para poder colocarlo en el curso como una página Web.

Es necesario que reciba capacitación para poder usar todas las facilidades de la plataforma Moodle.

Respecto a los docentes y la plataforma virtual es importante analizar dos aspectos la funcionalidad y la usabilidad. Para que los actores incorporen la funcionalidad es necesario diseñar cuidadosamente programas de formación continua los cursos de capacitación. Para aprender la usabilidad pensamos que deben implementarse espacios de socialización en los cuales los actores intercambien experiencias de uso y de esta forma continuar creciendo en esta nueva modalidad.

Mencionamos anteriormente que MOODLE permite incorporar diversas actividades entre las que cabe mencionar los objetos SCORM, los mismos subyacen bajo una filosofía muy específica de Objetos de Aprendizaje que en sí encierra un campo muy amplio de investigación. Para acercarnos a ellos, la unidad ejecutora del Proyecto de Investigación, inició trabajos conjuntos con otra unidad ejecutora de esta misma Universidad que se dedica al estudio de los mismos, cuya directora la Ing. Beatriz Rosanigo de la Facultad de Ingeniería, se convino en la pertinencia de realizar experiencias conjuntas en objetos SCORM y plataformas educativas.

Del trabajo de investigación conjunto, realizado durante el último año llegamos a la conclusión de que es necesario capacitar a los docentes en el diseño de páginas Web, para construir objetos de aprendizaje, la catalogación de los mismos para convertirlos en paquetes de contenidos SCORM y subirlos a las aulas virtuales.

En el marco del proyecto en curso hemos diseñado cursos de dos niveles para el uso de MOODLE. En el curso de primer nivel se enseñan los conceptos básicos de la plataforma y en el de segundo nivel la incorporación de la funcionalidad de Web 2.0 en la plataforma.

Los tutores

Los tutores recibirán la misma capacitación que los docentes redactores de contenido en cuanto al manejo de la plataforma, a su formación se agrega la funcionalidad que MOODLE ofrece para el seguimiento y la calificación de los alumnos, durante la realización de las distintas actividades que ofrecen los cursos.

Existe en la Facultad un Proyecto de Educación a Distancia, dirigido por la Profesora Adriana Fantini que considera la capacitación de los tutores en su función específica de tutor. Por lo que en este proyecto no se ocupa de tal formación.

Los alumnos

El mundo moderno y globalizado le exige al alumno de hoy habilidades como autoformación, capacidad de trabajar en grupo, ser crítico en cuanto al análisis de la información que recibe por los distintos medios, utilizar los entornos virtuales de aprendizaje.

La capacitación de los alumnos en el uso de la plataforma no está prevista en el contexto de este proyecto por considerar que las dudas que surjan a los alumnos serán evacuadas por los propios compañeros o los docentes de las cátedras.

Estado actual

En el año en curso es el segundo del proyecto, están creados los cursos de capacitación y se llevarán a cabo los mismos.

Se prevé realizar los ajustes necesarios para volverlos a impartir en años subsiguientes.

Hemos notado que las nuevas versiones de la plataforma ha mejorado mucho la configuración de las actividades y definidos nuevos roles. Lo que implica una adaptación de los cursos a las últimas modificaciones de la plataforma.

2. LINEAS DE INVESTIGACION y DESARROLLO

La línea de investigación es la aplicación de las TIC en educación superior.

Dentro de ella se ha incursionado en distintas aristas: construcción de multimedios, creación de videos didácticos digitales, la Web educativa y actualmente plataformas virtuales para cursos.

En el proyecto actual “Determinación de los actores y sus roles en los cursos de educación a distancia sobre la plataforma virtual MOODLE”. Se ajustan los roles y se define la capacitación necesaria

para que adquieran las habilidades, respecto a la plataforma los distintos actores.

3. RESULTADOS OBTENIDOS/ESPERADOS

En el corriente año por iniciativa propia se crearon más de 20 cursos nuevos. Los profesores responsables de las materias recibieron ayuda por correo electrónico, Skype y el material ya creado para los cursos diseñados.

Es importante destacar que estos profesores tienen un destacado interés por utilizar la herramienta de modo que su predisposición hace las dificultades que se presentan no son obstáculo para seguir utilizándola.

El curso inicial está totalmente armado en cuanto a material, tareas y programa. En estos días se lanzará una invitación para la inscripción.

El curso avanzado está en construcción, se espera lanzar para la segunda parte del año.

La mayoría de la unidad ejecutora compuesta por doce docentes utiliza la plataforma para sus cursos. Lo cual produce un enriquecimiento a la hora de proponer actividades en los cursos de formación de la herramienta.

Roles de la plataforma MOODLE

Los redactores de contenidos – Profesores y los tutores (eventualmente) deben tener la habilidad de:

- Configurar el curso
- Asignar roles
- Editar información
- Definir Grupos
- Realizar Copias de seguridad
- Restaurar cursos
- Importar cursos
- Reiniciar un curso
- Ver informes (seguimiento de los alumnos)
- Crear y editar las distintas actividades que ofrece la plataforma,
- Utilizar y definir escalas para calificar los Foros, Glosario .y Tareas.
- Cambiar la manera en que se calculan, se agregan y se muestran las calificaciones.
- Acceder a los Registros de accesos y participación
- Subir Archivos

Los alumnos

- Entrar y autenticarse
- Actualizar el perfil
- Definir el formato de e-mail
- Definir los modos de suscripción a los foros(autosuscribirse, darse de baja, etc)
- Elegir recibir todos los mensajes de Moodle en un mensaje diario.
- Activar Seguimiento de Foros: los mensajes del foro que no se hayan leído aparecerán resaltados,
- Acceder a los contenidos del curso
- Acceder a sus calificaciones

- Visualizar la lista de participantes.
- Participar en las distintas actividades propuestas.

Creadores de Cursos

- Crear un curso sin necesidad que intervenga el administrador del sitio Moodle
- Asignar profesores y actuar como profesor con privilegios de edición
- Sr el profesor principal, el jefe de departamento o el coordinador del programa

Administrador

- Gestionar el sitio Moodle
- Hacer cualquier cosa en cualquier parte del sitio.

Invitados

- Acceder a través de dos formas: en clave de acceso o sin ella
- Acceder siempre para “solo lectura”
- No pueden:
 - Enviar mensajes a discusiones
 - Editar páginas Wiki
 - Contestar cuestionarios
 - Enviar tareas
 - Contribuir al contenido o comentario de los glosarios
 - Ver el contenido de los paquetes SCORM

Correspondencia de los roles

Analizando las actividades que pueden realizar los distintos actores que implementa la plataforma, vemos que se adaptan a los roles existentes en el contexto de la Facultad de Ciencias Económicas.

Por otro lado cabe destacar que es posible redefinir los permisos de los roles de la plataforma, los mencionados aquí son los permisos por defecto que ya están asignados. Consideramos que los mismos son adecuados a cada función.

4. FORMACION DE RECURSOS HUMANOS

La unidad ejecutora se ha formado en el uso de la plataforma y en otras herramientas que complementan el uso de la misma como lo son los simuladores. La construcción de videos educativos, multimedios, etc.

Hasta el momento se han formado docentes de más de 20 cátedras.

La directora del presente proyecto es tutora de pasantes de la Facultad de Ingeniería de la carrera de Licenciado en Informática. Los alumnos se capacitan en el diseño Web, instalación, mantenimiento de distintas herramientas como son MOODLE y JOOMLA.

Se mantiene estrecha relación con otros grupos de la Universidad Nacional de la Patagonia, con docentes de otras universidades nacionales y con docentes de España.

Actualmente se está desarrollando un proyecto de aplicación de TIC en comunidades rurales patagónicas, la unidad ejecutora está formada por docentes investigadores de la Facultad de Ingeniería, de Humanidades y de Ciencias Económicas, con la UNED (Universidad Nacional de Educación a Distancia) de Madrid, financiado por la AECID (Agencia Española de Cooperación Internacional para el Desarrollo).

5. BIBLIOGRAFIA

MENA Marta (1996) “La educación a distancia en el sector público, manual para la elaboración de Proyectos”, INAP – DNC.

DANS, Marta (2003) “El medio maestro virtual de EaD” en la Facultad de Ciencias Económicas de la UNPSJB Anuario.

DANS, Marta (2003) “El método McCoVED para la construcción de videos didácticos digitales”.

MENA Marta (1996) “La educación a distancia en el sector público, manual para la elaboración de Proyectos”, INAP – DNC.

MUÑOZ, Miguel Ángel “Nuevos entornos y posibilidades telemáticas en educación”

VALZACCHI Jorge Rey (1996) “Internet y Educación” Ediciones Horizonte.

GARCÍA ARETIO, L. (2001). La educación a distancia. De la teoría a la práctica. Barcelona-España: Editorial Ariel.

ALONSO, C. M., GALLEGRO GIL, D.J. (2000) “Aprendizaje y Ordenador”. Madrid: Dykinson.

APROA (2005) “Aprendiendo con Repositorio de Objetos de Aprendizaje.” El proyecto Aproa es liderado por la Universidad de Chile, con el apoyo de Instituciones Ejecutoras y Contrapartes. Disponible en: <http://www.aproa.cl/1116/propertyvalue-5538.html> [consultado: 14-09-2007]

MOODLE DOC: <http://docs.moodle.org/es>

Repositorio de Objetos de Aprendizaje

Beatriz Rosanigo⁽¹⁾, Pedro Bramati⁽²⁾, Alicia Paur⁽¹⁾, Marta Saenz Lopez⁽¹⁾,
Gloria Bianchi⁽¹⁾, Ester Livigni⁽³⁾

brosanigo@yahoo.com.ar, pedrobramati@speedy.com.ar, abpaur@gmail.com, martasl@speedy.com.ar,
bianchi_gloria@yahoo.com.ar, elivigni@yahoo.com.ar

⁽¹⁾ Departamento de Informática – ⁽²⁾ Departamento de Ingeniería – ⁽³⁾ Departamento de Matemática

Facultad de Ingeniería – Sede Trelew – U.N.P.S.J.B. – Te-Fax (02965) 428402

CONTEXTO

Nuestra área de investigación apunta a mejorar el proceso de enseñanza-aprendizaje promoviendo el protagonismo del sujeto como ser que elabora activamente el conocimiento y facilitando el trabajo que para alumno y profesor supone la tarea de formación.

Buscamos integrar la interactividad que proporcionan las TIC en el proceso de enseñanza-aprendizaje, y facilitar la construcción y uso de material didáctico y herramientas educativas, aplicando conceptos modernos de la Ingeniería de Software y poniendo énfasis en la reusabilidad y flexibilidad de las soluciones, así como en la posibilidad de evolución y aseguramiento de la calidad.

Este grupo de investigación viene trabajando en la línea de Informática Educativa desde 1997 con los siguientes objetivos generales:

- ✓ Generar líneas de investigación en Informática aplicada a la Educación, que haga uso de tecnología de punta y procure la integración del saber.
- ✓ Generar conocimiento en áreas de interés para la Informática y la Educación, que permita un continuo desarrollo y actualización tanto de sus estudiantes como de los profesores.
- ✓ Complementar el proceso de enseñanza-aprendizaje con ayuda de las TIC.
- ✓ Promover el trabajo y aprendizaje colaborativo de todos los actores y beneficiarios en cada proyecto.
- ✓ Propender a una adecuada gestión del conocimiento y de la información.

Particularmente, en el proyecto actual nos hemos propuesto:

- ✓ Investigar sobre el diseño de OA.
- ✓ Determinar modelos de clasificación de OA que faciliten la secuenciación de contenidos.
- ✓ Diseñar un repositorio de OA.

RESUMEN

En este trabajo se presentan los objetivos y alcances del proyecto de investigación PI 620-10E067, “Hacia un repositorio de objetos de aprendizaje”.

Mediante este proyecto, nos proponemos *incentivar* el diseño de propuestas de enseñanza con materiales multimedia para la Web, que medien en el proceso educativo potenciando la construcción del conocimiento, *organizar y facilitar* el acceso a los materiales educativos de la Institución y a la gran cantidad de recursos Web que en este momento están en la red de forma desordenada, y *posibilitar* que los docentes universitarios tenga a mano herramientas de fácil acceso y de sencilla utilización para crear, utilizar y localizar material educativo.

Para ello, estamos trabajando en la implementación de un Repositorio de Objetos de Aprendizaje, con metadatos que siguen los estándares de la IEEE y especificaciones SCORM (Sharable Content Object Reference Model) para compartir, reutilizar, importar y exportar objetos de aprendizaje, compatibles con distintas plataformas educativas.

Palabras clave: Objeto de aprendizaje, Repositorio de Objetos de Aprendizaje, SCORM, metadatos

1. INTRODUCCIÓN

La innovación que se busca en los ambientes de enseñanza-aprendizaje con la utilización de TIC motoriza la investigación, análisis y evaluación de nuevas herramientas que permitan mejorar la formación de los futuros profesionales.

En los últimos años, con el interés de compartir y reutilizar recursos en el ámbito educativo, ha surgido el concepto de *Objetos de Aprendizaje* (OA). Este concepto se aplica a materiales digitales creados como pequeñas piezas de contenido o de información (Wiley, 2002), con la finalidad de maximizar el número de situaciones educativas en las que el mismo pueda ser utilizado.

La esencia de este concepto es compartir y reutilizar recursos educativos en procesos de aprendizaje apoyados por tecnología. Entre las ventajas de su utilización se destaca la notable reducción de costos y tiempos de producción y distribución.

Un OA “es un recurso docente digital empaquetado conjuntamente con un recubrimiento semántico estandarizado” (García Peñalvo, 2006), compatible con diversos ambientes y sistemas de administración de aprendizajes, fácil de migrar de una plataforma a otra, fácil de localizar, acceder, archivar y reutilizar.

Los OA posibilitan personalizar la educación, haciéndola más dúctil y adaptando los recursos didácticos a:

- ✓ las necesidades de cada momento: enseñanza, aprendizaje, revisión, afianzamiento y/o evaluación,
- ✓ las motivaciones individuales y/o grupales,
- ✓ los estilos de aprendizaje y enseñanza,
- ✓ la linealidad o no de los distintos temas,
- ✓ otros.

Para garantizar una fácil reusabilidad de los OA, poder compartirlos y agilizar la ubicación de los mismos debe establecerse un criterio de ensamble que desencadene en unidades de conocimiento más abarcativas. También es necesario describirlos mediante el uso de *metadatos* en formatos conocidos.

Los metadatos son un conjunto de atributos que describen el contenido, el contexto y la estructura del OA, permitiendo su localización y reutilización. La IEEE ha desarrollado un conjunto de metadatos para OA denominado LOM (Learning Object Metadata). Se utilizan en forma de descriptores del objeto de aprendizaje, almacenando información dividida en nueve categorías (IEEE, 2002), relativa a su procedencia, su idioma, autores, el propósito del objeto en sí, entre otros datos.

Utilizando los metadatos podemos clasificar y recuperar los OA para ciertos perfiles de usuario y temática o relacionarlos entre ellos para ampliar contenidos.

Una vez descrito el OA mediante esos metadatos, se cataloga en *Repositorios de objetos de aprendizaje* (ROA) para quedar disponible al público, y ser incorporado a diferentes experiencias de aprendizaje. Un ROA es una colección ordenada de objetos de aprendizaje que brinda facilidades para ubicarlos por contenidos, áreas, categorías y otros descriptores. Los ROA proporcionan un marco para uniformar mecanismos de distribución y acceso a los OA, posibilitando la colaboración entre autores y usuarios, contribuyendo a evitar la duplicación de esfuerzos.

Resulta necesario entonces, por un lado, tener los OA disponibles a los potenciales usuarios en la Web, almacenándolos en repositorios de diferentes tipos; y por otro, que los OA provean información semántica y pedagógica que facilite su hallazgo y reutilización en la Web, haciendo uso de ontologías que especifiquen de forma consistente los conceptos y estructura.

Para poder reutilizar los materiales en distintas plataformas y escenarios han surgido estándares que permiten la documentación, búsqueda y distribución de los contenidos educativos que se generan. Entre los más importantes se puede mencionar IMS desarrollado por el Global Learning Consortium (IMS, 2003), que propone especificaciones basadas en tecnologías abiertas (XML) para facilitar las actividades de aprendizaje sobre tecnología web, y SCORM desarrollado por Advanced Distributed

Learning Initiative (ADL, 2002), que es un conjunto de estándares y especificaciones para compartir, reutilizar, importar y exportar OA, es expandible e incluye a trabajos de IEEE, y de IMS para algunas de sus funciones.

SCORM proporciona un marco de trabajo y una referencia de implementación detallada, que indican cómo deben publicarse los contenidos, cómo deben usarse los metadatos y cómo debe representarse la estructura de los cursos.

Se caracteriza por posibilitar la creación de contenidos que pueden importarse dentro de diferentes Sistemas Gestores del Aprendizaje genéricamente conocido como LMS (Learning Management System), tales como Moodle, ATutor, Dokeos, y hacer uso de la Web como medio de instrucción.

La utilización de estándares permite centrarse en los aspectos educativos, dejando de lado los tecnológicos, produciendo contenidos de mayor calidad a la vez que se facilita la reutilización de los mismos de acuerdo a las diferentes estrategias de aprendizaje, permitiendo trabajar en diferentes plataformas educativas de forma transparente, gracias a la interoperabilidad entre ellas.

Un paquete SCORM es un archivo comprimido en formato ZIP que empaqueta los OA junto con un índice, contiene los recursos y el manifiesto, que es un archivo XML llamado *imsmanifest.xml* que contiene referencias a los recursos, la organización y los metadatos, y las hojas de estilo que permiten interpretarlo. Los recursos son los archivos digitales con el real contenido del OA. Pueden ser de cualquier tipo, desde páginas HTML hasta video digital.

Los metadatos usados por SCORM permiten describir el contenido de los recursos y se basan en el estándar LOM.

Un OA conforme a SCORM se denomina SCO y su estructura es similar a la de la figura 1. Las plataformas educativas que soportan SCORM, pueden desplegar automáticamente el SCO manteniendo el índice de contenidos junto con los enlaces para acceder a los recursos.

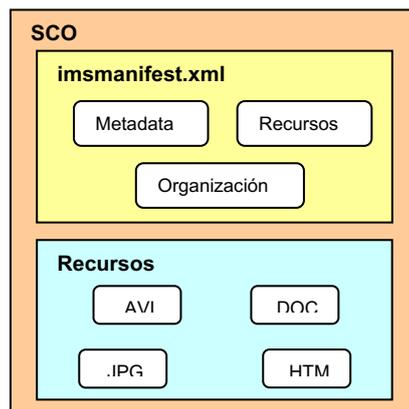


Figura 1: Estructura de un OA Scorm

Si disponemos de un repositorio bien gestionado, y que se pueda acceder desde un LMS, los docentes y alumnos podrían acceder a contenidos que refuerzan el aprendizaje de cierta área, y se facilitaría la comunicación entre ellos, convirtiendo la enseñanza en un campo más global y accesible.

En resumen, podemos decir que el principal beneficio que un OA aporta a la educación es compartir y reutilizar el material educativo sin restricciones tecnológicas ni pedagógicas. Mediante ello, se intenta no “reinventar la rueda”, aportando *flexibilidad*, ya que el mismo recurso puede utilizarse y combinarse en distintos contextos; *adaptabilidad*, pues facilita al diseñador poder seleccionar y componer recursos según las necesidades de aplicación; *administración del contenido*, facilitado por los metadatos que describen los recursos permitiendo su catalogación, almacenamiento y control; *interoperabilidad y portabilidad* que elimina los problemas de incompatibilidad entre plataformas, pudiendo trabajar en diferentes plataformas de hardware y software y moverse y almacenarse de manera transparente, sin cambio alguno en estructura o contenido.

En la gestión y utilización de OA intervienen varios procesos, tales como producción y control de calidad del OA, empaquetamiento, almacenamiento en el ROA, y carga en el LMS.

Para la producción de un OA se distinguen dos etapas principales: *Pedagógica*, en la que se definen objetivo educativo, contenidos, actividades de aprendizaje y/o evaluación, y

Tecnológica, en la cual se adecuan los contenidos y actividades para su presentación en Web y se genera el SCO para almacenarlo en un repositorio y publicarlo en un LMS.

Desde un punto de vista pedagógico, un OA abarca contenidos y considera los procesos de aprendizaje necesarios para el óptimo desarrollo del proceso de enseñanza y aprendizaje. Como recurso pedagógico, un OA debe atender a distintos tipos de usuarios considerando las características individuales de cada uno de ellos y flexibilizando las estrategias acorde a los estilos de aprendizaje.

Desde esta perspectiva el diseño de un OA resulta un desafío para el docente, quien debe elegir el contenido, crear la forma de presentación e interacción, apoyándose en las características de los usuarios o destinatarios.

Para completar la información de metadatos y empaquetado, existen varias herramientas que facilitan la tarea, de modo que el mismo docente que crea el material, podría realizarlo.

Cuando queremos armar un curso o lección basados en OA, primero debiéramos fijarnos si ya existen OA que puedan reutilizarse, y si no los hubiera, proceder a crearlo. La figura 2 muestra las etapas requeridas.

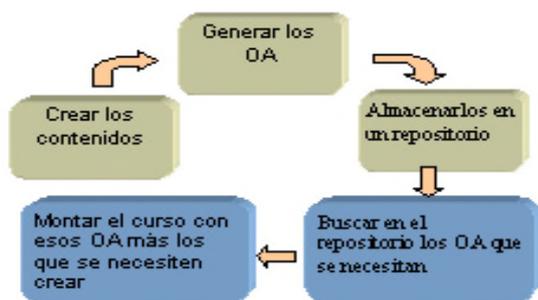


Figura 2: Etapas para creación y uso de OA

En el caso de poder reusar los OA, armar un nuevo curso sólo requiere de las fases azules (las dos últimas).

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Los principales temas abordados son:

- **Objetos de aprendizaje:** En este tema hemos analizado criterios y patrones para

su diseño, criterios de evaluación de la calidad del OA, consideraciones para la evaluación de los contenidos, criterios de clasificación, estudio de taxonomías y ontologías, estándares existentes. Se están desarrollando y produciendo OA en diferentes áreas y experimentando con diseño educativo basado en objetos de aprendizaje.

- **Repositorios de objetos de aprendizaje:** Análisis de repositorios existentes, con énfasis en la funcionalidad, prestaciones y distinción de aspectos positivos y negativos. Estudio y análisis de estándares. Modelos arquitectónicos y criterios de diseño de ROA. Técnicas de búsqueda. Protocolos de interoperabilidad.
- **Plataformas de código abierto para gestión de contenidos:** Análisis de plataformas existentes. Posibilidades de extensión y personalización. Instalación, personalización, configuración y experimentación en Moodle y ATutor. Mediación para el acceso a OA en Moodle y ATutor. Estudio de modelos y escenarios para la integración de objetos de aprendizaje y herramientas de gestión de contenido.
- **Herramientas de diseño de material educativo.** Investigación y experimentación con herramientas de software libre que facilitan la generación de materiales educativos, especialmente aquellas que puedan ser usadas por docentes con pocos conocimientos informáticos y que permitan crear actividades interactivas. Entre ellas, Hot Potatoes, PHP Webquest, JClíc.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se espera contribuir al desarrollo de un repositorio de objetos de aprendizaje, proporcionando una herramienta de código abierto y de fácil uso, para que las distintas unidades académicas cooperen entre sí con el objetivo común de mejorar el proceso de enseñanza-aprendizaje, concentrando los esfuerzos en mejorar la calidad.

Se ha llevado adelante la indagación bibliográfica y la capacitación en temas específicos: objetos de aprendizaje, técnicas avanzadas de diseño, técnicas de visualización y representación, tópicos avanzados de desarrollo en ambientes Web, diseño de material para educación a distancia, metodologías de enseñanza on line, estilos de aprendizaje.

Se han estudiado y analizado los estándares y especificaciones para e-Learning, adoptando IEEE-LOM y SCORM. Se han desarrollado varios SCO en diferentes temáticas de las carreras de la Facultad, con los que se han hecho experiencias de utilización en diferentes contextos y plataformas.

Se está trabajando en la definición de diferentes taxonomías para clasificación de los OA y en la modelación de ontologías tanto desde el punto de vista tecnológico como desde el punto de vista social relacionadas con el uso y aplicación de los OA, para facilitar los procesos automatizados de búsqueda y recuperación de recursos, propios de la web semántica.

4. FORMACIÓN DE RECURSOS HUMANOS

En julio 2008 se concluyó y defendió la tesis de magíster “Objetos de Aprendizaje. Factores que potencian su reusabilidad” que permitió a la integrante Alicia Paur obtener el título de Master en Ingeniería de Medios para la Educación, del programa Erasmus Mundus – Euromime.

Actualmente dos alumnos están trabajando en tesis de grado.

Se han incorporado alumnos e investigadores jóvenes en las distintas etapas.

Hay una fuerte interacción con otros grupos de investigación en áreas afines, tanto de nuestra propia universidad como de otras del país y de España, con los que se han realizado diferentes actividades.

5. BIBLIOGRAFÍA

ADL Advanced Distributed Learning (2002). “Emerging and Enabling Technologies for the design of Learning Object Repositories Report”. <http://xml.coverpages.org/ADLRepositoryTIR.pdf>, último acceso en Agosto 2007.

APROA (2005) “Aprendiendo con Repositorio de Objetos de Aprendizaje”, último acceso en Agosto 2007. <http://www.aproa.cl/1116/propertyvalue-5538.html>

Berners_Lee, T., Hendler, J., Lassila, O (2001) The Semantic Web. Scientific American, May Issue.

García Peñalvo Francisco José (2006), Entrevista en Learning Review: “Dos temas controvertidos en eLearning: Objetos de Aprendizaje y Calidad”. <http://www.learningreview.es/content/view/447/279/> Accedido en marzo 2008.

IEEE (2002) IEEE Standard 1484.12.1, Learning Technology Standards Committee. IEEE Standard for Learning Object Metadata., Institute of Electrical and Electronics Engineers, New York. http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf URL último acceso en Febrero 2008.

IMS (2003) Digital Repositories Interoperability Core Functions: Best Practice Guide, Information Model, and XML Binding Copyright 2003 IMS Global Learning Consortium, Inc. Version 1.0 Final Specification, Enero 2003.

López, C. (2005) “Los Repositorios de Objetos de Aprendizaje como soporte a un entorno e-learning”, Tesina doctoral, Universidad de Salamanca. (Director Francisco José García Peñalvo). Recuperado de <http://www.biblioweb.dgsca.unam.mx/libros/repositorio/s/index.htm>, Agosto 2006.

SCORM Sharable Content Object Reference Model. <http://www.adlnet.org/>.

Sicilia M.A. (2005) “Reusabilidad y reutilización de objetos didácticos: mitos, realidades y posibilidades”. RED: Revista de Educación a Distancia, <http://www.um.es/ead/red/M2/>

Smith Nash, Susan (2005) “Learning Objects, Learning Object Repositories, and Learning Theory: Preliminary Best Practices for Online Courses” Interdisciplinary Journal of Knowledge and Learning Objects Volume 1, 2005 Consultado en: <http://ijklo.org/Volume1/v1p217-228Nash.pdf>

Wiley David A. (2002) “Learning Objects”. En Kovalchick & Dawson (Eds.). Educational Technology. An Encyclopedia. Santa Bárbara: ABC-CLIO.

CARACTERIZACIÓN DE PROBLEMAS DE APRENDIZAJE BASADA EN EXPLOTACIÓN DE INFORMACIÓN

Elizabeth Jiménez Rey, Darío Rodríguez, Paola Britos, Ramón García-Martínez

Programa de Maestría en Tecnología Informática Aplicada a la Educación. Facultad de Informática. UNLP
Laboratorio de Sistemas Operativos y Base de Datos. Facultad de Ingeniería. UBA
Centro de Ingeniería de Software Ingeniería del Conocimiento. Escuela de Postgrado. ITBA
Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA
Área Ingeniería del Software. Licenciatura Sistemas. UNLa
Área Ingeniería del Software. Unidad Académica Río Gallegos. UNPA

ejimenezrey@yahoo.com.ar, {drodrigu,pbritos}@itba.edu.ar, rgarciamar@fi.uba.ar

CONTEXTO

La propuesta articula líneas de trabajo del Proyecto “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaría de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires

RESUMEN

Se presenta una línea de investigación que se focaliza en la aplicación de herramientas de Minería de Datos basada en Sistemas Inteligentes para la identificación y predicción de problemas de aprendizaje de los estudiantes (y sus causas).

El interés principal radica en la exploración de la posibilidad de construir un proceso que permita al Docente, por una parte, revisar la precedencia de los aprendizajes significativos considerados en el diseño del currículo y, por otra, reflexionar sobre las estrategias de su práctica docente para promover su mejoramiento.

Se aborda el estudio de los diversos métodos de la Minería de Datos: algoritmos de inducción, algoritmos genéticos, redes neuronales y redes bayesianas, para establecer el método que resultará más confiable para resolver el problema.

Actualmente, el trabajo de investigación está centrado en el estudio de los conceptos no comprendidos por los estudiantes de dos cursos de la asignatura Computación de carreras no informáticas de la Facultad de Ingeniería de la Universidad de Buenos Aires.

Palabras clave: *Explotación de Información aplicada en Educación, Problemas de Aprendizaje en Educación.*

1. INTRODUCCIÓN

En esta línea de investigación, el interés se centra en permitir a los Docentes la revisión de sus estrategias de enseñanza para favorecer en los estudiantes el

desarrollo de la cognición, en su más amplio sentido: la comprensión y la capacidad para pensar y conocer cada vez mejor [Litwin, 2008]. En este escenario, se explorará la posibilidad de construir un proceso que permita al Docente revisar la precedencia de los aprendizajes significativos [Joyce y Weil, 2002] de los estudiantes considerados en el diseño del currículo. La solución propuesta es una segmentación de los procesos intuitivos del Docente (de construcción artesanal y fuertemente ligados a su experiencia) que se estructura en un proceso de varias etapas: 1) Construcción del Mapa de Aprendizajes Significativos, que se articulará en concepto-atributo-valor [García Martínez y Britos, 2004] a determinar su existencia en los instrumentos de evaluación. 2) Explotación de Información sobre la base de evaluación. 3) Ratificación o Rectificación del Mapa de Aprendizajes Significativos.

Se estudia actualmente los casos correspondientes a dos cursos de la asignatura Computación de carreras no informáticas de la Facultad de Ingeniería de la Universidad de Buenos Aires.

La asignatura Computación es formativa, básica y obligatoria para alumnos de todas las carreras de Ingeniería, excepto para aquellos que cursan Ingeniería Electrónica e Ingeniería en Informática. Es cuatrimestral y tiene una carga horaria de cuatro horas por semana (sesenta y cuatro horas por cuatrimestre). La única asignatura correlativa posterior es Análisis Numérico.

Los alumnos pueden cursar la asignatura Computación en distintas etapas de avance en sus planes de estudio pero casi todos los alumnos la cursan en los primeros años de su carrera y tienen una edad comprendida entre 18 y 20 años. Algunos pocos alumnos provienen de escuelas técnicas y tienen conocimiento y experiencia previa en programación pero la mayoría de los alumnos nunca programó. La algoritmia representa para los alumnos un nuevo paradigma para resolver problemas y les produce un fuerte impacto en su predisposición al aprendizaje que, en muchos casos, se traduce en falta de motivación o en rechazo.

Son los objetivos *generales* de enseñanza de la asignatura:

- Brindar al alumno una visión global de la Computación, de manera que comprenda conceptos y técnicas de la disciplina que en su futuro profesional lo habilitarían para interactuar en forma interdisciplinaria con pares y profesionales en informática sin problemas de comunicación.
- Lograr que el alumno se compenetre con las tecnologías y herramientas fundamentales de la Computación, de manera que aprenda a usar la computadora como herramienta de trabajo, conociendo su precisión, capacidad y limitaciones.

Son los objetivos *específicos* de enseñanza de la asignatura:

- Concientizar al alumno de la importancia de la Algoritmia como paradigma de resolución de problemas y de la Programación como práctica y ejercitación en la resolución de problemas, como promotores del desarrollo de la capacidad de abstracción, es decir, la capacidad de relacionar esquemas de solución con la resolución de problemas algorítmicos, con hincapié en el método científico para lograr ese objetivo.
- Enseñar al alumno de Carreras de Ingeniería el Análisis, la Sistematización, Programación y Procesamiento de distintos problemas de tipo técnico-científicos, a fin de que dichos conocimientos le resulten de utilidad ya sea en el desarrollo de la carrera como así también en su actividad profesional.
- Sustentar la estrategia de enseñanza y de aprendizaje de los conceptos y procedimientos en los recursos tecnológicos, como mediadores de los procesos y las habilidades cognitivas a inducir en el alumno.

Son los objetivos de aprendizaje de la asignatura:

- Conocimiento Declarativo (SABER que el alumno debe adquirir): capacidad de resolución de problemas con la computadora. El alumno deberá conocer QUÉ es el Hardware (computadora) y QUÉ es el Software (programas).
- Conocimiento Procedimental (SABER HACER que el alumno debe desarrollar): habilidad para encontrar la Solución (representada por un programa) a un Problema (representado por un enunciado), es decir, CÓMO construir un programa solución a partir de un enunciado problema.
- Conocimiento Actitudinal (SABER ESTAR): dominio de actitudes de compromiso, responsabilidad, solidaridad y respeto del estudiante en el proceso de construcción de su conocimiento. La interacción continua del alumno con sus pares, con el Docente y con el material de estudio, favorecerá la comprensión, la reflexión, la acción y la comunicación, aspectos claves que el alumno deberá poner en juego en su proceso de aprendizaje. La estrategia de enseñanza y de aprendizaje de la asignatura Computación se focaliza en la adopción del proceso de resolución de problemas de Polya

[Polya, 1945] como sustento para la construcción de programas, es decir, la resolución de problemas con la computadora. Se representa dicho proceso, en el contexto de la creación de programas, por medio de un Mapa Conceptual [Novack, Gowin, y Otero, 2002]. El Mapa Conceptual, utilizado iterativamente durante el desarrollo del curso, a manera de ciclo, es el instrumento que, por una parte, permite adoptar el método científico como estrategia para el desarrollo del contenido (el qué y el cómo enseñar y aprender) en forma evolutiva e incremental [Jiménez Rey, 2005] y, por otra parte, es la herramienta estratégica educativa que permite implementar el proceso de aprendizaje humano (modelo de la memoria de dos almacenes) [Shunk, 1997], facilitando al estudiante la comprensión del conocimiento conceptual, procedimental y operativo.

La concepción del Docente es que un programa se construye siguiendo un proceso evolutivo (Modelo de Polya de las Cuatro Fases para la Resolución de Problemas) que abarca cuatro fases:

Análisis: El estudiante debe comprender en qué consiste el problema a resolver con la construcción del programa. Debe ser capaz de interpretar el enunciado del problema a resolver con la computadora, es decir, identificar y definir el objetivo del programa.

Diseño: El estudiante debe idear un plan para construir el programa. Debe ser capaz de entender cuál es la naturaleza del problema a resolver, es decir, identificar los recursos necesarios a utilizar y el procedimiento de transformación de esos recursos que conduce al alcance del objetivo.

Codificación: El estudiante debe implementar el plan para construir el programa, o sea, representar los recursos y el procedimiento necesarios en un lenguaje de programación para que la computadora pueda resolver el problema. Debe ser capaz de comprender la semántica del lenguaje Pascal y expresar el algoritmo en la sintaxis correspondiente, es decir, expresar la solución del problema como programa concreto verificable.

Evaluación: El estudiante debe comprobar que el programa diseñado para resolver el problema especificado en el enunciado es una solución ejecutable y correcta. Debe ser capaz de diseñar los juegos de datos de prueba representativos del problema para que la ejecución del programa conduzca al resultado deseado en todas las situaciones posibles del problema, independientemente de los datos particulares de una ejecución (eliminar y depurar errores). La explotación de información ha sido señalada como una manera efectiva de descubrir nuevo conocimiento desde conjuntos de datos de procesos educacionales, datos generados por sistemas de aprendizaje o experimentos de aprendizaje, así como la información descubierta puede ser usada para probar la personalización y la adaptación [Britos, 2008]. Entre los problemas interesantes que la

explotación de información puede ayudar a resolver: determinación de cuáles son los estilos o estrategias de aprendizaje comunes [Britos *et al.*, 2008a], predicción del conocimiento e intereses de un usuario basado en su comportamiento previo, particionamiento de un grupo heterogéneo de usuarios en clusters homogéneos o detección de conceptos no comprendidos en procesos de aprendizaje [Jiménez Rey *et al.*, 2008; Britos *et al.*, 2008c].

Una de las técnicas más comunes de explotación de información son los árboles de decisión (TDIDT), usados para descubrir conocimiento en forma de reglas, las cuales constituyen un modelo que representa el dominio del conocimiento subyacente a los ejemplos disponibles sobre el mismo. Una red bayesiana es un gráfico acíclico direccionado en el cual cada nodo representa una variable y cada arco representa una dependencia probabilística, la cual especifica la probabilidad condicional de cada variable dada a sus padres; la variable a la cual el arco apunta es dependiente (causa – efecto) de la variable en el origen de ésta [Britos *et al.*, 2008b].

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Las líneas de investigación abarcan los siguientes ejes temáticos:

- Enseñanza de Computación en Carreras de Ingeniería
- Enseñanza para la Comprensión
- Aprendizaje basado en Problemas
- Redes Conceptuales
- Teoría Triárquica de la Inteligencia
- Análisis del Error
- Nuevas Tecnologías de la Información y de la Comunicación en la Enseñanza de Algoritmia y Programación
- Educación Basada en Web
- Aplicaciones de Explotación de Información basada en Sistemas Inteligentes a Educación

3. RESULTADOS OBTENIDOS/ESPERADOS

Se ha identificado y clasificado, una lista de componentes de conceptos no comprendidos del modelo del estudiante a ser evaluado, teniendo en consideración tres aspectos fundamentales en la creación de programas: Metodología de Desarrollo, Funcionalidad del Programa y Calidad del Diseño.

Para descubrir conceptos mal comprendidos de programación de los estudiantes, se focalizó la investigación en el uso de herramientas de bases de datos de sistemas inteligentes: inducción de reglas por algoritmos TDIDT y redes bayesianas. Estas reglas se utilizaron en un proceso de descubrimiento del conocimiento en tres pasos:

- Construcción de una base de datos sobre una caracterización estándar de cada estudiante y su estilo y conceptos mal aprendidos de programación.

- Descubrimiento de reglas (mediante algoritmos TDIDT) las cuales establecieron una relación entre conceptos no comprendidos de programación y sus posibles causas.

- Descubrimiento del peso que cada causa tiene sobre cada concepto mal aprendido (mediante redes bayesianas).

En la actualidad, con el objetivo a preparar una metodología escalable, se está revisando los modelos de representación de precedencia de aprendizajes significativos, previéndose realizar un “tuning” del procedimiento experimental seguido para las pruebas de concepto.

4. FORMACIÓN DE RECURSOS HUMANOS

En el marco de este proyecto se está desarrollando una tesis de maestría y dos trabajos de especialidad radicadas en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires.

5. BIBLIOGRAFÍA

- Britos, P. (2008). *Procesos de Explotación de Información Basados en Sistemas Inteligentes*. Tesis de Doctorado en Ciencias Informáticas. Facultad de Informática. UNLP.
- Britos, P., Cataldi, Z., Sierra, E., García-Martínez, R. (2008a). *Pedagogical Protocols Selection Automatic Assistance*. LNAI, 5027: 331-336.
- Britos, P., Felgaer, P., García-Martínez, R. (2008b). *Bayesian Networks Optimization Based on Induction Learning Techniques*. IFIP Series, 276: 439-443.
- Britos, P., Jiménez Rey, E., Rodríguez, D., García-Martínez, R. (2008c). *Work in Progress: Programming Misunderstandings Discovering Process Based On Intelligent Data Mining Tools*. Proceedings 38th ASEE/IEEE Frontiers in Education Conference. Session F4H: Assessing and Understanding Student Learning. ISBN 978-1-4244-1970-8.
- García Martínez, R. y Britos, P. (2004). *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4.
- Jiménez Rey, E. (2005). *Un Enfoque Procedimental para la Enseñanza de Computación en Carreras de Ingeniería*. Proceedings JEITICs 2005: 35-39.
- Jiménez Rey, E., Rodríguez, D., Britos, P., García-Martínez, R. (2008). *Identificación de Problemas de Aprendizaje de Programación con Explotación de Información*. Proceedings del XIV CACIC 2008. Artículo 1881. ISBN 978-987-24611-0-2.
- Joyce, B. y Weil, M. (2002). *Modelos de Enseñanza*. Gedisa. ISBN 84-7432-780-6.
- Litwin, E. (2008). *El Oficio de Enseñar, Condiciones y Contextos*. Paidós. ISBN 978-950-12-1513-7.
- Polya, G. (1945). *How to Solve It: A New Aspect of Mathematical Method*. Princeton. ISBN 0-691-08097-6.

HERRAMIENTAS DIDÁCTICAS PARA EL DICTADO DE INFORMÁTICA TEÓRICA

Ing. María Alejandra Paz Menvielle/ **Facultad Regional Córdoba/ Universidad Tecnológica Nacional**
Ing. Marcelo Marciszack / **Facultad Regional Córdoba/ Universidad Tecnológica Nacional**
Ramiro / **Facultad Regional Córdoba/ Universidad Tecnológica Nacional**
Marina Cardenas / **Facultad Regional Córdoba/ Universidad Tecnológica Nacional**

CONTEXTO

El presente proyecto es la continuación (fase II) de otro anterior y se encuentra consolidado dentro de la línea de investigación que busca mejorar las herramientas didácticas de que disponen los docentes de la asignatura Sintaxis y Semántica del Lenguaje en el Dpto. de Sistemas de Información de la UTN FRC, para el dictado de la misma y que además brinde al alumno un recurso didáctico importante y único para lograr alcanzar el aprendizaje significativo deseado por la Institución.

RESUMEN

Los contenidos de Informática Teórica en carreras de informática no deberían presentar dificultades en cuanto a su transposición didáctica, ya que constituyen, el núcleo central de estudio de su disciplina. Para que los estudiantes logren un aprendizaje significativo de los conceptos de Informática Teórica, se ha elaborado como estrategia la creación de un conjunto integrado de herramientas didácticas accesibles desde la web, las cuales son utilizadas en los procesos de enseñanza, aprendizaje, posibilitando de esta manera que los estudiantes fijen los contenidos curriculares de teoría de autómatas, gramáticas formales, traducción de lenguajes y compiladores, realizando prácticas efectivas de los mismos, permitiendo además que sea el propio alumno que evalúe el progreso de su aprendizaje mediante el uso de un test de autoevaluación.

Por lo tanto, la estrategia es que el estudiante disponga de un conjunto integrado de herramientas y materiales de estudio, en ambiente web, sentando bases sólidas de conocimientos teóricos, con una gran ejercitación y aplicabilidad de los mismos, permitiendo autoevaluar sus conocimientos teóricos y prácticos a través de simuladores y test de autoevaluación.

Palabras Clave: Informática Teórica, Máquinas Abstractas, Gramáticas Formales, Autómatas, Estrategias didácticas

1. INTRODUCCION

Los contenidos de Informática Teórica, incluidos en el núcleo común de las carreras con informática tales como lenguajes y gramáticas formales, teoría de autómatas, traducción de lenguajes, máquinas computadoras teóricas, y compiladores son difíciles de relacionar por los estudiantes, por su naturaleza matemática y fuertemente abstracta.

Además los estudiantes efectúan prácticas de diseño de máquinas abstractas, sin contar con herramientas para la experimentación en computadora y exploración de contenidos, que les permita en forma clara aplicarlos en la resolución de problemas reales. Esta dificultad manifiesta de relacionar la teoría y sus posibles aplicaciones, se ha advertido durante las evaluaciones de los mismos.

Para poder asimilar en forma completa estos contenidos, consideramos que se deben realizar prácticas efectivas de los mismos utilizando simuladores de máquinas abstractas, estudiando su funcionamiento y

los algoritmos de las herramientas respectivas. Es necesario además que el educando pueda explorar estos conocimientos y además que, mediante un simulador de evaluaciones de contenidos teóricos (el que incluye la generación de un instrumento de evaluación automatizado que basado en un repositorio de preguntas y respuestas en ambiente web), le sea posible generar en forma aleatoria un test de autoevaluación. La importancia de una completa asimilación de estos contenidos, radica en que los mismos no sólo tienen efectiva aplicabilidad en la construcción de compiladores, sino en temas tan diversos dentro de la disciplina informática como son la ingeniería de requerimientos o programación de dispositivos móviles y embebidos.

La dificultad manifiesta de relacionar la teoría y sus posibles aplicaciones, se ha advertido durante el dictado de los contenidos y se pone en evidencia en el proceso de evaluación.

Existe una gran cantidad de bibliografía y herramientas de software para el estudio de estos temas, pero las mismas los abordan en forma parcializada, sin integración adecuada, con diferentes nomenclaturas y con definiciones dispares, por lo que no resultan didácticamente útiles para llevar adelante el proceso de enseñanza-aprendizaje, sin permitir la correcta transposición de los contenidos teóricos en aplicaciones de la realidad.

En este proyecto, se busca además, lograr profundo conocimiento teórico y práctico, de las técnicas y herramientas conceptuales de la Informática Teórica relacionadas con la construcción de compiladores, e involucrar a alumnos en el estudio y la investigación de temas fundamentales de su carrera.

Los aspectos teóricos, para el abordaje de la temática planteada, están muy difundidos y son bastante generalizados: sobre los contenidos de informática Teórica, es completo lo publicado en [1] [2] [3] [4] [5], pero los mismos no se ajustan en lo didáctico, para ser transferido en forma directa a los alumnos en una carrera de grado en informática.

Siguiendo los lineamientos de un proyecto anterior a éste, tendrá como misión fundamental

continuar lo realizado en la fase previa a través de la revisión, reformulación y ampliación el conjunto de herramientas desarrolladas, valiéndonos de los resultados obtenidos y de los cambios introducidos en la Cátedra, para mejorarlas y así propender a lograr en los alumnos aprendizaje significativo. Se utilizarán técnicas y metodologías actuales que posibiliten a los miembros del proyecto incrementar sus habilidades en diseño, desarrollo y pruebas de los productos construidos, basándose en la bibliografía mencionada en [6] [7] [8].

En resumen, en este trabajo se propone construir un conjunto de herramientas disponibles para ejecutar desde la web “simuladores” que permitan al alumno realizar una práctica efectiva, utilizando estos simuladores de máquinas abstractas, estudiando y comprendiendo el funcionamiento de los programas fuente de los mismos, y en todo momento poder autoevaluar sus aprendizajes.

2. LINEAS DE INVESTIGACION y DESARROLLO

El presente proyecto se encuentra consolidado dentro de la línea de investigación que comenzó con la fase I desarrollado en el Dpto. de Sistemas de Información de la UTN FRC, el cual tuvo como objetivo la construcción del sitio web en su etapa inicial y el desarrollo de los primeros simuladores. En esta fase II se está perfeccionando el sitio y se están desarrollando mayores funcionalidades y nuevos simuladores.

3. RESULTADOS OBTENIDOS/ESPERADOS

Para la realización de este trabajo se utilizó el método científico, con la aplicación de un desarrollo experimental, el cuál comenzó con la recolección del material sobre esta temática que se encuentra publicado, continuando con la selección y clasificación del mismo.

Se procedió a la unificación de la nomenclatura y simbología utilizada.

Se estableció la estrategia didáctica para abordar los contenidos de manera de facilitar el aprendizaje de los alumnos incluyendo la

estructura del sitio web se incluyen todos los contenidos teóricos, ejercitación práctica y las herramientas de simulación para que los alumnos realicen las prácticas, y la evaluación de sus aprendizajes.

Detallaremos a continuación las características esenciales de los productos construidos:

Sitio Web

A través de él se puede acceder en forma integrada a todos los contenidos teóricos, conjuntamente con la correspondiente ejercitación práctica con ejercicios resueltos y a resolver por los alumnos. Conjuntamente con los links para acceder a los simuladores y a los tests de evaluación.

Pantalla de acceso Sitio Web



El sitio web puede ser accedido desde la dirección www.institucional.frc.utn.edu.ar/sistemas/ghd

Herramientas de Simulación

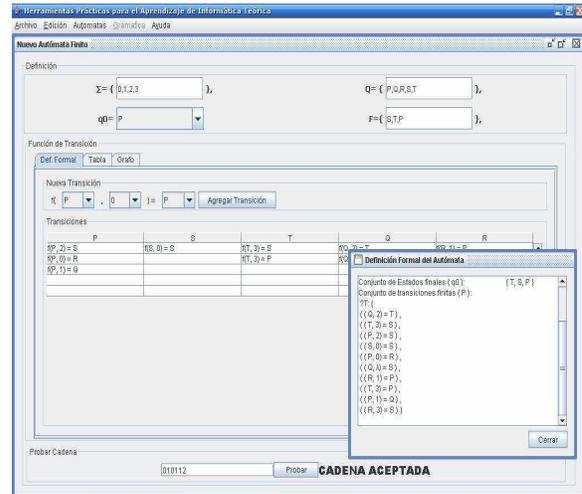
Para la ejercitación práctica de los contenidos, se desarrollaron las siguientes herramientas:

- Generadores de gramáticas formales con su correspondiente ejercitación práctica.
- Simuladores de autómatas finitos, con pila, linealmente acotados y máquina de Turing, para que los alumnos evalúen las soluciones a los problemas planteados en las clases prácticas.

- Generador de analizadores léxicos a partir de una gramática regular que muestre la generación automática de código desde una especificación formal y las distintas formas de tratar con el no determinismo del autómata finito obtenido: conversión a autómata finito determinista, búsqueda en árboles generados por estados posibles y procesamiento en paralelo de los distintos caminos deterministas generados por el no determinismo del autómata.

La siguiente es una pantalla que muestra uno de los simuladores (Autómatas Finitos), con los que los alumnos pueden realizar simulaciones y ejercitar.

Simulador de Autómatas Finitos



Test de Autoevaluación

Si bien existen un gran cantidad de herramientas para confeccionar cuestionarios de evaluación de aprendizajes (incluidas en plataformas para e-learning, incluso gran cantidad de ellas de software libre), se construyó una herramienta de manera de poder integrarla al sitio y que cumpliera con funcionalidades adicionales como el de poder generar en forma aleatoria el instrumento de evaluación con las siguientes particularidades: definir cantidad de preguntas por cada unidad temática y sub-temas a incluir en el cuestionario de evaluación, selección aleatoria de los diferentes enunciados para una misma pregunta, selección aleatoria de opciones

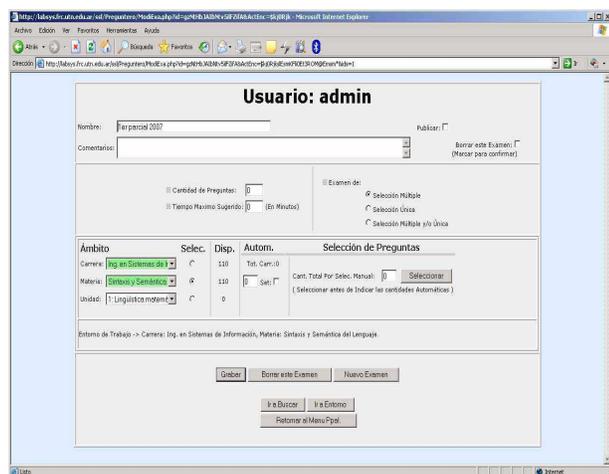
correctas e incorrectas dentro de un conjunto de opciones dentro de cada pregunta, cantidad de opciones correctas que puede tener una pregunta y permitir valorizaciones diferenciadas para cada pregunta.

Esto permite entre otras cosas la imposibilidad de que un instrumento resulte igual a otro ya practicado por el alumno.

A continuación se muestran algunas pantallas generadas por la herramienta de autoevaluación.

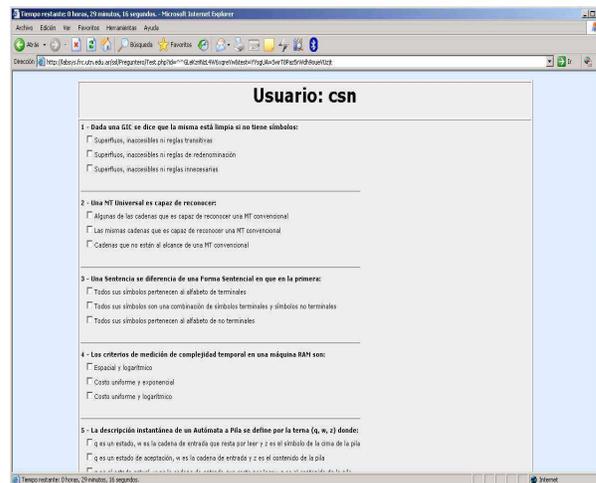
Para la Generación del Instrumento

En esta pantalla, el Docente, luego de haber ingresado el conjunto de preguntas, con sus diferentes enunciados y opciones de respuestas correctas e incorrectas, puede configurar diversos parámetros, tales como: selección de contenidos de las unidades temáticas y dentro de estas los sub-temas a incluir en el cuestionario en forma conjunta con la cantidad de preguntas, cantidad de opciones por preguntas, y opciones de valorización de las mismas.



Instrumento generado

A continuación se muestra parcialmente la visualización de un cuestionario generado por la herramienta de autoevaluación en donde se evidencia la simplicidad para responder las preguntas por parte del alumno.



Implementación

Estas herramientas, aunque aún en proceso de desarrollo y continuo agregado de nuevas funcionalidades, durante el ciclo 2008 han sido puestas en práctica sobre un conjunto de alumnos, mas precisamente en la comisión 2K4 en el dictado de la asignatura Sintaxis y Semántica del Lenguaje de la carrera de grado de Ingeniería en Sistemas de Información en la Facultad Regional Córdoba de la Universidad Tecnológica Nacional, estando disponibles desde un primer momento los contenidos teóricos, la ejercitación práctica, las herramientas construidas, y el test de autoevaluación. El sitio web, alojado en la intranet de la Facultad Regional Córdoba, a modo de prueba piloto para este conjunto de alumnos, se hará efectivo para todas las comisiones en ciclos lectivos posteriores.

Si bien, no ha estado completamente implementado al inicio del ciclo lectivo, ya que durante el desarrollo del dictado de la asignatura se han introducido cambios y nuevas funcionalidades, se evidencia que el resultado obtenido ha sido satisfactorio, desde las siguientes dimensiones de análisis: la primera, que resulta subjetiva ya que no existen parámetros que de medición que la justifiquen, es que se ha reflejado con respecto a ciclos anteriores que los alumnos se encuentran mas motivados al poder verificar con los simuladores sus ejercitaciones, y la segunda que si es verificable, ya que aplicar la misma metodología

de evaluación, durante el mismo período y sobre idénticos contenidos, al efectuar una comparación de los rendimientos académicos con ciclos lectivos anteriores, el promedio de las calificaciones obtenidas por los estudiantes se ha visto incrementado en algo superior a un punto en una escala de diez.

4. FORMACION DE RECURSOS HUMANOS

La realización de este proyecto sirve a la Cátedra de Sintaxis y Semántica del Lenguaje para impartir los conocimientos de la dicha asignatura.

El producto desarrollado está disponible en la web para que cualquier alumno lo acceda, el producto se instalará en las máquinas de un laboratorio de la facultad a los fines de ser utilizado por alumnos de la cátedra de Sintaxis y Semántica del Lenguaje para mostrarles su funcionamiento, enseñarles a construir los diferentes autómatas y que realicen los prácticos asignados por la cátedra

Esta línea de investigación forma parte del proyecto mencionado en la sección 2 el cual también se encuentra enmarcado dentro del ámbito de la temática adoptada por uno de sus integrantes como parte de su tesis de maestría en ingeniería de software y como parte de un informe técnico de investigación de su carrera de doctorado.

Además, se ha incorporado como parte integrante del proyecto, una beca BINIT para fomentar el acercamiento de jóvenes graduados a las actividades de investigación.

Por otra parte, también participan del proyecto, alumnos del último nivel la carrera de Ingeniería en Sistemas de Información de la UTN FRC, próximos a recibirse con lo cual, uno de los objetivos del proyecto es el contribuir que puedan hacer su Práctica Profesional Supervidada.

9. BIBLIOGRAFIA

- [1] J. E. Hopcroft / Rajeev Motwani / J.D. Ullman (2008), Teoría de Autómatas, Lenguajes y Computación, PEARSON Addison-Wesley
- [2] Kenneth Loudon (2004), Lenguajes de Programación, Thomson Internacional
- [3] Pascual J. Iranzo (2005), Lógica Simbólica para Informáticos, Alfaomega Group Editor
- [4] John C. Martin (2004), Lenguajes Formales y Teoría de la Computación, McGraw-Hill
- [5] P. Isasi / P. Martínez / D. Borrajo (1997), Lenguajes, Gramáticas y Autómatas, Addison-Wesley
- [6] Borsotti Carlos (2006), Temas de Metodología de la investigación, Ed MNO Y DAVILA
- [7] Pressman Roger (2005), Ingeniería del software: un enfoque práctico, McGraw Hill
- [8] Norma ISO 9001:2000, Calidad en el desarrollo de software.

Modelos predictivos y técnicas de minería de datos para la identificación de factores asociados al rendimiento académico de alumnos universitarios

Porcel, Eduardo; Dapozo, Gladys; López, María V.

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura
Universidad Nacional del Nordeste. 9 de Julio N° 1449. CP 3400. Corrientes. Argentina.

TE: (03783) 423126 - (03783) 473930 Fax)

{gndapozo, eporcel, mvlopez}@exa.unne.edu.ar

CONTEXTO

Las líneas de I/D presentadas en este trabajo forman parte de las actividades definidas en el marco del proyecto F008-2008: "Rendimiento Académico de alumnos de la FACENA – UNNE: Su análisis mediante métodos cuantitativos", acreditado por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste. El mencionado proyecto tiene como objetivo fundamental construir modelos predictivos del rendimiento académico de los alumnos de las carreras de la Facultad de Ciencias Exactas y Naturales y Agrimensura (FACENA). A tal fin, puede ser considerado una extensión del proyecto de investigación (PI 005/06) anteriormente desarrollado por el mismo grupo de investigación.

Los resultados obtenidos en el mencionado proyecto, han permitido determinar, principalmente, la estrecha vinculación que existe entre el rendimiento académico de los alumnos del primer año de todas las carreras con el nivel de conocimientos matemáticos previos y con las condiciones socioeconómicas de los mismos. Así también, para las carreras de formación docente y licenciaturas en ciencias básicas, se pudo observar que el escaso nivel de avance en los estudios y las prolongadas estadías de estos alumnos en el sistema guardan relación con la estructura curricular, fundamentalmente con el ordenamiento de los contenidos de enseñanza (correlatividades).

Sin embargo, por la permanente mutación que sufren los factores mencionados (fluctuaciones en las condiciones sociales y económicas de los alumnos, desarrollo de programas de mejora de la calidad de la enseñanza en los niveles educativos previos, reformas de los planes de estudios, entre otros), es factible pensar que las variables correspondientes a estas dimensiones, no permanezcan estables con el devenir del tiempo y, por lo tanto puedan ser reconstruidas las relaciones y tipologías obtenidas.

De acuerdo a lo expresado anteriormente, en las líneas de investigación presentadas en este trabajo, se ha puesto énfasis en la construcción de modelos matemáticos que permitan predecir el rendimiento académico futuro de los estudiantes, tomando como base la información de las cohortes 2001 – 2008. Esta predicción permitirá conocer el rendimiento

académico del alumno a priori con sólo disponer de la información de los mismos referida a sus condiciones iniciales (socioeconómicas y/o de conocimientos matemáticos previos), detectar con anticipación cuáles son las acciones pertinentes para contribuir a que los estudiantes superen los obstáculos que actualmente les impide avanzar en sus estudios y finalizarlos en menos tiempo que el que hoy día emplean.

RESUMEN

Este proyecto tiene por objetivo construir modelos predictivos del rendimiento académico de los estudiantes de las diversas carreras de la FACENA de la UNNE. Las variables a incorporar en los modelos serán seleccionadas de acuerdo a los resultados obtenidos a partir de los siguientes análisis: a) Resultados del test de diagnóstico de conocimientos matemáticos previos; b) Condiciones socioeconómicas de los alumnos de las distintas carreras y datos obtenidos de encuesta directa a los alumnos de primer año. Para la formulación y ajustes de los modelos de predicción, se utilizarán alternativamente, técnicas de minería de datos clásicas y métodos simbólicos o inteligentes, evaluando su desempeño en la predicción del rendimiento académico de los alumnos. Los resultados obtenidos a partir del desarrollo de este proyecto, constituirán un aporte significativo para los procesos de evaluación y acreditación universitarios, considerando que la reflexión sobre todos los elementos proporcionados por el análisis del rendimiento del alumnado contribuirá a la mejora de la calidad del sistema educativo.

Palabras clave: Rendimiento académico. Educación Superior. Minería de datos. Métodos simbólicos. Métodos estadísticos.

1. INTRODUCCION

A partir de la década del '80 surge en las universidades de todo el mundo la preocupación por la calidad del servicio educativo que prestan. Esto dio lugar a procesos de evaluación a fin de detectar las debilidades y fortalezas institucionales y generar acciones correctivas de las deficiencias encontradas. En nuestro país, en la década del '90, el Estado Nacional incluye en su agenda de política educativa

la evaluación de la calidad del accionar universitario, y la mayoría de las universidades nacionales inician procesos de evaluación institucional.

En 1996, se conocen los primeros resultados referidos al rendimiento académico de los estudiantes de las trece carreras que por entonces podían cursarse en la FACENA. Dicha información hace referencia a elevados índices de desgranamiento en todos los años de estudios pero, fundamentalmente, al término del primer cuatrimestre del primer año de estudios. Asimismo, da cuentas de que el retraso promedio en el egreso de todas las carreras alcanza al 50% de la duración teórica de las mismas, llegando en algunas a superarlo.

Ahora bien, ¿a qué se hace referencia con el término rendimiento académico?

El rendimiento académico es un claro indicador del avance exitoso en la carrera de estudios de algún alumno en un momento particular, y a su vez también es un pronosticador de la posibilidad de completar exitosamente dicha carrera de estudios.

El término "rendimiento" tiene muchas implicancias, principalmente si se considera a las notas obtenidas por los alumnos como el referente casi exclusivo. Esta información puede generar, incluso una lectura ingenua, que centra sólo la responsabilidad académica en el alumno. Sin embargo, la responsabilidad institucional es clave para evaluar lo que se entiende por rendimiento. Más allá de las condiciones internas a las instituciones y de las prácticas docentes, resulta imprescindible también conocer las características que aportan quienes son los receptores de la labor docente. Esta información puede contribuir a estimar algunas de las razones que inciden en el rendimiento y la deserción de los alumnos universitarios (Toer, 2000).

Debe tenerse en cuenta que se trata de un constructo teórico complejo y multidimensional, atravesado y determinado por múltiples factores sociales, económicos, históricos, institucionales e individuales. Por tal motivo el rendimiento académico ha sido representado de diferentes maneras en los diversos estudios que han abordado el tema. En algunos, está representado sólo por el número de materias aprobadas por un alumno en una carrera, en otros por los resultados de tests específicamente diseñados o el promedio de notas de las asignaturas cursadas. Esta variedad de interpretaciones del concepto de rendimiento académico está ligada a las particularidades de las investigaciones en cuestión, principalmente al momento histórico en que se realiza la investigación y las concepciones de quienes llevan a cabo y financian la misma. Restringir el concepto a uno solo de estos indicadores, supone una postura ingenua y hasta simplista de lo que el acto educativo significa, pues equivale a descontextualizarlo, aislándolo de la realidad social e histórica de la que

forma parte. En una mirada contextualizada, el rendimiento académico es el producto de condiciones institucionales (diseño curricular, práctica docente, valores y concepciones institucionales, etc.), socioeconómicas (situación laboral, estado civil, nivel educativo del grupo familiar, entre otras) e individuales (formación previa, hábitos de estudios, etc.) de los estudiantes.

Para la construcción de modelos de predicción del rendimiento académico de los estudiantes, se utilizará información referida a la situación socioeconómica y el desempeño académico en el período 2001- 2008 de los estudiantes de las diversas carreras de FACENA. Para el análisis de esta información se utilizarán técnicas de minería de datos.

La minería de datos se define como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en diferentes formatos. El objetivo es encontrar modelos inteligibles a partir de los datos, descubrir patrones cuya utilización apoye decisiones que reporten beneficios a la organización (Hernández Orallo et al, 2004).

Para cumplir sus objetivos, son dos los retos de la minería de datos: por un lado, trabajar con grandes volúmenes de datos, que proceden generalmente de sistemas de información, con los problemas que ello conlleva (ruido, datos ausentes, intratabilidad, volatilidad de los datos, etc.), y por el otro, usar técnicas adecuadas para analizar los mismos y extraer conocimiento novedoso y útil. En muchos casos, la utilidad del conocimiento extraído está relacionada con la comprensibilidad del modelo inferido (Hernández Orallo et al, 2004).

Esta tecnología emergente combina el análisis estadísticos y la gestión de las bases de datos para extraer información desde los datos, y se presenta como un campo multidisciplinar que se ha desarrollado en paralelo o como prolongación de otras tecnologías (Thuraisingham, 2000).

Las técnicas que conforman el campo de la Minería de Datos buscan descubrir, en forma automática, el conocimiento contenido en la información almacenada en las bases de datos de las organizaciones. Por medio del análisis de datos, se pretende descubrir patrones, perfiles y tendencias. Es importante que estas técnicas sean las adecuadas al problema abordado. En este sentido, se pueden establecer dos grandes grupos de técnicas ó métodos analíticos: los métodos simbólicos y los métodos estadísticos (Britos, 2005).

Entre los métodos simbólicos se incluyen a las Redes Neuronales, Algoritmos Genéticos, Reglas de Asociación, Lógica Difusa, entre otros. Estos derivan del campo de la Inteligencia Artificial.

Los métodos estadísticos están constituidos por las técnicas del Análisis Multivariante de Datos, tales como Regresión Lineal simple y Múltiple, Regresión

No Lineal, Regresión Logística, Análisis Discriminante, Árboles de Regresión, entre otras. Las técnicas de esta categoría, de alguna manera, constituyen la piedra basal de la Minería de Datos (Britos, 2005).

En este estudio se utilizarán las siguientes metodologías para analizar las variables socioeconómicas relacionadas con el rendimiento académico: La Regresión Logística, los Árboles de Decisión y las Redes Neuronales.

El modelo de Regresión Logística es un método lineal que intenta modelizar la probabilidad de ocurrencia de un evento de interés. La variable dependiente es categórica dicotómica o policotómica, a los efectos de facilitar la interpretación (Britos, 2005).

Es una técnica adecuada cuando se pretende hacer una clasificación basada en las características de los datos. Una ventaja adicional de esta técnica es que no requiere la normalidad estricta de los datos, además muchos estudios han evidenciado otras características que hacen de la regresión logística una buena herramienta para la categorización (García Jiménez et al, 2000).

Las Redes Neuronales son modelos computacionales inspirados en las características neurofisiológicas del cerebro humano y están formadas por un gran número de neuronas dispuestas en varias capas e interconectadas entre sí mediante conexiones con pesos. Una neurona sobre un conjunto de nodos N es una tripleta (X, f, Y), donde X es un subconjunto de N, Y es un único nodo de N y f es una función neuronal que calcula un valor de salida para Y basado en una combinación de los valores de los

$$y = f\left(\sum_{x_i \in N} w_i x_i\right)$$

componentes de X, es decir w_i . Los pesos w_i pueden ser positivos o negativos, reproduciendo el carácter excitador o inhibitorio de la sinapsis de las neuronas. Las redes neuronales usan un proceso de aprendizaje por analogía donde los pesos de las conexiones son ajustados para reproducir un conjunto de datos representativo del problema a aprender. Las redes neuronales constituyen herramientas analíticas que permiten examinar los datos con el objeto de descubrir y modelar las relaciones funcionales existentes entre las variables. Pueden comportarse como técnicas de aproximación o de clasificación universales (Castillo et al, 1999).

Como antecedentes de aplicación de la técnica de redes neuronales en el ámbito de educación pueden mencionarse los trabajos de González (1999), Salgueiro et al (2006), Borracci y Arribalzaga (2005).

Los árboles de decisión son una serie de decisiones o condiciones organizadas de forma jerárquica, a modo de árbol. Son muy útiles para encontrar estructuras en espacios de alta dimensionalidad y en

problemas que mezclan datos categóricos y numéricos.

Básicamente, un árbol de decisión es un árbol donde cada nodo representa una condición o test sobre algún atributo y cada rama que parte de ese nodo corresponde a un posible valor para ese atributo. Finalmente, las hojas representan el valor de la variable predicha. Esta técnica se usa en tareas de clasificación, agrupamiento y regresión. Los árboles de decisión que se usan para predecir variables categóricas se llaman árboles de clasificación, mientras que los árboles de decisión que se utilizan para predecir variables continuas se llaman árboles de regresión (Alcover et al, 2007).

Como antecedente de aplicación de la técnica de árboles de regresión aplicado el rendimiento de alumnos universitarios puede mencionarse a Bacallao Gallestey et al (2004).

2. LINEAS DE INVESTIGACION y DESARROLLO

- a) **Preprocesamiento de los datos: Actualización de la base de datos oportunamente diseñada dentro del marco del proyecto PI 005/06, con los datos socioeconómicos y del estado académico de los alumnos de las carreras de FACENA a diciembre de 2008**, a fin de proveer la información que constituye el soporte de las actividades del proyecto. Esto implica incorporar los datos del estado académico de los alumnos, que provee el Departamento Estudios de la FACENA, y los datos que surgen del formulario de ingreso que completan los alumnos al momento de ingresar a la UNNE.
- b) **Modelado de datos y predicción del rendimiento académico en el primer cuatrimestre del primer año de estudios en función del nivel de conocimientos matemáticos previos de los ingresantes a la FACENA:** Los datos del diagnóstico se utilizarán como variables explicativas en un modelo que permita predecir el rendimiento de los alumnos al finalizar el primer cuatrimestre, empleando métodos simbólicos y estadísticos.
- c) **Formulación y ajuste de modelos para predecir el rendimiento académico de los alumnos en función de las características socioeconómicas de los mismos**, empleando métodos estadísticos y simbólicos. Numerosas investigaciones han encontrado vinculación entre las condiciones socioeconómicas y personales de un individuo (tales como, edad, sexo, lugar de procedencia, etc.), y su rendimiento académico. Para el análisis de esta relación se requiere, previamente, una etapa de preprocesamiento de los datos, que comprende los siguientes pasos: Integración, Reconocimiento y Limpieza, Transformación y Reducción. Esta etapa tiene por objeto mejorar

la calidad de los datos, teniendo en cuenta que, a lo largo del período de estudio, se ha modificado, en más de una oportunidad, el diseño del formulario de ingreso, situación que exige un análisis detallado para la determinación de equivalencias entre los distintos valores de las variables en estudio.

Para la construcción y ajuste de los modelos de predicción del rendimiento académico de los alumnos se utilizarán métodos de minería de datos simbólicos y estadísticos, previéndose realizar un estudio comparativo entre ambos grupos de metodologías, con el objeto de contrastar el desempeño y la eficiencia de las mismas en el problema de la predicción del rendimiento académico de los estudiantes.

3. RESULTADOS OBTENIDOS/ESPERADOS

Dentro de las líneas de trabajo mencionadas, se han obtenido los siguientes resultados:

- a) Se han estudiado técnicas de preprocesado de datos para mejorar la calidad de la información obtenida desde los sistemas de información existentes, y para mantener actualizado un repositorio con toda la información sistematizada existente en la unidad académica respecto del desempeño de los alumnos (Dapozo et al, 2007).
- b) Se ha analizado el perfil socioeconómico y educativo de los alumnos ingresantes de la FACENA y su relación con su rendimiento académico, medido en términos de su desempeño en la primera asignatura de Matemática, en el primer año de carrera universitaria, utilizando técnicas clásicas de minería de datos (Porcel et al, 2008).
- c) Se ha analizado el rendimiento académico de los alumnos de las trece carreras de grado de la FACENA, utilizando indicadores basados en la relación entre el número de exámenes rendidos y el número de asignaturas aprobadas por los mismos, estimados mediante regresión lineal paramétrica y no paramétrica. Se trazaron gráficos de dispersión para cada carrera que permitieron observar la eficiencia de los alumnos en los exámenes, y se calculó además una matriz de correlación de los indicadores. (Porcel et al, 2009)

Como resultados esperados, se espera poder predecir el rendimiento académico de los estudiantes de FACENA, para lo cual se ha planificado la construcción de modelos cuantitativos predictivos del desempeño estudiantil en base a la información disponible sobre las condiciones socioeconómicas de los estudiantes, a sus conocimientos matemáticos previos y los datos aportados por los mismos

alumnos a través de instrumentos diseñados especialmente.

Los resultados de la investigación aportarán a un mayor conocimiento de los posibles factores que inciden en el desempeño de los alumnos. Esta información permitirá que, desde la gestión institucional, se aborden mecanismos correctivos o superadores que contribuyan al mejoramiento de los índices de desgranamiento, abandono, bajo rendimiento académico y prolongación excesiva de la duración de la carrera, comunes en el ambiente aniversario nacional, principalmente notorios en la actuación de los alumnos en el primer año de carrera.

Por otra parte, en el marco actual de las universidades nacionales, los procesos de evaluación y acreditación de títulos se basan en la construcción de indicadores que permiten descubrir fortalezas y debilidades de la formación universitaria, por lo que “reflexionar sobre todos los elementos que la evaluación del rendimiento del alumnado proporciona se convierte en un mecanismo claro para la mejora de la calidad del proceso educativo” (Muñoz, 2005).

Finalmente, a nivel de las políticas educativas nacionales, la formación en Informática, Ingeniería y Ciencias Básicas, se considera prioritaria. Se destinan recursos especiales como becas, planes de tutorías y otras acciones tendientes a promover titulaciones en estas áreas, así como también, mejorar el índice de graduados y de retención de alumnos. Conocer las causas que subyacen en el rendimiento académico, permitirá mejorar estas iniciativas por el aporte de mayor información.

4. FORMACION DE RECURSOS HUMANOS

En este proyecto se enfatiza el enfoque interdisciplinario dado que sus integrantes proceden de distintas disciplinas: Matemática, Estadística e Informática. Esto permite un abordaje sistémico de los problemas de la investigación, a la vez que se complementan y enriquecen las distintas miradas disciplinares. Los alumnos en proceso de formación de la Licenciatura en Matemática y la Licenciatura en Sistemas de Información que se suman al proyecto en calidad de becarios o para la realización del Trabajo Final de Aplicación, requisito académico de la carrera de Sistemas, tienen la oportunidad de aplicar y ampliar sus conocimientos en estas aplicaciones interdisciplinarias que posibilita la Minería de Datos.

5. BIBLIOGRAFIA

- ALCOVER, R., BENLLOCH J., BLESA, P., CALDUCH, M., CELMA, M., FERRI C., HERNÁNDEZ ORALLO, J., Y OTROS “Análisis del rendimiento académico en los estudios de Informática de la Universidad Politécnica de Valencia aplicando técnicas de Minería de Datos”. XII Jornadas de Enseñanza Universitaria de la Informática 2007. Disponible En <http://bioinfo.uib.es/~joemiro/aenui/procjenui/jen2007/alanal.pdf>
- BACALLAO GALLESTEY C., PARAPAR DE LA RISTRA, J., ROQUE GIL M., BACALLOA GUERRA J. “Arboles de regresión y otras opciones metodológicas aplicadas a la predicción del rendimiento académico”. Revista Cubana de Educación Médica Superior, vol.18, N°3. 2004. Disponible en: http://bvs.sld.cu/revistas/ems/vol18_3_04/ems_02304.htm
- BORRACCI, R. A., ARRIBALZAGA, E. B. “Aplicación de análisis de conglomerados y redes neuronales artificiales para la clasificación y selección de candidatos a residencias médicas”. Educación Médica, Vol 8, N° 1. ISSN 1575-1813. Barcelona. 2005.
- BRITOS, P. V. “Minería de Datos”. Buenos Aires: Nueva Librería. 2005.
- CASTILLO, E., COBO, A., GUTIÉRREZ, J. M., PRUNEDA, R. E. “Introducción a las Redes Funcionales con Aplicaciones. Un Nuevo Paradigma Neuronal”. Editorial Paraninfo S.A. Madrid. España. 1999.
- DAPOZO, G., PORCEL, E., LÓPEZ, M. V.; BOGADO, V. “Técnicas de preprocesamiento para mejorar la calidad de los datos en un estudio de caracterización de ingresantes universitarios”. IX Workshop de Investigadores en Ciencias de la Computación (WICC 2007). Trelew. Chubut. Argentina. 2007.
- GARCÍA JIMÉNEZ, M. V., ALVARADO IZQUIERDO, J. M. y JIMÉNEZ BLANCO, A. “La predicción del rendimiento académico: regresión lineal versus regresión logística”. Psicothema, 12 (2), 248-252. 2000. Disponible en http://redalyc.uaemex.mx/redalyc/pdf/727/727_97059.pdf.
- GONZÁLEZ, D.S. “Detección de alumnos de riesgo y medición de la eficiencia de centros escolares mediante redes neuronales”. Biblioteca de Económicas y Empresariales. Servicios de Internet. Universidad Complutense de Madrid. 1999.
- HERNÁNDEZ ORALLO, J., FERRI RAMÍREZ, C. y RAMÍREZ QUINTANA M. J. “Introducción a la Minería de Datos”. España: Prentice Hall. Pearson Education. 2004.
- MUÑOZ, S. “Indicadores de rendimiento académico del alumnado de la universidad de La Laguna. Jornadas sobre Políticas de Calidad en la Universidad de La Laguna”, 2005.
- PORCEL E. A., DAPOZO GLADYS N., LÓPEZ M. “Técnicas clásicas de minería de datos aplicadas al estudio del rendimiento académico de alumnos de primer año de carreras de la FACENA”. Comunicaciones Científicas y Tecnológicas 2008. Universidad Nacional del Nordeste. Corrientes. Argentina. 2008.
- PORCEL, E., LÓPEZ, M. V., DAPOZO, G., CAPUTO, L. “Relación entre el número de exámenes rendidos y el número de asignaturas aprobadas como indicador del rendimiento académico de alumnos universitarios”. XXII Encuentro Nacional de Docentes de Investigación Operativa (ENDIO). XX Escuela de Perfeccionamiento en Investigación Operativa (EPIO). Buenos Aires. Argentina. 2009.
- SALGUEIRO, F., COSTA, G., CÁNENA, S., LAGE, F., KRAUS, G., FIGUEROA, N., CATALDI, Z. “Redes Neuronales para predecir la aptitud del alumno y sugerir acciones”. VIII Workshop de Investigadores en Ciencias de la Computación (WICC 2006). Buenos Aires. Argentina. 2006.
- THURASINGHAM, B. “A primer for understanding and applying Datamining”. IT Professional, 2 (1), 28-31. 2000.
- TOER, M. “El caso de los ingresantes de 1998 al Ciclo Básico Común de la Universidad de Buenos Aires, para seguir carreras de la Universidad de Derecho, Ciencias Económicas y Ciencias Sociales”. Buenos Aires, Argentina: Instituto de Investigaciones Gino Germani, FCSoc., Ciclo Básico Común, Universidad de Buenos Aires. 2000. Disponible en <http://caraya.cbc.uba.ar/dat/sbe/perfil/perfil.html#1>.

SITIO WEB EDUCATIVO CON ENLACES CONCEPTUALES Y APLICATIVOS DE DIVERSAS MATERIAS DE LAS CARRERAS DE INFORMÁTICA

Ing. José I. Gallardo †
Prof. Titular- Dpto. Informática
Director de P.I. N°686/07- UNPSJB
jgallardo@ing.unp.edu.ar

Ing. Nilda Belcastro
Integrante P.I. N°686/07- UNPSJB
nildab@arnet.com.ar

Ing. Adriana Désima †
Aux. Docente- Dpto. Matemática
Integrante P.I. N°686/07- UNPSJB
desima@unpata.edu.ar

Lic. Ángela Belcastro †
Prof. Adjunta- Dpto. Informática
CoDirectora de P.I. N°686/07- UNPSJB
angelab@ing.unp.edu.ar

Lic. Raúl I. Jáuregui †
Aux. Docente- Dpto. Informática
Integrante P.I. N°686/07- UNPSJB
rjauregui@ing.unp.edu.ar

† **Fac. de Ingeniería - Sede Cro. Rivadavia - Univ. Nac. Patagonia "San Juan Bosco"**
Tel/Fax (0297) 4550836

CONTEXTO:

Esta línea de investigación forma parte del Proyecto de Investigación PI N°686/2007- UNPSJB en ejecución, "Análisis y Desarrollo de un sistema de enlaces conceptuales y aplicativos de contenidos de diversas materias de carreras de Informática, como herramienta de apoyo al aprendizaje del alumno".

RESUMEN:

El objetivo del proyecto es el de mejorar la calidad de enseñanza-aprendizaje, fortaleciendo los enlaces conceptuales y aplicativos de algunas materias de las carreras del Departamento de Informática de la UNPSJB, mediante la construcción y uso de un sistema de apoyo de integración conceptual y aplicada de contenidos del plan de las carreras. Se intenta además, brindar a docentes de diferentes cátedras un sistema genérico de integración, que les permita incorporar elementos aplicativos y conceptuales relevantes, enlazando contenidos de diversas asignaturas, proporcionando luego este sistema a los estudiantes.

El trabajo está orientado a examinar la forma en la que las personas aprenden, relacionando nuevos conceptos con aquellos de los que se han apropiado, asociándolos a los objetivos específicos de aprendizaje de diferentes temas de asignaturas correlativas, identificando además distintos formatos de soporte de información.

Palabras Claves: Tecnología informática aplicada en educación. Software Educativo. Informática. Aprendizaje.

1. INTRODUCCION

Los estudiantes cursan una materia específica que se encuentra dentro del plan de una carrera, en un contexto definido, si no está cursando primer año, han cursado o aprobado materias de años anteriores, y muchas veces desconocen la utilidad de cada una de ellas, o piensan que no todas son indispensables, o que no tienen relaciones entre ellas. Algunas veces se debe a que no disponen de referencias a temas o contenidos de la misma en las demás asignaturas de la carrera. Esta situación se presenta inclusive en casos en los que ya han obtenido el cursado o aprobación de una materia, muchas veces porque las asignaturas pertenecen a distintas áreas, y las palabras esenciales y contenidos mínimos de cada una de ellas no son compartidos.

Por ello, es importante establecer enlaces conceptuales y aplicativos con materias anteriores de la carrera, que permitan al estudiante detectar claramente en qué casos puede aplicar cada contenido mínimo definido en el plan de la carrera y ha sido analizado en otras materias de la misma, de años anteriores o en temas iniciales de asignaturas que se dictan paralelamente.

Con la construcción de un sistema de integración de conceptos y aplicaciones de contenidos de diferentes materias iniciales del plan de las carreras Licenciatura en Informática y Analista Programador Universitario, se intenta fortalecer la integración de contenidos, permitiéndoles a los estudiantes acceder a módulos con software educativo y material didáctico buscando por conceptos o asignaturas, identificando objetivos y/o soluciones propuestas, y bibliografía de base seleccionada.

Como lo analizan los alumnos al considerar el tema “La información en las Organizaciones”, la información se usa una y otra vez y no pierde valor, aumenta la cantidad de veces que una persona refresca un conocimiento del que dispone, y lo usa, y aumenta además la confiabilidad de esta información. Para adquirir un nuevo conocimiento, el sujeto debe poseer una cantidad básica de información respecto al mismo, esquemas cognitivos relacionales y no acumulativos.

Algunas de las actividades que los docentes pueden proponer para concretar una buena enseñanza son las actividades de consolidación, en las que se contrastan las nuevas ideas con las previas de los alumnos, y se aplican nuevos aprendizajes. También las actividades de introducción-motivación, introducen la realidad que se quiere aprender, y las de ampliación que permiten seguir construyendo nuevos conocimientos. Estos distintos tipos de actividades pueden plantearse sobre la base de los contenidos que el estudiante ha examinado al ir avanzando en la carrera.

En segundo término, el equipo del PI pretende construir un sistema genérico, logrando la transferencia del mismo a otras cátedras, llegando a un árbol en el que se enlazan temas a tratar, con materias en los que se examinan, con lista de ejercicios planteados explicativos, y otros propuestos para resolución por parte del alumno.

Los ejes centrales del trabajo se basan en la tecnología educativa y las herramientas de apoyo al aprendizaje de diversos temas de las carreras Analista Programador Universitario y Licenciatura en Informática.

2. LINEAS DE INVESTIGACION Y DESARROLLO

El objetivo principal es el de mejorar la calidad educativa de nivel Universitario, favoreciendo la apropiación de los distintos temas esenciales de carreras de Informática, proporcionando a los estudiantes un sitio con material didáctico y software educativo, que les permitirá resaltar la utilidad de ciertos contenidos que son contenidos mínimos de materias de carreras de Informática.

3. RESULTADOS OBTENIDOS/ ESPERADOS

El producto resultante de este proyecto será de utilidad inicialmente para estudiantes y docentes de materias del Departamento de Informática, y posteriormente también para aquellas cátedras de otras carreras que deseen adoptar el sistema, incorporando los elementos y relaciones de enlace en base a contenidos y aplicaciones del plan, y de la carrera en la que se encuentra inserta dicha asignatura, incluyendo material educativo en cada tema del árbol de enlaces importantes.

ELEMENTOS EN CONSIDERACION:

3.1. SÍNTESIS DE ESPECIFICACIONES DEL SISTEMA GENÉRICO

Las características básicas del sitio al que accederán alumnos y docentes, para brindar aportes con módulos y direcciones Web, ó para bajar módulos con material didáctico de diferentes formatos, se describen en función de los ocho paneles propuestos, a saber:

- Panel de Bienvenida: presenta los objetivos del sitio, Texto del Director y Detalle del PI.
- Panel de Control de Acceso: en función de los roles y funciones de Dirección, Codirección, Administración del sitio, Integrantes del proyecto, Docente externo al proyecto, Alumno, Invitado externo, Docente de otra universidad ó Usuario normal.
- Panel de Árbol de Contenidos: Raíz: Nombre de la carrera, Hojas de Nivel 1: Año y Cuatrimestre, de Nivel 2: Materia y de Nivel 3: Contenidos mínimos y temas de la asignatura.
- Panel de Módulos: presenta un icono asociado, título y breve descripción del módulo, archivo de Módulo y link de descarga, con estimación del tiempo estimado para bajarlo. Si la cátedra lo desea, se incorporará un icono que destaque, para el material específico, que es el propuesto para el estudio del tema, en la asignatura para la cursada actual.
- Panel de Orígenes de los módulos, con sus autores.
- Panel de Recomendaciones de uso.
- Panel de Mensajería, para comunicación entre los distintos actores del proceso, mencionados al citar el panel de control de acceso.
- Panel de Normas del sitio.

3.2. SISTEMA KARMIN (de Karnaugh Minimizer)

El software Karmin es una herramienta interactiva, desarrollado en lenguaje C#, que permite al usuario simplificar funciones booleanas, las cuales pueden ser ingresadas mediante dos métodos distintos:

- Mediante su tabla de verdad
- Mediante un mapa o tabla de Karnaugh

Independientemente del método de ingreso, el software simplificará la función booleana para obtener una posible expresión mínima.

Para tal fin se eligió implementar el algoritmo de **Quine-McCluskey**, método de simplificación de funciones booleanas desarrollado por Willard Van Orman Quine y Edward J. McCluskey, que es funcionalmente idéntico a la utilización del mapa de Karnaugh, pero su forma tabular lo hace más eficiente para su implementación en lenguajes computacionales, y provee un método determinista para conseguir la mínima expresión de una función booleana.

Actualmente se encuentra implementado en un 50%, puede simplificar funciones booleanas a través de su tabla de verdad, independientemente si la misma posee 2 a 4 variables, sin redundancias, y consta de un asistente mediante el cual se pretende mejorar la funcionalidad del mismo. Se prevé la finalización del mismo para fines de agosto del corriente año, para poder ser utilizado en las asignaturas “Elementos de Lógica y Matemática Discreta” y “Arquitectura de los Sistemas de Cómputo”, donde se desarrollan estos temas.

3.3. MATERIAL DIDACTICO EN FORMATO PDF

Se ha preparado material didáctico en formato pdf sobre distintos temas, entre ellos:

- “Estadística Descriptiva: medidas de posición: media aritmética, mediana y modo”, y
- “Estadística Descriptiva: diagrama de cajas y bigotes o box-plot para un grupo de datos”.

Los mismos incluyen la elaboración de material teórico con ejercicios resueltos y propuestos integrados por los siguientes módulos: desarrollo conceptual, guía de conceptos desarrollados en la teoría, bibliografía utilizada en el tema, desarrollo ejemplificativo, que incluye enunciado, objetivos y solución, práctica, y referencias en Internet.

Este material didáctico tiene su enlace aplicativo con materias de informática que plantean actividades de programación, identifican cuál es el proceso, qué entradas se necesitan, en algunos

casos establecen requisitos para el procesamiento de la información, y para la consistencia de entradas. También especifican elementos que integran la interfaz grafica del usuario (IGU).

3.4. SISTEMA SAAF (DE APOYO AL APRENDIZAJE DE AUTÓMATAS FINITOS)

Sistema implementado en Toolbook II, cuyo objetivo es el de asistir al estudiante en el proceso de enseñanza-aprendizaje de interpretación de autómatas finitos, examinados en la asignatura “Fundamentos Teóricos de la Informática”. Permite identificar, recordar y repasar los sistemas binarios y su utilidad en el ámbito informático, vistos en “Introducción a la Computación”, así como asociar ejemplos concretos de autómatas finitos con el lenguaje que reconocen.

Brinda un prototipo, permitiéndole observar el comportamiento de un sistema, y diseñarlo paso a paso, en otro lenguaje, bajo el paradigma orientado a objetos, en Java, al realizar el desarrollo durante la cursada de la materia “Programación Avanzada”.

Se identifican las opciones del menú que implementa cada grupo de trabajo, y variantes explícitas definidas inicialmente, para cada uno de ellos. Por último, permite examinar aspectos claves del tema “Interfaz grafica del usuario”, ejemplificados con el software SAAF.

Este sistema ha sido utilizado en 2008 en la asignatura “Fundamentos Teóricos de la Informática”, al cerrar temas puntuales asociados a autómatas finitos deterministas y no deterministas.

3.5. SISTEMA EduAFin

Este sistema interactivo, implementado en Java, presenta cada ejercitación estructurada en tres módulos, el primero, plantea objetivos, lecturas recomendadas y conocimientos previos, el segundo, presenta un ejercicio interactivo, y el tercero, proporciona elementos adicionales de apoyo al aprendizaje. Se ha utilizado en dos oportunidades en la cursada 2009 de “Fundamentos Teóricos de la Informática”, por el grupo completo de alumnos. La metodología de utilización puede ser durante la clase para el grupo de alumnos, ó como herramienta de entrenamiento individual.

Es una herramienta de apoyo al aprendizaje orientada al estudio de los diagramas de transición de autómatas finitos deterministas. El equipo de proyecto está confeccionando el software educativo de enlace, que se basará en el entrenamiento en el tema “Grafos”, que se examina en la asignatura “Elementos de Lógica y Matemática Discreta”, y se emplean al confeccionar diagramas de transición de autómatas finitos deterministas.

Cada ejercicio tiene un objetivo definido explícitamente en la herramienta.

- 1) Ejercicio Uno: tiene como objetivo identificar el significado de una notación para describir un modelo. Brinda aportes adicionales con animación simple, en la que una serie de imágenes con contenido explicativo aparecen ordenadas lógicamente. Para preparar material para este tipo de ejercitación, es necesario considerar el modelo que se ejercitará e identificar el objetivo del ejercicio, ideando los elementos que se presentan en la interfaz del sistema.
- 2) Ejercicio Dos: permite realizar una actividad de clasificación, por lo que para construir otros ejercicios asociados a otros temas, es necesario identificar el objetivo, y preparar una serie de imágenes y mensajes de texto, con respuestas ante errores y ante aciertos.
- 3) Ejercicio Tres: lleva a establecer relaciones entre elementos. Brinda resultados interactivos, a medida que el usuario ingresa una asociación, presenta resultados logrados por el estudiante, y respuestas completas. Por lo que para confeccionar un ejercicio de este tipo, es necesario identificar los objetivos, y preparar la lista de imágenes, destacando soluciones correctas, y mensajes ante aciertos y ante errores.

Los aportes, en los tres ejercicios, llevan a incorporar material educativo adicional de contenidos íntimamente relacionados al tema bajo ejercitación.

El software EduAFin se está utilizando como diseño base, cuyo formato, permitirá confeccionar otros módulos de software educativo, de apoyo al aprendizaje de otras temáticas, entre ellas: la ya

mencionada, el tema “Grafos”, ”Estadística Descriptiva: diagrama de cajas y bigotes o box-plot para dos o mas grupos de datos. Comparación de resultados”, “Probabilidades: resolución por diagramas de Benn” y “Estadística Descriptiva: cálculo de la moda para distribuciones cuantitativas continuas”.

4. FORMACION DE RECURSOS HUMANOS:

Se han desarrollado experiencias educativas aplicando módulos destacados, en la asignatura “Fundamentos Teóricos de la Informática”, con alumnos de la carrera de Licenciatura en Informática.

Si bien el grupo es heterogéneo en la composición y antecedentes de sus integrantes, el hecho común de pertenecer a diferentes cátedras, abre la posibilidad de no solamente incorporar nuevas herramientas de apoyo al proceso de enseñanza-aprendizaje, sino de capacitar a los docentes de las mismas, que están desarrollando estos sistemas.

Asimismo, el Director del PI está realizando su Tesis de Maestría en Redes de Datos-UNLP; la Codirectora está cursando actualmente el Magister en Tecnología Informática aplicada en Educación-UNLP, donde han examinado, entre otros temas, los distintos tipos de inteligencia, estudios aplicados en el proyecto, algunos reflejados en el anterior material de participación en WICC 2008 (Sistema SAAF). Otro de los integrantes del PI, se recibió de Ingeniero en Sistemas-UNPA y comenzó sus estudios de Doctorado en Ingeniería de Software en la Universidad de Málaga (España) y otro integrante se recibió recientemente de Licenciado en Informática-UNPSJB.

5. BIBLIOGRAFIA

- [1] Teoría de la Computación. J. Glenn Brookshear. Addison Wesley. 1993.
- [2] Teoría de Automatas y Lenguajes Formales. Pedro García, Tomás Perez, José Ruis, E. Segarra. Alfaomega. 2001.
- [3] Lenguajes, gramáticas y autómatas. Un enfoque práctico. Pedro Isasi. Paloma Martínez. Daniel Borrajo. Addison Wesley. 1997.
- [4] Compiladores: teoría y construcción. F.J.Sanchis Llorca, C. Galan Pascual. Paraninfo. 1986.
- [5] Lenguajes formales y teoría de la computación. 3º edición. Martin John C. Mc Graw Hill. 2004.
- [6] Introducción a la teoría de autómatas lenguajes y computación. John Hopcroft, Rajeev Motwani, Jeffrey D. Ullman. Addison Wesley. 2002.
- [7] Lenguajes, gramáticas y autómatas. Curso básico. Rafael Cases Muñoz. Lluís Márquez Villodre. Alfaomega. 2002.
- [8] Java Cómo Programar. Quinta edición. Deitel y Deitel. Pearson-Prentice Hall. 2004.
- [9] Java a través de ejemplos. Jesús Bobadilla. Alfaomega. Ra-Ma. 2003.
- [10] Estructuras de datos en java. Mark Allen Weiss. Addison Wesley. 2000.
- [11] Cómo programar en Java. Deitel y Deitel. Prentice Hall. 1998.
- [12] C/C++ y java como programar. Deitel. Pearson. 2004.
- [13] Análisis y Diseño de sistemas. Sexta Edición. Kendall & Kendall. Pearson-Prentice Hall. 2005.
- [14] Enseñanza asistida y diseño de sitios Web con ToolBook II. José Carlos Mota. Julia Castillo. Alfaomega. Ra-Ma. 1999.
- [15] ToolBook. Crear Multimedia con PC. E. Álvarez Sáiz. J. I. Álvaro González. 1998.
- [16] Mapas conceptuales. La gestión del conocimiento en la didáctica. Virgilio Hernández Fortes. Alfaomega. 2005.

Línea de investigación: Las TICs y la didáctica en la enseñanza de la química en cursos universitarios iniciales

Zulma Cataldi, M. Cristina Donnamaría y Fernando J. Lage

zcataldi@posgrado.frba.utn.edu.ar, donna@iflysib.unlp.edu.ar, flage@fi.uba.ar,

Escuela de Educación de Posgrado. Facultad Regional Buenos Aires Universidad Tecnológica Nacional.
LIEMA Laboratorio de Informática Educativa y Medios Audiovisuales. Facultad de Ingeniería. Universidad de Buenos Aires. Ciudad de Buenos Aires.

Instituto de Física de Líquidos y Sistemas Biológicos (IFLYSIB) y Comisión de Investigaciones Científicas Provincia de BsAs. (CICPBA) La Plata.

CONTEXTO

En esta comunicación se describen los avances del Proyecto de Investigación y Desarrollo: *La Didáctica de la Química y el uso de TICs en su enseñanza en cursos universitarios iniciales, periodo 2008-2010 TEUTNBA933* de la Secretaría de Ciencia y Tecnología de Programa “*Tecnología Educativa y enseñanza de la Ingeniería*”. Este Proyecto está radicado en la Escuela de Educación de Posgrado, la Carrera de Licenciatura en Ciencias Aplicadas y la UDB Química. (s/Res. 2573/08) de la Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.

RESUMEN

En el laboratorio químico real los tiempos de trabajo siempre son escasos, como así los productos químicos y los docentes auxiliares disponibles. Surge así un nuevo paradigma de trabajo virtual que se complementa con las clases y la labor en el aula. Esta nueva forma de interacción tiene tres componentes básicos para la herramienta software: los laboratorios virtuales, los programas de modelación y los simuladores que apoyan los procesos de enseñanza y facilitan la tarea al docente. Por este motivo, se estudia la didáctica en la enseñanza de la química con el uso de los simuladores y los laboratorios virtuales. El software permite el trabajo en un ambiente protegido que facilita la tarea y convierte al laboratorio en una aventura sin riesgos permitiendo que los estudiantes ensayen, prueben y se arriesguen a equivocarse sin miedos. Los programas ayudan también a representar eventos del mundo real lo más cercanos posibles a como aparecen en la realidad. El uso de programas de aplicación permite incrementar el interés de los estudiantes al “aprender haciendo”. Así se busca que recuperen la satisfacción respecto de sus aprendizajes a través de la adopción de estos complementos virtuales que les abren nuevas opciones y en alguna medida, se pueda revertir la idea de que la química es difícil como normalmente la conciben y que la puedan aprender con motivación.

Palabras clave: *Simuladores, enseñanza de química, laboratorios virtuales*

1. INTRODUCCION

La química es una disciplina que forma parte del diseño curricular de un gran número de carreras de universitarias y está presente en todos los aspectos de la vida cotidiana por lo que no se puede obviar su existencia. Con el uso de las computadoras han aparecido nuevas formas de aprendizaje para la enseñanza de la química que posibilitan su acercamiento a alumnos para quienes les resulta una asignatura muy poco interesante.

En primera instancia se busca analizar desde los marcos teóricos del aprendizaje haciendo basado en ideas de Dewey (1989), desde cambio conceptual y desde la pedagogía de la comprensión algunas de las aplicaciones disponibles hoy día que permiten que los estudiantes puedan llevar a cabo sus experiencias químicas sin peligro, dado que el laboratorio virtual o el simulador son verdaderos ambientes protegidos. Un ambiente protegido es aquel donde no existe riesgo de exposición física donde la informática ha creado herramientas como el correo electrónico o chat que permiten interactuar sin riesgos físicos. Así los software ya sean simuladores o laboratorios virtuales se constituyen en un ambiente protegido (Lage, 2001). Por otra parte, dado que se trata de un software que facilita la tarea, convierte al trabajo de laboratorio y sus precauciones por accidentes en una opción de aprendizaje donde el alumno puede equivocarse y repetirla las veces que fueran necesarias hasta apropiarse de los conceptos en juego ya que la inversión es por demás baja, lo que no sería posible en un laboratorio real (Cabero, 2007).

Por otra parte la computadora, permite cambiar la imagen negativa que el alumno tiene de la química, así la recibe de una manera más interesante buscando explorar el nuevo ambiente promoviendo así el autoaprendizaje y la aplicación de las capacidades de análisis, síntesis y evaluación, fomentando el pensamiento crítico usando los laboratorios virtuales y la estrategia de aprendizaje basado en problemas con problemas semejantes a los reales, favoreciendo la adquisición de técnicas de aprendizaje, con la posibilidad de transferencia otras áreas y promover en el estudiante, la comprensión de mecanismos de reacción química, la motivación e interés en experimentos de química. En primer lugar,

es de interés que los estudiantes recuperen la satisfacción respecto de sus aprendizajes y por otra también interesa la adopción de los complementos virtuales les ofrezcan nuevas opciones, para poder revertir la idea de que la química es difícil como normalmente la conciben y que la puedan aprender con motivación. Así, los estudiantes expuestos a experiencias de simulaciones interactivas de laboratorios a través de multimedia se ha visto que mejoran el dominio del material de laboratorio y de los procedimientos que deberán aplicar en las prácticas reales. De este modo basado en sistemas informáticos se pueden desarrollar y aplicar simuladores y laboratorios para la acción y el descubrimiento, es decir aprender haciendo, investigando y superando el conflicto cognitivo que conlleve al cambio conceptual desde una pedagogía para la comprensión.

1.1. La investigación en Didáctica de la Química

La investigación se inscribe en un objetivo general que es: *generar un cuerpo de conocimientos en las áreas de interés de la didáctica específica de la Química.*

El acto de enseñar implica enseñar algo con una epistemología del conocimiento enseñado, donde confluyen conocimientos prácticos y propios de la disciplina. Desde la didáctica, más allá de lo que pase con el aprendizaje, el docente es responsable de la buena propuesta y de su acción para una buena enseñanza. Así se analizará la problemática de la relación entre una propuesta didáctica y las condiciones de su puesta en práctica en contextos específicos partiendo de suponer que la enseñanza es una actividad eminentemente práctica que requiere la reflexión y toma de decisiones pedagógicas por parte del docente (Camilloni, 1998). Surgen as las preguntas de investigación: ¿Cuáles de las categorías de estrategias didácticas aparecen en las prácticas?, ¿Cuáles son las formas de enseñar de cada docente?, ¿Cuáles son comunes o recurrentes?, ¿Cuales son las creencias que surgen?, ¿Cómo se evidencian éstas cuestiones?, ¿Cómo se visualiza en currículum?, ¿Usan recursos tecnológicos?, ¿Cómo se usan, dentro de qué marcos teóricos?, ¿En qué momento y con qué finalidad?, ¿Que técnicas utilizan para abordar un contenido con tecnología?, ¿Cómo evalúan el proceso de enseñanza?, ¿En forma continua?, ¿Se autoevalúa el docente?, ¿Y el alumno?, ¿Qué piensan los alumnos?, ¿Qué estrategias les resultan más eficientes?. Basado en la necesidad de una didáctica de la Química surgen tres caminos posibles que se entrelazan como posibles *objetivos* para recortar el objeto de investigación: 1) Sistematizar los conocimientos significativos sobre la didáctica de la química, desde la perspectiva de la epistemología de la ciencia, la psicología del aprendizaje, la didáctica general y la didáctica de las ciencias, 2) Analizar las dificultades del aprendizaje

de los distintos temas de química de la enseñanza universitaria, a partir de las ideas que plantean los alumnos y 3) Interpretar y analizar los modelos didácticos de los profesores en formación a fin de aplicar a los mismos una metodología de cambio conceptual. Y Analizar las estrategias didácticas docentes y el uso de la incorporación de TICs en la enseñanza. Por lo tanto, para el estudio de las tecnologías informáticas aplicadas a la educación en química los objetivos que interesan son: a) *Indagar acerca de la inclusión de TICs en la enseñanza en cursos iniciales y su influencia en la metodología de educación en ciencias químicas y b) Analizar las estrategias didácticas docentes y el uso de la incorporación de TICs en la enseñanza.*

1.2. Algunas consideraciones sobre al aprendizaje de las ciencias: Aprender haciendo, el cambio conceptual y la pedagogía de la comprensión

La idea básica en Dewey (1989), es que el individuo se desarrolla en interacción con el contexto, y el papel de la educación es el de hacer que los individuos organicen sus experiencias aprendiendo en la interacción con el medio y su vinculación con la vida. Es una perspectiva que se apoya en la racionalidad de las ciencias, y que cree en la posibilidad de cambiar la sociedad a través de la educación. Desde su concepción del currículum Dewey remarcó la necesidad de que la escuela fuera una promotora de experiencias de cultura democrática. Pero, aprender haciendo supone que los aprendices se deban enfrentar con problemas “reales”, que tengan que ver con su propia implicación personal, perceptiva, cognitiva, y que no sean problemas de tipo “académicos”, alejados de cualquier situación práctica. Detrás de la idea de aprender haciendo, como en otros tipos de aprendizaje, hay concepciones más concretas de lo que es el aprendizaje.

El verdadero aprendizaje de las ciencias hace uso de la abstracción, siendo las operaciones abstractas las que se llevan a cabo cuando se comprende la estructura y funcionamiento de los sistemas en estudio y se interpretan las evidencias que llegan de objetos o fenómenos que no se pueden presenciar, y cuando se busca predecir acontecimientos por deducción a partir de las leyes de comportamiento. La abstracción sólo se adquiere tras un largo proceso evolutivo de desarrollo psíquico e intelectual, que comienza al nacer y finaliza con la etapa del pensamiento formal en la edad adulta. Es decir, cada individuo requiere un tiempo propio para llegar a esta etapa dependiendo de su predisposición genética, así como de estímulos, intereses y experiencias de aprendizaje que el ambiente les haya propiciado (Piaget, 1970).

Es así que más allá de la abstracción, para pasar del mundo macroscópico al submicroscópico para entender algunos de los temas esenciales, el docente

se encuentra en una situación donde debe buscar un enfoque motivador que supere las dificultades de recibir un grupo de estudiantes con una formación heterogénea y con preconceptos arraigados desde un modelo intuitivo y hasta ingenuo.

Los obstáculos al cambio conceptual, se podrían identificar con las restricciones estructurales del conocimiento implícito, se trataría de ayudar al estudiante a construir las estructuras conceptuales adecuadas (interacción, equilibrio, cuantificación, etc.) para asimilar las teorías científicas incompatibles con sus conocimientos implícitos (teorías implícitas).

La comprensión implica traducir o asimilar una información nueva a los conocimientos previos. El aprendizaje requiere que se activen estructuras de conocimiento previas a las cuales poder asimilar la nueva información. Pero, *“la asimilación de esa información nueva tiende a producir cambios en esas estructuras de conocimiento, generando conceptos más específicos por procesos de diferenciación o principios más generales, por procesos de generalización”* (Perkins, 1995). Sin embargo, a veces la comprensión o asimilación de una nueva información no es posible porque el estudiante no dispone de conocimientos previos relevantes o los que activa no son los apropiados y en ese caso, cuando no existen conocimientos previos adecuados se requiere un verdadero cambio conceptual y no tan solo la comprensión de un concepto.

Cuando los conceptos nuevos son relacionados de un modo sustancial, con lo que el alumno ya sabe y a través de los conocimientos previos, se produce la asimilación cognitiva y la memorización comprensiva. Perkins (1999) dice que *“comprender es la habilidad de pensar y actuar con flexibilidad a partir de lo que uno sabe ... la capacidad de desempeño flexible es la comprensión”*.

“Comprender es pensar con lo sabido y aplicarlo con flexibilidad en el mundo (...). No es simplemente tener conocimientos, como muchas veces se cree, sino tener la habilidad de pensar con lo que se sabe y poder aplicarlo flexiblemente en el mundo. Entendemos la comprensión como una habilidad para desempeñarse con el conocimiento que se tiene” (Stone Wiske, 2007a,b).

1.3. Metodología

Dada la naturaleza del problema en estudio, los objetivos que se persiguen en el proyecto y que la investigación actual en el área requiere de especialistas en la disciplina y en didáctica. El presente proyecto se inscribe bajo un enfoque cualitativo, en tanto se entiende que los hechos educativos deben ser considerados dentro de un contexto, por lo que el investigador se “sumerge” en el ámbito seleccionado. Se busca en una primera instancia conocer el ámbito en el que se inscribe la

misma y las características de los sujetos en estudio: docentes que enseñan química básica a fin de conocer sus trayectorias previas en la profesión, formación en docencia y su visión acerca de las necesidades de los estudiantes para una formación básica en programación desde su propia reflexión.

Los pasos a seguir serán. a) Análisis de las estrategias didácticas docentes y el uso de las TICs, b) Generación de espacios de intercambio e interacción para actualización docente, c) identificación de las concepciones metodológicas de intercambio de información que prioricen los temas, d) Elaboración de bases para una didáctica de la química y cuerpo de contenido propio.

2. LINEAS DE INVESTIGACION y DESARROLLO

2.1. Laboratorios virtuales y simuladores: software de aplicación para aprendizaje

La realización de prácticas en *laboratorios*, es uno de los objetivos más importantes que debe perseguir la enseñanza de la química ya que *además de ayudar a comprender los conceptos, permite a los alumnos incursionar en el método científico*, todas las prácticas en los laboratorios reales o virtuales, requieren que el estudiante desarrolle capacidades y destrezas como la autopreparación, a través de una serie de documentos impresos o electrónicos, la ejecución, la obtención de resultados, su evaluación y comunicación a través de un informe. Estos medios tecnológicos facilitan la tarea, convirtiendo al trabajo de laboratorio y sus precauciones por accidentes en una opción de aprendizaje donde el alumno puede equivocarse y repetirla con una inversión por demás baja, que no sería posible en un laboratorio real. La computadora por otra parte, permite cambiar la imagen negativa que el alumno tiene de la química, así la recibe de una manera más interesante buscando explorar el nuevo ambiente.

La realización de experimentos químicos sin la necesidad de comprar equipo y materiales químicos que son costosos o peligrosos brindan algunas ventajas que impactan en el proceso de aprendizaje. Experimentar en química a través de simulaciones en una computadora personal y sobre todo resolviendo problemas previos permite: a) Promover en los estudiantes el autoaprendizaje y la aplicación de las capacidades de análisis, síntesis y evaluación, b) Fomentar el pensamiento crítico usando los laboratorios virtuales y la estrategia de aprendizaje basado en problemas con problemas semejantes a los reales, c) Favorecer la adquisición de técnicas de aprendizaje, con la posibilidad de transferencia a otras áreas, d) Promover en el estudiante, la comprensión de mecanismos de reacción química, la motivación e interés en experimentos de química.

Un *simulador* no es otra cosa que conjunto de ecuaciones matemáticas que modelan en forma ideal situaciones del mundo real, ya sea por su dificultad experimentar o comprender. La tecnología ha proporcionado las herramientas y métodos para que el ambiente de simulación se transforme en un ambiente donde convivan vídeos, animaciones, gráficos interactivos, audio, narraciones, etc. (Casanovas, 2005). “Los programas de simulación construyen modelos en los cuales se representan objetos, atributos de los objetos y relaciones entre predicados científicos” (Lion, 2006). Estos programas permiten modificar algunos parámetros, posiciones relativas, procesos, Las posibilidades de los laboratorios virtuales y las simulaciones se verán plasmadas en un futuro no muy lejano, con el uso de Internet 2 a través de la *teleinmersión* (Cataldi et al., 2007) que permite que las personas que se encuentran en puntos distantes se puedan sumergir en contextos virtuales a través de dispositivos ópticos y manipular datos, compartir simulaciones y experiencias como si estuvieran juntas.

“Los programas de simulación permiten que los estudiantes ensayen, prueben y se arriesguen a equivocarse. Ayudan a representar eventos del mundo real lo más cercanos posibles a como aparecen en la realidad” (Lion, 2006). De acuerdo a lo señalado, la construcción de los simuladores requiere de muchas habilidades de los diseñadores, no es una tarea sencilla, y su costo está en proporción al tiempo que insume su producción. Así como se vio en los laboratorios, los estudiantes pueden experimentar manipulando variables sin peligro alguno. “Las simulaciones e caracterizan por la presencia de un modelo preestablecido de antemano que no es accesible y transparente al alumno. De esta manera los estudiantes no pueden modificar el modelo que subyace a la simulación. Esta predeterminación en la programación y su incapacidad para producir una situación inesperada o azarosa resulta una limitante para el aprendizaje. Es por ello que los docentes deben recuperar el trabajo con el azar y la aleatoriedad propias de toda situación científica” (Lion, 2006).

La importancia de las simulaciones reside en hacer partícipe al usuario de una vivencia para permitirle desarrollar hábitos, destrezas, esquemas mentales, etc. que influyan en su conducta, por lo que hace falta también controlar el tiempo de respuesta del usuario ya que en función de éste y de lo acertado de la decisión solucionará la situación simulada. Se pueden resumir en los siguientes principios básicos para el diseño de simuladores (a los que se les deberá sumar las características didácticas deseadas a través de indicadores específicos: a) Concurrencia de canales, auditivos y visuales, en forma proporcionada y sin redundancia, b) Interactividad debida a la toma de decisiones entre alternativas, c) Libre recorrido dentro de las posibilidades preestablecidas, d) Procesamiento de información en

tiempo real para mostrar el resultado en forma instantánea, e) Amigabilidad de la interface con el usuario, f) Valor agregado debido a la relevancia del resultado obtenido por su uso, g) Aspecto atractivo de modo que el estudiante se sienta impactado, h) Unicidad de diseño basado en una técnica y un estilo aplicados a las producciones multimedia y definidos previamente. “Los docentes entienden que los problemas de comprensión no surgen del trabajo con las computadoras sino que tiene que ver con dificultades, en este caso, en la transferencia de conceptos en el proceso de resolución de problemas. Hemos encontrado que los programas de simulación favorecen además la transferencia por que trabajan con una operatividad cercana a la vida cotidiana”. Se trata de transferencia cercana cuando se trabaja en papel y computadora, y lejana ya que establecen pautas y lineamientos orientados a la práctica profesional (Lion, 2006).

3. RESULTADOS OBTENIDOS/ESPERADOS

Hoy día se cuenta con programas de simulación¹ muy completos de un laboratorio de química como: a) *Model ChemLab*², en el cual se utilizan equipos y procedimientos de laboratorio para simular los pasos necesarios a fin de efectuar los experimentos de laboratorio y b) *Virtual Laboratory*³ es un Laboratorio Virtual desarrollado a través del proyecto IrYdium de la Universidad Carnegie Mellon que presenta dos versiones, una para trabajar en línea desde el navegador de Internet y otra que se puede descargar e instalar.

Existen programas que están disponibles en sus sitios de Internet que permiten elaborar modelos moleculares y efectuar su visualización, entre los que debido a su nivel de aplicación y facilidad de uso se pueden mencionar:

a) *ChemSketch*⁴ es un programa de uso sencillo, que se puede bajar de Internet en forma gratuita para poder construir ecuaciones químicas, estructuras moleculares y diagramas de laboratorio. Es una aplicación adecuada para crear en forma sencilla moléculas de compuestos orgánicos; experimentar con algunos instrumentos de laboratorio; resolver ejercicios y visualizar u ocultar enlaces.

b) *RasMol*⁵ es un programa para representación gráfica de moléculas grandes (proteínas y ácidos nucleicos) y de pequeñas. Es una herramienta que permite visualizar imágenes que serían imposibles de dibujar en un plano por ser muy complejas, tales como estructuras de ADN y las proteínas.

¹ Se seleccionaron dos casos de los más completos, existe otros también de gran utilidad como VLabQ (Laboratorio virtual del química) <http://www.colombiaaprende.edu.co/html/mediateca/1607/article-73438.html>

² http://www.modelscience.com/products_sp.html

³ <http://ir.chem.cmu.edu/irproject/>

⁴ <http://www.acdlabs.com/download/chemsk.html>

⁵ <http://openrasmol.org>

c) *Chime*⁶ es un módulo de programa (plug in) gratuito, que permite manipular representaciones tridimensionales en los navegadores Internet Explorer y Netscape. Su instalación habilita al navegador para trabajar con archivos de moléculas en formato pdb. Funciona de manera similar al programa RasMol y es muy útil cuando el docente necesita explicar a sus estudiantes algunas moléculas complejas. Con este programa se pueden ver, rotar y animar moléculas y cristales.

d) *3D Angles*⁷ Es un programa visualizador de estructuras tridimensionales muy fácil de utilizar; sobre todo para temas como hibridación, que presenta inconvenientes cuando se intenta representar en forma bidimensional las moléculas tridimensionales de compuestos orgánicos. La rotación de las moléculas facilita su visualización desde diferentes ángulos; los tipos de enlaces se indican con distintos colores que ayudan a identificar estereoisómeros.

En el sitio *Chemconnections*⁸ de la universidad de Berkeley (creado por the National Science Foundation para reestructurar el curriculum de química de grado), se tienen simulaciones que son muy útiles de utilizar a través de *applets* de Java y son apropiadas para varios temas de química básica, tal como las leyes de los gases, simulador de reacciones químicas, procesos termodinámicos o procesos de entálpicos y entrópicos. El acceso es fácil y rápido y aunque cada simulación tiene las explicaciones, hace falta mayor información respecto a las variables.

Así los estudiantes expuestos a experiencias de simulaciones interactivas de laboratorios a través de multimedia se ha visto que mejoran el dominio del material de laboratorio y de los procedimientos que deberán aplicar en las prácticas reales. El alumno puede volver a consultar las instrucciones de trabajo, tantas veces como lo requiera hasta entender el procedimiento a seguir, y exigiendo así en el estudiante, la habilidad de observar y tomar notas para realizar posteriormente el experimento en el laboratorio real. El simulador permite un ahorro en recursos monetarios y de instalaciones debido a que no se usan productos químicos, equipos e instalaciones físicas, dado que se realiza en la computadora.

4. FORMACION DE RECURSOS HUMANOS

La producción del grupo de investigación se plasma a través de comunicaciones a eventos y artículos Cataldi *et al.* (2008a,b,c; 2009), formación de los investigadores y tecnólogos participantes en el proyecto y dos tesis de grado en desarrollo en: *Diseño y Evaluación de Laboratorios Virtuales* de Diego Chiarenza (UTN-FRBA) y *Determinación de*

calidad de productos a través descenso crioscópico de Esther Voiro (UTN-FRH).

5. BIBLIOGRAFIA

- Cabero, J. (2008) *Las TICs en la enseñanza de la química: aportaciones desde la Tecnología Educativa*. En BODALO, A. y otros (eds) (2007): *Química: vida y progreso* (ISBN 978-84-690-781, Murcia, Asociación de químicos de Murcia.
- Camillioni, A; Celman, S; Litwin, E y Palou de Maté, M (1998). *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Paidós, Buenos Aires.
- Casanovas, I. (2005) *La didáctica en el diseño de simuladores digitales para la formación universitaria en la toma de decisiones*: Tesis para Magíster en Docencia Universitaria, UTN, Bs. As.
- Cataldi, Z. y Lage, F. (2007). *Innovaciones tecnológicas para el desarrollo de interacciones colaborativas en tiempo real: La teleinmersión. Comunicación y Pedagogía* N° 217. TIC en la sociedad de la información. marzo. ISSN: 1136-7733. Páginas 63-70.
- Cataldi, Z. y Lage, F. 2009. *Las TICs en la enseñanza de la química: experimentando y descubriendo en los laboratorios virtuales, con modelos y simulaciones*. ICECE 2009. Buenos Aires. ITBA 8-11 de marzo.
- Cataldi, Z., Dominighini, C.; Gottardo, M. y Donnamaría, D. (2008^a) *La investigación educativa en didáctica de la química y la formación docente*. Congreso Nacional de Formación Docente. UNNE. 30 y 31 de octubre.
- Cataldi, Z., Dominighini, C.; Gottardo, M. y Lage, F. (2008^b). *La enseñanza de la ingeniería: reflexiones sobre la investigación en estrategias de enseñanza*. Congreso Nacional de Formación Docente. Universidad Nacional del Nordeste. 30 y 31 de octubre.
- Cataldi, Z.; Donnamaría, C. y Lage, F. (2008^c). *Simuladores y laboratorios químicos virtuales: Educación para la acción en ambientes protegidos*. Cuadernos Digitales Número 55, diciembre. Páginas 1-10.
- Contreras, D. (1990) *El Currículum como formación, Cuadernos de Pedagogía* 194, Ed. Fontalba.
- Contreras, D. (1994) *Enseñanza, Currículum y Profesorado*, Akal Universitaria, Madrid.
- Dewey, J. (1989) *Como pensamos. Nueva exposición de la relación entre pensamiento reflexivo y proceso educativo*. Paidós. Barcelona.
- Lage (2001) *Ambiente distribuido aplicado a la formación/capacitación de RR HH. Un modelo de aprendizaje cooperativo-colaborativo* Tesis de Magíster en Informática. Facultad de Informática. UNLP
- Lion, C. (2006) *Imaginar con tecnología*. Editorial Stella. La Crujía Eds.
- Perkins, D. (1995) *La Escuela inteligente*. Gedisa.
- Perkins, D. (1999) *¿Qué es la comprensión*, en Stone Wiske, M. (comp) *La enseñanza para la comprensión*, Paidós.
- Piaget, J. (1978): *La Equilibración de las Estructuras Cognitivas*. Madrid, Ed. Siglo XXI.
- Pozo, J. I. (1998): *Aprendices y maestros*. Alianza
- Rodríguez Illera, J. L. (2004) *El aprendizaje virtual. Enseñar y aprender en la era digital*. Homo Sapiens Ediciones
- Stone Wiske, M. (2007) *Conferencia Enseñar para la comprensión con nuevas tecnologías*. Universidad de San Andrés. 8 de mayo.
- Stone Wiske, M. (2007) *Entrevista Clarín* 27 mayo

⁶ <http://www.mdlchime.com/downloads/downloadable/index.jsp>

⁷ <http://www.molsci.ucla.edu/pub/explorations.html>

⁸ <http://mc2.cchem.berkeley.edu/>

Línea de investigación: Sistemas Tutores Inteligentes orientados a la enseñanza para la comprensión

Zulma Cataldi, Cristina Donnamaría y Fernando J. Lage

zcataldi@posgrado.frba.utn.edu.ar, donna@iflysib.unlp.edu.ar, flage@fi.uba.ar

Escuela de Educación de Posgrado. Facultad Regional Buenos Aires Universidad Tecnológica Nacional.
LIEMA Laboratorio de Informática Educativa y Medios Audiovisuales. Facultad de Ingeniería. Universidad de Buenos Aires. Ciudad de Buenos Aires.
Instituto de Física de Líquidos y Sistemas Biológicos (IFLYSIB) y Comisión de Investigaciones Científicas Provincia de BsAs. (CICPBA) La Plata.

CONTEXTO

En esta comunicación se da cuenta de los avances del Proyecto de investigación: *Modelado del tutor basado en redes neuronales para un Sistema Tutor Inteligente* Código 25/C099, correspondiente al Programa de Incentivos Secretaría de Políticas Universitarias del Ministerio de Ciencia y Tecnología de la Nación. Está radicado en la Escuela de Posgrado (Carrera de Maestría en Ingeniería de Sistemas de Información) y en las Carreras de Licenciatura en Tecnología Educativa y en Ciencias Aplicadas de la Universidad Tecnológica Nacional Facultad Regional Buenos Aires, en convenio con el Laboratorio de Informática Educativa y Medios Audiovisuales de la Facultad de Ingeniería de la Universidad de Buenos Aires y se inscribe en el: *Programa "Electrónica, Informática y Telecomunicaciones"*.

RESUMEN

Un sistema tutor inteligente actúa como un tutor particular del estudiante, por lo que debe poseer libertad para actuar de acuerdo a las necesidades del estudiante. Por ese motivo se busca diseñar un sistema adaptable de acuerdo a los conocimientos previos y a la capacidad de evolución de cada estudiante y las concepciones epistemológicas que subyacen en las prácticas de enseñanza.

Palabras clave: *Sistemas tutores inteligentes, enseñanza para la comprensión.*

1. INTRODUCCION

Los sistemas tutores inteligentes (STI) comenzaron a desarrollarse en los años ochenta con la idea de poder impartir el conocimiento usando alguna de inteligencia para poder asistir y guiar al estudiante en su proceso de aprendizaje. Se buscó emular el comportamiento de un tutor humano, es decir a través de un sistema que pudiera adaptarse al comportamiento del estudiante, identificando la forma en que el mismo resuelve un problema a fin de poder brindarle ayudas cognitivas cuando lo requiera.

Un tutor inteligente, por lo tanto: *"es un sistema de software que utiliza técnicas de inteligencia artificial (IA) para representar el conocimiento e interactúa con los estudiantes para enseñárselo"*. Son *"sistemas que modelan la enseñanza, el aprendizaje, la comunicación y el dominio del conocimiento del especialista y el entendimiento del estudiante sobre ese dominio"*. (VanLehn, 1988). (Wolf, 1984) e incorporan *"técnicas de IA (Inteligencia Artificial) a fin de crear un ambiente que considere los diversos estilos cognitivos de los alumnos que utilizan el programa"* (Giraffa, 1997).

En los 90, los avances de la psicología cognitiva, las neurociencias y los nuevos paradigmas de programación, han permitido la evolución de los STI desde una propuesta instructiva conductista inicial hacia entornos de descubrimiento y experimentación del nuevo conocimiento (Bruner, 1991; Perkins, 1995, Pozo; 1998) desde la pedagogía de la comprensión (Perkins, 1995 Stone Wiske, 2007, 2008). Las dificultades de representación se centran en la identificación de los diferentes estadios evolutivos del estudiante y en el reconocimiento de los preconceptos o concepciones erróneas. Así, las teorías ingenuas o intuitivas, se basan en ideas que en general no coinciden con las explicaciones científicas. Gardner (2000) dice que para remover estas concepciones *"sólo una investigación en profundidad pondrá en evidencia los defectos de esas ideas falsas iniciales, y solo una exploración a fondo de estos temas, bajo la supervisión de alguien capaz de pensar de manera disciplinaria, puede fomentar el desarrollo de una comprensión más sofisticada"*.

"Algunas propuestas recientes para promover el cambio conceptual, (...) están dirigiéndose hacia una instrucción basada en la contrastación de modelos o teorías alternativas por parte del aprendiz con el fin de reestructurar su conocimiento. La idea es que el cambio conceptual está más vinculado a la diferenciación y reorganización de las posiciones teóricas que a la existencia de datos empíricos a favor o en contra" (Pozo, 1998).

Se ha observado que la mayor parte de los STI no presentan el nivel esperado de *"inteligencia"* debido

a la dificultad para el modelado del funcionamiento de la mente humana, más allá de la aplicación de las técnicas de programación más avanzadas. La orientación actual de las investigaciones se centra en proveer una alternativa al tutor humano, cuando no puede dedicar más tiempo a sus estudiantes y para los estudiantes que buscan aprender en forma más autónoma.

Un STI actúa como un tutor particular del estudiante ya que como un entrenador humano, posee libertad para actuar de acuerdo a las necesidades más complejas del estudiante. Los STI aún no proveen de un modo de aprendizaje lo suficientemente adaptables de acuerdo a los conocimientos previos y a la capacidad de evolución de cada estudiante y las concepciones epistemológicas que subyacen en las prácticas de enseñanza.

2. LINEAS DE INVESTIGACION y DESARROLLO

Los STI permiten la emulación de un tutor humano para determinar *qué enseñar, cómo enseñar y a quién enseñar* a través de un *módulo del dominio*: que define el dominio del conocimiento (ver Figura 1), un *módulo del estudiante*: que es capaz de definir el conocimiento del estudiante en cada punto durante la sesión de trabajo, un *módulo del tutor*: que genera las interacciones de aprendizaje basadas en las discrepancias entre el especialista y el estudiante y finalmente *la interface* con el usuario: que permite la interacción del estudiante con un STI de una manera eficiente (conocimiento sobre *cómo presentar* los contenidos). Para la interface se siguen los principios del diseño, implementación y evaluación de sistemas computacionales interactivos para su utilización por seres humanos (HCI: Human Computer Interaction), es decir que estudian y buscan de poner en práctica procesos orientados a la construcción de interfaces siguiendo el criterio de usabilidad, es decir con alto grado de facilidad en el uso del sistema interactivo de acuerdo al estándar ISO 92401 de requisitos ergonómicos para el trabajo de oficina con terminales visuales y normas asociadas. Se basan en aplicación de las leyes gestálticas que están relacionadas con los criterios de Smith y Mosier (1992) y las normas ISO 9241 (1998) y 11064 (2000) para el diseño de interfaces y ergonomía.

En 1912 Wertheimer, y luego Kohler, Koffka y Lewin, inicia en Alemania una concepción opuesta al asociacionismo denominada Gestalt. Las ideas de este movimiento se centran en la unidad mínima de análisis la estructura o la globalidad con significado propio (Gestalt) rechazando la concepción atomista del conocimiento en el que este es una suma de partes preexistentes. La Gestalt se preocupa de los problemas perceptuales, en cambio el cognitivismo conductista elabora a partir de ellos una teoría

psicológica completa incluyendo los posibles conocimientos. El conocimiento es una síntesis de la forma y contenido recibido por las percepciones, las cuales son relativas, individuales e influidas por la historia, actitud y motivación del individuo. Tiene raíces filosóficas con conceptos como: *contemporaneidad, interacción simultánea y mutua con el ambiente, relatividad de la percepción e intencionalidad de la conducta.*

A través de la interacción entre los módulos básicos, los STI son capaces de determinar lo que sabe el estudiante y cómo va en su progreso, por lo que la enseñanza, se puede ajustar según las necesidades del estudiante, sin la presencia de un tutor humano.

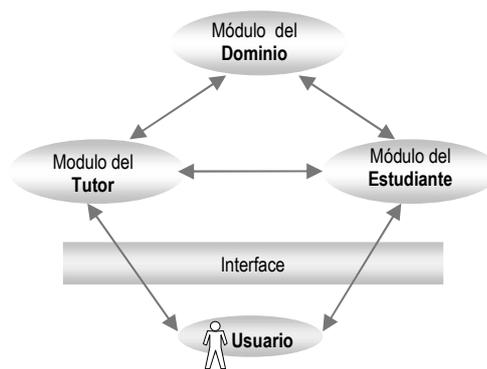


Figura 1: Interacción de los Módulos de un Sistema Tutor inteligente.

El *Modulo Tutor* del STI es quien define y aplica una estrategia pedagógica de enseñanza, contiene los objetivos a ser alcanzados y los planes utilizados para alcanzarlos. Selecciona los problemas, monitorea el desempeño, provee asistencia y selecciona el material de aprendizaje para el estudiante. Integra el conocimiento acerca del método de enseñanza, las técnicas didácticas y del dominio a ser enseñado (ver Figura 2).

Consta de: a) *Protocolos Pedagógicos*: almacenados en una base de datos, con un gestor para la misma, b) *Planificador de Lección*: que organiza los contenidos de la misma y c) *Analizador de Perfil*: analiza las características del alumno, seleccionando la estrategia pedagógica más conveniente.

El *Módulo Estudiante* del STI tiene por objetivo realizar el diagnóstico cognitivo del alumno, y el modelado del mismo para una adecuada retroalimentación del sistema.

Para el módulo estudiante se han planteado los siguientes submódulos (los datos se almacenan en una base de datos del estudiante a través del uso de un gestor) (ver Figura 2):

a) *Estilos de aprendizaje*: Está compuesto por una base de datos con los estilos de aprendizajes disponibles en el sistema, los métodos de selección de estilos y las características de cada uno de ellos.

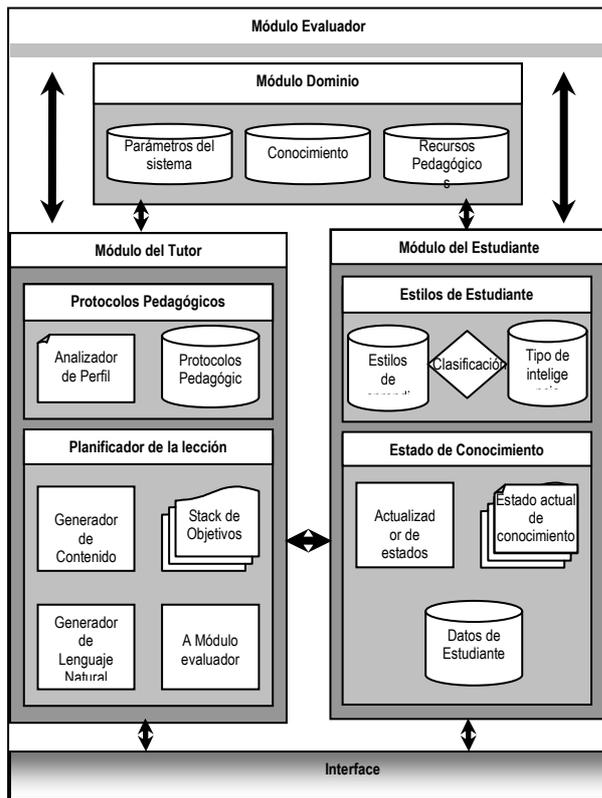


Figura 2: Esquema de un STI con sus módulos principales

Un estilo de aprendizaje es la forma de clasificar el comportamiento de un estudiante de acuerdo a la manera en que toma la información, forma las estrategias para aprender, cómo entiende y cómo le gusta analizar la información que está utilizando para acceder a un conocimiento determinado. En otras palabras, es una forma agrupar o clasificar un estudiante de acuerdo a un perfil en relación con la información, ya que este estilo evoluciona y cambia de acuerdo a las variables de entorno y ambientales que afectan al estudiante.

b) *Estado de conocimientos:* Contiene el mapa de conocimientos obtenido inicialmente a partir del módulo del dominio y que el actualizador de conocimientos irá modificando progresivamente a través de los resultados obtenidos en las evaluaciones efectuadas por el módulo del tutor quien le enviará dichos resultados procesados y

c) *Perfil psico-sociológico del estudiante:* Para determinar el perfil psico-sociológico se usa la Teoría de las Inteligencias Múltiples de Gardner (1993, 2001) quien señala no existe una inteligencia única en el ser humano, sino una diversidad de inteligencias que evidencian las potencialidades y aspectos más significativos de cada individuo, en función de sus fortalezas y debilidades para la expansión de la inteligencia. Señala que las inteligencias trabajan juntas para: a) *resolver problemas cotidianos*, b) *crear productos* o c) *para ofrecer servicios dentro del propio ámbito cultural*.

El *Módulo Dominio* tiene el objetivo global de almacenar todos los conocimientos dependientes e independientes del campo de aplicación del STI (ver Figura 2).

Entre sus submódulos están los siguientes: a) *Parámetros Básicos del Sistema:* los cuales se almacenan en una base de datos, b) *Conocimientos:* son los contenidos que deben cargarse en el sistema, a través de los conceptos, las preguntas, los ejercicios, los problemas y las relaciones, c) *Elementos Didácticos:* Son las imágenes, videos, sonidos, es decir material multimedia que se requiere para facilitarle al alumno apropiarse de conocimiento en la sesión pedagógica. Los temas relacionados con el almacenamiento de conocimiento han sido tratados ampliamente por (Russell y Norvig, 2003; Genesereth and Nilsson, 2001).

El problema que da lugar a esta investigación se centra entonces en que los STI, en general no proveen de un modo de aprendizaje que se pueda ajustar a los conocimientos previos, a la capacidad de evolución de cada estudiante (Millán *et al.*, 2000) y a las concepciones epistemológicas que subyacen en las prácticas de enseñanza.

Por otra parte, cada estudiante debería poder elegir las características del procedimiento aplicado por el tutor de acuerdo a sus preferencias, entre los diferentes métodos que éste utilice: instruccional, orientador, socrático u otros (Perkins, 1995), y si lo deseara debería poder cambiarlo de acuerdo a sus propios requerimientos.

3. Aplicaciones de Sistemas Inteligentes

En el contexto de los sistemas inteligentes se encuentran las redes neuronales, que son interconexiones masivas en paralelo de elementos simples y que responden a una cierta jerarquía intentando interactuar con los objetos reales tal como lo haría un sistema neuronal psicológico (Kohonen, 1988, 1998, 2001). Las redes neuronales poseen la característica de asimilar conocimiento en base a las experiencias mediante la generalización de casos, que las convierte en una herramienta interesante en el desarrollo de los modelos de la presente investigación (Haykin, 1999; Nilsson, 2001).

Por ejemplo, para efectuar la predicción del rendimiento académico, se puede usar una red neuronal de tipo backpropagation tomando como datos de entrada los resultados de las evaluaciones parciales desagregados en dos formas. a) tomando el caso de resolución por ejercicios y b) tomando ejercicios en función de los logros cognitivos, usando datos provenientes de las evaluaciones parciales de los estudiantes a fin de poder predecir futuros rendimientos.

Las redes bayesianas son herramientas estadísticas orientadas a la inferencia probabilística y en el

ámbito de la tutorización electrónica se pueden utilizar para modelar la incertidumbre asociada al estudiante y su nivel de conocimientos. Los algoritmos genéticos, se fundamentan en el concepto biológico de la evolución natural y son utilizados en procesos de optimización (Davis, 1991; Falkenauer, 1999). Se fundamentan en los mecanismos de la selección natural, por los que sólo sobreviven los individuos más aptos, luego de la interacción entre los mismos, pertenecientes a una población de posibles soluciones.

La minería de datos se centra en la búsqueda de patrones sugerentes y regularidades importantes en grandes bases de datos, denominado conocimiento cualitativo. La minería se puede aplicar con métodos de sistemas inteligentes y otros métodos asociados, para descubrir y detallar patrones presentes en los datos. Se pueden obtener agrupaciones en un conjunto de datos, sin tener relaciones o clases predefinidas, basándose en la similitud de los valores de los atributos de los distintos datos. La minería de datos se puede aplicar incluyendo algoritmos de inducción, algoritmos genéticos, redes neuronales y redes bayesianas; de acuerdo al problema a resolver.

Perspectiva de los STI y la enseñanza para la comprensión

La intención de esta investigación es la de proporcionar las bases para un sistema que pueda emular a un tutor humano, con un enfoque orientado hacia la psicología cognitiva, es decir, teniendo en cuenta como señala Perkins (1995), los estilos más apropiados de enseñanza tales como la instrucción didáctica, el entrenamiento y la enseñanza socrática que son compatibles con su "*Teoría Uno*".

Los nuevos desarrollos de los STI, se deben caracterizan por la inclusión de experiencia adicional basada en el entorno de aprendizaje del estudiante y en los métodos y técnicas de enseñanza. Esto permitirá sistemas más flexibles, adaptados a los intereses del estudiante y con métodos pedagógicos que faciliten el proceso de aprendizaje a través de ayuda cognitivas.

Los STI son un intento para proveer de nuevas oportunidades a los estudiantes permitiéndoles desarrollar procesos mentales de índole superior tales como la resolución de problemas (Vigotsky, 1978). De este modo podrían aportar una nueva perspectiva a los sistemas educativos que no proveen de oportunidades de enseñanza diferentes para los estudiantes con necesidades diferentes. Estas nuevas formas de interacción posibilitarán a los estudiantes adentrarse en una de las condiciones esenciales de la educación continua permitiendo la relación de su aprendizaje con los problemas de la vida real. Por otra parte, se pueden concebir tutores que trabajen para eliminar paulatinamente los conceptos erróneos (misconceptions) a fin de poder reelaborar el cambio conceptual (Perkins, 1995; Pozo, 1998).

La comprensión implica traducir o asimilar una información nueva a los conocimientos previos, así el aprendizaje requiere que se activen estructuras de conocimiento previas a las cuales poder asimilar la nueva información. Pero, "la asimilación de esa información nueva tiende a producir cambios en esas estructuras de conocimiento, generando conceptos más específicos por procesos de diferenciación o principios más generales, a través de los procesos de generalización" (Perkins, 1995). "*Comprender es pensar con lo sabido y aplicarlo con flexibilidad en el mundo (...). No es simplemente tener conocimientos, como muchas veces se cree, sino tener la habilidad de pensar con lo que se sabe y poder aplicarlo flexiblemente en el mundo. Entendemos la comprensión como una habilidad para desempeñarse con el conocimiento que se tiene*" (Stone Wiske, 2007).

Sin embargo, a veces la comprensión o asimilación de una nueva información no es posible porque el estudiante no dispone de conocimientos previos relevantes o los que activa no son los apropiados y en ese caso, *cuando no existen conocimientos previos adecuados se requiere un verdadero cambio conceptual* y no tan solo la comprensión de un concepto.

3. RESULTADOS OBTENIDOS/ESPERADOS

En esta línea de investigación se busca definir un marco teórico que sustente el diseño y la evaluación de los STI, presentando las diferentes visiones existentes acerca de los diseños y desarrollos, delineando un marco teórico general con base en la ingeniería de software, los sistemas inteligentes, la psicología cognitiva y las ciencias de la educación y elaborando una extensión metodológica específica que cautele los aspectos inherentes para el diseño de STI orientados al tutorizado. En este contexto se ha trabajado en los aspectos metodológicos de diseño (Cataldi, 2004), en proponer una arquitectura de STI (Salgueiro *et al*, 2005c), en identificar modelos del estudiante (Costa *et al*, 2005; Cataldi *et al*, 2007) y de selección del tutorizado (Cataldi *et al*, 2005), se ha investigado en el uso de redes neuronales para selección del protocolo pedagógico (Salgueiro *et al*, 2005a; 2005b; Cataldi *et al*, 2006; Cataldi *et al*, 2007), Cataldi y Lage 2009.

4. FORMACION DE RECURSOS HUMANOS

La producción del grupo de investigación se plasma en comunicaciones a eventos y artículos (Cataldi *et al*. (2004, 2005, 2006, 2007a,b y 2009), formación de los investigadores y tecnólogos participantes en el proyecto (Fernando Salgueiro, Guido Costa) y dos tesis de grado finalizadas (Ing. Fernando Salgueiro y Lic. Arnaldo Odorico), cuatro tesis de grado en curso (María Paz Colla, Nancy Szyszco, Mónica Mollo, Adriana Barbella) y una tesis de Maestría en finalizada (Ing. Martín Valiente) y otras dos en curso

(Ing. Octavio Rodríguez Angulo e Ing. Sebastián Carnotta) y una tesis doctoral en curso (MD. Oscar Bruno).

5. BIBLIOGRAFIA

- Bruner, J. (1991). *Actos de significado. Más allá de la revolución cognitiva*. Alianza. Madrid. 2002.
- Cataldi, Z. (2004). Metodología para diseño y evaluación de sistemas tutores inteligentes Plan de Tesis de Doctorado en Ciencias Informáticas. Facultad de informática UNLP .
- Cataldi, Z., Salgueiro, F., Lage, F. y García-Martínez, R. (2005) *Sistemas Tutores Inteligentes. Los Estilos del Estudiante para Selección del Tutorizado*. Proceedings del VII Workshop de Investigadores en Ciencias de la Computación. WICC 2005. Pág. 66-70. 2005.
- Cataldi, Z., Salgueiro, F., Britos, P., Sierra, E. Y García Martínez, R. (2006). *Selecting Pedagogical Protocols using SOM*. Research in Computing Science Journal, 21: 205-214.
- Cataldi, Z. y Lage, F. (2007a). *El problema del modelado del estudiante en Sistemas Tutores Inteligentes*. II Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET'07. 12-15 de junio. Facultad de Informática, Universidad Nacional de La Plata.
- Cataldi, Z.; Salgueiro, F. Y Lage, F. (2007b). *Fundamentos para el Submódulo Evaluador en Sistemas Tutores Inteligentes: Diagnóstico, predicción y autoevaluación*. CACIC 2007. 1-5 de octubre. Universidad Nacional del Nordeste Facultad de Ciencias Exactas y Naturales y Agrimensura. Corrientes y Universidad Tecnológica Nacional, Facultad Regional Resistencia.
- Cataldi, Z. y Lage, Fernando J. (2009). Sistemas tutores inteligentes orientados a la enseñanza para la comprensión *Edutec-E Revista Electrónica de Tecnología Educativa*. Núm. 28, marzo. ISSN 1135-9250
- Costa, G.; Salgueiro, F. A., Cataldi, Z., García Martínez, R. Y Lage, F. J. 2005. *Sistemas inteligentes para el modelado del estudiante* Proc. GCETE'2005, CD. marzo 13-15.
- Davis, L. (1991). *Handbook of Genetic Algorithms*. New York. Van Nostrand Reihold.
- Falkenauer, E. (1999). *Evolutionary Algorithms: Applying Genetic Algorithms to Real-World Problems*. Springer, New York, Pag 65-88.
- Gardner, H (1993). *Inteligencias Múltiples: La teoría en la práctica*. Paidós. Barcelona, Buenos Aires, México.
- Gardner, H (2001) *La inteligencia reformulada: las inteligencias múltiples en el siglo XXI*. Barcelona. Paidós.
- Gardner, H. (2000) *La educación de la mente y el conocimiento de las disciplinas*, Paidós.
- genesereth, M. and Nilsson, N. (1987) *Logical Foundations of Artificial Intelligence*. San Francisco: Morgan Kaufmann.
- Giraffa, L.M.M.; Nunes, M. A.; Viccari, R.M. (1997) *Multi-Ecological: an Learning Environment using Multi-Agent architecture*. Proc. MASTA'97: Coimbra: DE-Universidade de Coimbra.
- Haykin, S. (1999). *Neural Networks: A comprehensive foundation*. Prentice Hall 2nd. edition.
- ISO 11064-1 (2000) *Diseño ergonómico de los centros de control*.
- ISO 9241 (1996) *Requisitos ergonómicos para trabajos de oficina con pantallas de visualización de datos (PVD)*
- Kohonen, T. (1988). *Self-Organizing Maps Springer Series in Information Sciences*. Vol. 30, Springer, Berlin, Heidelberg, NY. P. 236.
- Kohonen, T. (1998). *An introduction to neural computing. Neural networks*. Vol 1. p. 3-16.
- Kohonen, T. (2001). *Self-Organizing Maps, third edition. Springer series in informarion sciences*. Ed. Springer. Helsinki University of Technology Neural Networks Research Centre 286-310. Pitman, London.
- Millán, E. (2000) *Sistema bayesiano para modelado del alumno*. Tesis Doctoral Universidad de Málaga.
- Nilsson, N. (1998) *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann Publishers.
- Perkins, D. (1995) *La escuela inteligente*. Gedisa.
- Piaget, J. (1978) *La Equilibración de las Estructuras Cognitivas*, Madrid, Ed. Siglo XXI.
- Pozo, J. I. (1998). *Aprendices y maestros*. Alianza
- Russell, S. J. Y Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2nd Edition). Prentice Hall.
- Salgueiro, F. (2005). *Sistemas Inteligentes para el Modelado del Tutor*. Tesis de Grado en Ingeniería Informática. Facultad de Ingeniería. Universidad de Buenos Aires.
- Salgueiro, F. A, Costa, G., Cataldi, Z., García Martínez, R. Y Lage, F. J. (2005a). *Sistemas inteligentes para el modelado del tutor*. GCETE'2005, Global Congress on Engineering and Technology Education. marzo 13-15.
- Salgueiro, F., Cataldi, F., Lage, F., García-Martínez, R. (2005) *Sistemas Tutores Inteligentes: Redes Neuronales para Selección del Protocolo Pedagógico*. Proceedings del IV Workshop de Tecnología Informática Aplicada en Educación del X Congreso Argentino de Ciencias de la Computación. Pág. 255-266.
- Salgueiro, F., Cataldi, Z., García-Martínez, R. (2005a). *Los Estilos Pedagógicos en el Modelado del Tutor para Sistemas Tutores Inteligentes*. Revista de Informática Educativa y Medios Audiovisuales 2(4):70-79
- Salgueiro, F., Costa, G., Cataldi, Z., Lage, F. García-Martínez, R. (2005b). *Nuevo Enfoque Metodológico para el Diseño de los Sistemas Tutores Inteligentes un Acercamiento Distribuido*. Revista de Informática Educativa y Medios Audiovisuales 2(5):25-32
- Salgueiro, F., Costa, G., Cataldi, Z., Lage, F. Y García-Martínez, R. (2005c). *Redefinition of Basic Modules of an Intelligent Tutoring System: The Tutor Module*. Proceedings del VII Workshop de Investigadores en Ciencias de la Computación. Pág. 444-448.
- Salgueiro, F; Costa, G., Cataldi, Z., Lage, F., García-Martínez, R. (2005b). *Redefinition of basic modules of an intelligent tutoring system: the tutor module*. WICC 2005. 13 y 14 de mayo. UN de Río Cuarto. Córdoba.
- Smith, S. Y Mosier, J. (1996) *Guidelines for Designing User Interface Software*, Ma, MITRE Corp.
- Stone Wiske, M. (2007A) Conferencia *Enseñar para la comprensión con nuevas tecnologías*. Universidad de San Andrés. 8 de mayo.
- Stone Wiske, M. (2007b) Entrevista Clarín 27 mayo
- Vanlehn, K (1988). Student Modelling. M. Polson. Foundations of Intelligent Tutoring systems. Hillsdale. N.J. Lawrence Erlbaum Associates, 55-78
- Vygotsky, L. (1978) *El desarrollo de los procesos psicológicos superiores*, Harvard University Press.
- Wolf, B. (1984). Context Dependent Planning in a Machine Tutor. Ph.D. Dissertation, University of Massachusetts, Amherst, Massachusetts.

ENTORNOS PARA EL APRENDIZAJE VIRTUAL: PRODUCCIÓN DE CONTENIDO ESTÁNDAR PARA CARRERAS INFORMÁTICAS Y SU GESTIÓN MEDIANTE REPOSITORIOS DE OBJETOS DE APRENDIZAJE

Berta García, Marcela Chiarani, Irma Pianucci, Vanesa Torres, Paola Allende, Hugo Viano, Viviana Ponce, Guillermo Leguizamón

Proyecto: Herramientas Informáticas Avanzadas para Gestión de Contenido de Carreras de grado en Informática

Departamento de Informática - Universidad Nacional de San Luis

Ejército de los Andes 950 - (5700) San Luis

02652-426747 int 256

e-mail: { bgarcia,mcchi, pianucci, legui}@unsl.edu.ar

CONTEXTO

Uno de los enfoques aceptados en la aplicación de las TIC en educación se basa en un nuevo paradigma de diseño de actividades de aprendizaje, que hace énfasis en la reutilización de contenidos. Como consecuencia, surge la necesidad de gestionarlos digitalmente, para ser compartidos.

Nuestro grupo de investigación viene trabajando en la temática de las Tics aplicadas a la educación, específicamente en la educación superior desde hace varios años, con la idea de integrar aspectos técnicos y pedagógicos e impactar en los procesos de enseñanza y de aprendizaje, tanto en clases presenciales como virtuales.

El contexto de aplicación se concreta dentro de la carrera del Profesorado en Ciencias de la Computación de la Universidad Nacional de San Luis. En el tramo pedagógico y aplicado se trabaja desde el año 2000 con la Plataforma Virtual ILIAS, versión 3.10.4 desarrollada como "open source" por la Universidad de Colonia, Alemania.

La experiencia obtenida en este sentido sirve de Input al proyecto, ha permitido redefinir objetivos, y resignificar resultados. En la actualidad, tomando en cuenta problemáticas comunes en el ingreso a carreras de Informática de distintas universidades de nuestro país, y considerando el esfuerzo que implica que los docentes produzcan materiales educativos adecuados, se tiene como objetivo investigar y desarrollar herramientas de software que permitan crear una red de conocimientos específicos para carreras de grado en informática.

En el contexto de una realidad cambiante y dinámica la incorporación de las Nuevas Tecnologías en ámbitos educacionales debería contribuir a mejorar la calidad de los procesos de enseñanza. Si bien las especificaciones han alcanzado un desarrollo considerable desde el punto de vista técnico, aún resultan insuficientes para contemplar la amplia variedad de modelos pedagógicos factibles de ser aplicados en una propuesta específica de enseñanza.

Se busca desde este proyecto propiciar un marco de colaboración para el desarrollo de las actividades académicas en carreras afines a modo de optimizar la generación y reusabilidad de contenidos educativos.

Este artículo presenta una descripción del estado actual del proyecto, objetivos alcanzados hasta la fecha, acciones a realizar en el año en curso y actividades de transferencia desarrolladas dentro y fuera del ámbito de la Universidad Nacional de San Luis.

RESUMEN

El costo de desarrollar recursos de aprendizaje ha generado la necesidad de compartirlos y reutilizarlos de manera eficiente. Esto, unido al desarrollo de especificaciones y estándares para solucionar el problema de incompatibilidad entre diversas plataformas para la gestión de los aprendizajes (LMS's), ha impulsado la aparición del concepto de objeto de aprendizaje (OA) y

como consecuencia, la aparición de Repositorios para la gestión de los mismos. Investigar en estos temas para aportar soluciones eficientes es uno de los grandes desafíos que enfrentan las instituciones de nivel superior que quieren adoptar la enseñanza virtual como alternativa de formación de grado o de postgrado.

Este proyecto busca investigar y desarrollar herramientas de software que permitan crear una red de conocimientos focalizados en carreras de grado en informática, posibilitando un marco de colaboración para el desarrollo de las actividades académicas en carreras afines a Ciencias de la Computación de modo de optimizar la generación y reutilización de los recursos.

Palabras clave: plataformas virtuales, educación virtual, OA, repositorios.

1. INTRODUCCIÓN

Estudiar a distancia requiere de un esfuerzo mucho mayor que el necesario para la educación presencial. Con las nuevas tecnologías aparece la posibilidad de utilizar la computadora e Internet como mediadores de los aprendizajes. Surgen el e-learning y las plataformas para darle soporte. Estas aplicaciones facilitan la creación de entornos para la enseñanza, integrando materiales didácticos, herramientas de comunicación y colaboración y herramientas de gestión educativa. El estudiante se encuentra ante múltiples dificultades: el sentimiento de soledad, la falta de motivación, la necesidad de autodisciplina y habilidades mínimas para utilizar medios informáticos: manejar la plataforma y sus herramientas. Implica, además, el conocimiento de las posibilidades cognitivas personales, la comprensión del proceso que permite construir conocimiento y la reflexión acerca del mismo.

Por parte del docente, elaborar propuestas didácticas acordes a estos nuevos requerimientos requiere de un esfuerzo significativo, y en la mayoría de los casos implica un elevado costo en tiempo y recursos.

La disponibilidad de recursos y/o materiales didácticos que actualmente encontramos en Internet nos lleva a pensar en emplearlos para nuestras asignaturas, en próximos cursos, ponerlos a disposición de otros docentes, y desagregarlos en caso que fuera necesario, para reutilizarlos en la producción de nuevos materiales.

Debido a que en la elaboración de contenidos de enseñanza se observan diferencias importantes en cuanto a diseño, desarrollo y distribución, diferentes grupos de investigación trabajan en la estandarización de los mismos.

Para cumplir con este propósito, las características principales a tener en cuenta son: formato, perdurabilidad, modularidad y metaetiquetado de recursos (para hacer más eficiente la búsqueda de los mismos).

La concepción de objeto de aprendizaje que aceptamos implica partir de unidades de aprendizaje como contenidos de formación mínima, que puedan ser reutilizados, con independencia del medio, y que deben poseer las siguientes características:

- Granulidad, de manera que no se puedan subdividir en unidades más pequeñas.
- Independencia, con sentido en sí mismas.
- Interoperabilidad: los componentes puedan ser utilizados en distintas plataformas y soportes,
- Durabilidad: capaces de soportar cambios tecnológicos sin necesidad de volver a ser rediseñados.
- Reusabilidad y flexibilidad: para incorporar componentes formativos desde diversas aplicaciones.

Desde el punto de vista pedagógico, los elementos generales a tener en cuenta al momento de pensar en un OA como pequeñas unidades instructivas deben ser:

- Objetivos formativos claros y evaluables
- Contenidos formativos indivisibles
- Estrategias didácticas específicas

- Un sistema de autoevaluación

De este modo, cada OA deberá tener:

- Objetivos instruccionales
- Contenidos
- Actividades de aprendizaje
- Autoevaluación

Es clave considerar la granularidad adecuada y los contextos posibles de uso, para propiciar la reutilización de los OA. El diseño basado en estándares posibilita la posterior incorporación de los OA como recurso en una plataforma virtual y facilita su gestión en almacenes digitales o repositorios.

Considerando lo antes expuesto se analizaron repositorios existentes, y teniendo en cuenta el material publicado, podemos decir que no todos los recursos disponibles son OA y aquellos que lo son, en su mayoría están disponibles en idioma inglés.

Teniendo en cuenta las recomendaciones del JISC (Joint Information Systems Comitee) puntualmente relacionadas con el usuario, los repositorios deberían basarse fundamentalmente en las necesidades de la comunidad de usuarios finales. Como consecuencia de lo mencionado anteriormente, surge la necesidad de contar con un repositorio de OA temático, referido a contenidos informáticos, y en idioma español.

Luego de la evaluación realizada sobre repositorios existentes se procuró seleccionar el más adecuado a nuestras necesidades, por tal motivo, se optó por DOOR. La elección se basó principalmente en que está desarrollado en PHP y MySQL, por lo que resulta compatible con la plataforma utilizada en el proyecto para sus producciones. Otro factor importante considerado es que DOOR se desarrolló siguiendo los estándares internacionales para desarrollo de contenidos educativos IMS Metadata 1.2.1 y Content Package 1.1.3.

Se realizaron modificaciones al código original, y la versión modificada fue traducida al español. Se cambió el look and feel y se diseñó un logo que representara al repositorio que alojaría objetos de aprendizajes para carreras informáticas, denominado ROI (Repositorio de Objetos de aprendizaje para carreras de Informática). Entre las funcionalidades incorporadas podemos destacar: administración de nuevo usuarios, opciones Agregar, Buscar y Extraer para usuarios autores de OA y catálogo de clasificación para OA en base a las áreas de conocimiento correspondientes a las carreras informáticas.

La implementación de un repositorio de OA debiera posibilitar la integración de los materiales didácticos en la web para que estén a disposición de los estudiantes de carreras informáticas, y faciliten el aprendizaje de manera viable, eficaz y atractiva.

En síntesis, este proyecto de investigación tiene como objetivos principales:

1. Estudiar los alcances de la educación virtual.
2. Analizar, diseñar e implementar nuevas herramientas informáticas para enriquecer los procesos formativos en entornos virtuales de aprendizaje.
3. Analizar, diseñar e implementar nuevas herramientas informáticas para desarrollar repositorios de objetos y unidades de aprendizajes correspondientes a las temáticas de las carreras de Computación.
4. Compartir un espacio virtual de contenidos que permita enriquecer la diversidad de conocimientos de los profesores de la Universidad Nacional de San Luis y de las distintas Universidades que integran la Red UNCI.

2. RESULTADOS OBTENIDOS/ESPERADOS

Entre los avances y desarrollos más significativos alcanzados hasta el momento podemos mencionar:

1. La incorporación de materias y cursos correspondientes a otras carreras de la UNSL a **evirtual**. Se crearon en la plataforma los cursos correspondientes a dos materias optativas, una materia de la carrera Psicología y un curso de Cisco.

2. El estudio de otros LMSs que permitan ofrecer un ambiente multiplataforma: en este caso se analizaron las plataformas Dokeos y Moodle.
3. La puesta en funcionamiento del repositorio ROI, un espacio virtual con material didáctico para Docentes de las carreras Informáticas.
4. Instalación, configuración y actualización de la plataforma virtual Ilias a la versión 3.10.4. Para realizar esta tarea se instaló una nueva versión de Linux, Debian 32 bits, en servidor propio y se instalaron paquetes necesarios para el funcionamiento de la nueva versión. La plataforma sigue con el nombre de **evirtual**, y se puede acceder a ella a través de la dirección www.evirtual.unsl.edu.ar, correspondiente a un subdominio del sitio de la UNSL.
5. Selección y estudio de las herramientas informáticas a utilizar para el desarrollo de material conforme a estándares. Entre las herramientas utilizadas podemos mencionar: Reload, E-learning, Hot Potatoes.
6. Estudio de compatibilidad entre las herramientas de autor anteriormente citadas y las plataformas ILIAS, Moodle y Dokeos.
7. Reutilización de los contenidos de la carrera del Profesorado en Ciencias de la Computación en **evirtual** en su modalidad semi-presencial.
8. Evaluación del funcionamiento y desempeño de la plataforma **evirtual** a partir de su utilización con los alumnos del profesorado durante el ciclo lectivo 2008, mediante las estadísticas proporcionadas por la plataforma y las categorías de análisis definidas previamente.
9. Análisis y evaluación de los aportes de los alumnos sobre la utilización del **evirtual**
10. Estudio del módulo de evaluación de la plataforma ILIAS, incluyendo: tipo de preguntas que se pueden incluir, recursos que complementan o describen la pregunta, estructuración o forma de organización de las preguntas y proceso de generación y reutilización de las mismas.
11. Estudio de la factibilidad de aplicar la estrategia del aprendizaje basado en problemas (ABP) en los sistemas de gestión de aprendizaje de código abierto mencionados anteriormente, utilizando las características que los mismos poseen.

Dentro de las actividades previstas para el ciclo lectivo 2009 se contemplan las siguientes:

1. Análisis de nuevas herramientas para la producción de objetos de aprendizaje (OA).
2. Diseño de objetos desde la perspectiva de la ingeniería de software.
3. Diseño de objetos desde la perspectiva del Modelo por Competencias.
4. Diseño de objetos desde la perspectiva del Diseños de aprendizaje (LD)
5. Estudio de los Modelos de referencia y estándares internacionales para la producción de objetos de aprendizaje.
6. Testing del repositorios ROI con los OA generados.
7. Testing de los objetos desarrollados en los Sistemas de gestión del aprendizaje que soportan OA
8. Utilización de Metadatos y otras tecnologías de la Web Semántica aplicadas al diseño y desarrollo de OA.
9. Traducción del repositorio ROI al idioma español.
10. Adaptación o creación de un tesoro para el ámbito informático, que permita estandarizar el vocabulario a utilizar en la catalogación de los recursos a incorporar.
11. Agregado a ROI de un índice de repositorios disponibles en la Web, facilitando el acceso a potenciales recursos educativos.

Al finalizar el proyecto, se prevé contar con un repositorio de objetos de aprendizaje y herramientas de integración de contenidos en funcionamiento, lo que permitirá nuclear la producción académica de Docentes Universitarios de carreras informáticas afines.

Este trabajo, que consolidará el uso de la plataforma **evirtual**, permitirá además compartir experiencias y conocimientos sobre temas comunes en espacios de discusión colaborativos.

3. FORMACION DE RECURSOS HUMANOS

Se encuentran en desarrollo tesis de grado y posgrado, como culminación de los estudios de las carreras: Licenciatura en Cs de la Computación, Maestría en Ingeniería del Software y Maestría en Educación Superior.

Cabe destacar que se realizaron numerosas actividades de transferencia, se trabajó con otras cátedras y materias, dentro y fuera de la Universidad.

1. Dentro de la Universidad:

- con la Facultad de Cs Humanas de la Universidad Nacional de S Luis, facilitando y gestionando el acceso a la plataforma, y asesorando en el uso del campus para el dictado de la materia Psicología Educativa- Línea Psicoanalítica.
- En nuestra Facultad, facilitando la plataforma y asesorando para el dictado de las materias optativas:
 - Administración de Redes Locales, de la Tecnicatura en Redes
 - Calidad de Servicio, de la Tecnicatura en Redes y la Lic. en Cs. de la Computación.
- Con el programa de capacitación de Cisco, facilitando el uso del **evirtual** para el dictado del curso CCNA 1 EXPLORATION

2. Fuera de la Universidad:

- En el marco de la formación de docentes provinciales, programa RTIC - Villa Mercedes. Uso de la plataforma como espacio de comunicación e intercambio entre docentes y capacitadores de los centros educativos de Villa Mercedes.

4. REFERENCIAS BIBLIOGRÁFICAS

1. Alvarez G. Luis A., Gallardo G. Mónica del C.: "*Diseño de un Repositorio de Objetos de Apoyo al Aprendizaje Colaborativo*". CISC 2004. Orlando - USA. 21 y 25 de julio de 2004. http://www.inf.uach.cl/alvarez/publicaciones/CISC_2004.pdf
2. Chan Ma. E., Martínez J., Morales R., Sánchez V., *Prototipo de patrimonio público de recursos educativos basados en una red institucional y un repositorio distribuido de objetos de aprendizaje*. (CUDI 2004) Colima, México. http://www.cudi.edu.mx/primavera_2004/presentaciones/MaElena_Chan.pdf
3. Chiarani Marcela Cristina, Pianucci Irma, Ponce Viviana. *ROI: Repositorio de Objetos de Aprendizaje Informáticos*. Cacic 2007, Corrientes y Resistencia, Argentina
4. Friesen, N. Interoperability and Learning Objects: an overview of elearning standardization. 2005 Int. Journal of Knowledge and Learning Objects, 1.
5. García B., Leguizamón G., Lucero Ma., Pianucci I.: *Aplicación de un Estándar de contenidos de aprendizaje en plataformas virtuales de código abierto*. Congreso Argentino en Ciencias de la computación CACIC04. La Matanza (2004) disponible en CD
6. Iriarte Navarro, L. et al. - Generación de una biblioteca de objetos de aprendizaje (LO) a partir de contenidos preexistentes. Publicado en revista electrónica Red (Revista de Educación a Distancia) - Consultado el 30/03/2009 en <http://www.um.es/ead/red/M2/>
7. Recommendations to JISC for future research and development. CD-LOR. 2007 Consultado el 30/03/2009 en <http://academy.gcal.ac.uk/cdlor/>
8. MOREA (Múltiples Objetos Reutilizables para el desarrollo de experiencias de Enseñanza y Aprendizaje) <http://www.usc.es/morea>
9. Sicilia, M. A. Reusabilidad y Reutilización de Objetos Didácticos: Mitos, Realidades y Posibilidades. RED: Revista de Educación a distancia, número monográfico II. (2005). Consultado el 15/05/2008 en: <http://www.um.es/ead/red/M2/sicilia46.pdf>
10. Wiley, D. A.. "Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy," in D. A. Wiley, ed., *The Instructional Use of Learning Objects: Online Version*. (2000). <http://reusability.org/read/chapters/wiley.doc>

“E-LEARNING”

Sanz Cecilia, Madoz Cristina, Gorga Gladys, Zangara Alejandra, Gonzalez Alejandro, Ibáñez Eduardo, Martorelli Sabrina, Moralejo Lucrecia, Guisen Andrea, Galdamez Nicolás

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

{csanz, cmadoz, ggorga, alezan@elsitio.net, agonzalez, eibanez, smartorelli, lmoralejo, aguisen, ngaldamez}@lidi.info.unlp.edu.ar

CONTEXTO

Esta línea de Investigación forma parte del Proyecto “Sistemas de Software Distribuidos. Aplicaciones en procesos industriales, E-government y E-learning”, del Instituto de Investigación en Informática LIDI acreditado por la UNLP.

RESUMEN

Este subproyecto tiene como objetivo realizar investigación en áreas de Tecnología Informática aplicada a Educación. Se presentarán aquí los avances de éste respecto de los objetivos planteados y las líneas de investigación correspondientes.

Como avances respecto de la ponencia presentada en WICC del año 2008 se han obtenido tres títulos de magister (uno de un integrante del proyecto y los otros dos dirigidos por un miembro del proyecto) y varios de especialización en Tecnología Informática Aplicada en Educación. Se han presentado propuestas de doctorado y se está trabajando con alumnos en varias tesis de grado en temas vinculados al proyecto. De igual modo se han abordado diferentes experiencias de inclusión de tecnología en ámbitos educativos.

Durante 2008 y el corriente año, se abordaron los siguientes temas:

- Desarrollo de metodologías para trabajo colaborativo usando TICs. En este sentido se ha trabajado en el ámbito del postgrado y se han utilizado diferentes herramientas (ej. E-groupware)
- Estudio de la calidad en Educación a Distancia en sistemas que incluyen la utilización de TICs. Análisis de Métricas e Indicadores para medir calidad.
- Análisis y desarrollo de materiales educativos multimediales e hipermediales, teniendo en cuenta la inclusión de metáforas visuales y personajes. En esta área se desarrollaron dos de las tesis de maestría.
- Análisis de ambientes educativos 3D.

- Análisis e implementación de laboratorios virtuales y remotos.
- Uso de la tecnología móvil integrada a entornos virtuales.
- Investigación sobre entornos virtuales de enseñanza y aprendizaje y posibilidades que se generan a partir de su utilización. Cabe aclarar que este tema viene siendo estudiado desde el inicio de este subproyecto. Se han realizado ajustes al entorno desarrollado en el Instituto y se abren nuevas líneas de investigación relacionadas (se detalla posteriormente).
- Incorporación de tecnología informática en experiencias en diferentes ámbitos educativos. Evaluación de su impacto. Se ha trabajado particularmente en la Escuela 502 de Berisso con docentes y alumnos. Esto se detallará en una sección posterior.

Palabras claves: Entornos Web, E-learning, Calidad, Laboratorios remotos, Laboratorios virtuales, Ambientes educativos 3D.

1. INTRODUCCION

En el presente proyecto se abordan diferentes aspectos vinculados a la incorporación de tecnología en el ámbito educativo. Esto va desde estudios más teóricos a experiencias concretas donde se diseñan metodologías específicas y luego se evalúa impacto de dichas experiencias.

A continuación, y sólo de manera introductoria, se mencionan algunos temas que se vinculan al proyecto, de manera tal de dar contexto a esta presentación.

Con el avance de la Web2.0 y sus posibilidades para lograr la colaboración entre personas, este tema ha adquirido especial interés en el ámbito académico. En la maestría de Tecnología Informática Aplicada a Educación, se han desarrollado varias experiencias de trabajo colaborativo utilizando herramientas

informáticas. El planteo de un trabajo colaborativo en el proceso de aprendizaje implica el diseño de una consigna, análisis del rol que tendrán los docentes y los alumnos, ofrecimiento de los medios necesarios para colaborar más allá de la presencialidad, etc. En este último caso toman un especial interés aquellas tecnologías que median el proceso de colaboración a distancia. Se han estudiado y analizado diferentes posibilidades en este sentido e incluso abordado temas más vinculados a la interdisciplinariedad en el aprendizaje colaborativo mediado por TICs [1]. También es de destacar que esto se vincula con las tendencias actuales de evolución de los entornos virtuales de enseñanza y aprendizaje.

Se ha buscado evolucionar el entorno virtual de enseñanza y aprendizaje, desarrollado en el III LIDI, de manera de ir adaptándolo e integrándolo con otras herramientas tecnológicas. Así por ejemplo, se ha avanzado en su integración con un laboratorio remoto vinculado al tema de Procesamiento Paralelo Multi-Cluster [2, 3]. Además, se ha presentado un proyecto conjunto con la Universidad Complutense de Madrid y la UNED de España para trabajar en temas vinculados a la evaluación de calidad de EVEAs (se detalla posteriormente junto con otros proyectos).

Por otra parte, el aumento de la potencia y rendimiento de los ordenadores personales ha hecho que en los últimos años se popularicen las aplicaciones de entornos virtuales en 3D. Las interfaces 3D se vuelven más cercanas al lenguaje del ser humano, lo visual se vuelve relevante, las representaciones gráficas, las imágenes y las animaciones se convierten en elementos fundamentales de la interacción persona-ordenador. Estos modernos sistemas permiten a diferentes usuarios interactuar en un escenario virtual común. Cada uno de ellos es cliente de la simulación y puede estar localizado físicamente en un punto remoto. El ámbito de aplicación de este tipo de sistemas es bastante amplio. Se han realizado aproximaciones enfocadas a diseño colaborativo, a simulaciones para el entrenamiento civil o militar, aprendizaje interactivo a distancia o los conocidos juegos en red multiusuario. Es nuestro interés, estudiar las posibilidades en el ámbito educativo de entornos virtuales de enseñanza y aprendizaje 3D, que incluyan personajes y/o avatares con diferentes funcionalidades. Se ha iniciado una investigación teórica respecto al tema en conjunto con otras universidades de España [4, 5, 6].

También se continúa trabajando en la línea de laboratorios virtuales y remotos. Hay una tesis de maestría aprobada vinculada a este tema, y se está

iniciando un trabajo de grado relacionado. Este tiene como objetivo la realización de un microscopio virtual, lo que tendría un aporte significativo en el área de la enseñanza de varias disciplinas donde esta herramienta está adquiriendo popularidad [7].

Todos los aspectos teóricos que se estudian en este proyecto o los diseños que se realizan se plasman en experiencias educativas concretas. Es por ello que uno de los puntos centrales que hemos trabajado en estos últimos años ha sido la implementación de proyectos tanto en escuelas, como en facultades (grado y postgrado). Esto nos permite tener una retroalimentación respecto de las hipótesis que planteamos.

En la siguiente sección presentaremos el listado completo de las líneas de investigación y desarrollo vinculadas a este proyecto.

2. LINEAS DE INVESTIGACION / DESARROLLO

- Educación a distancia.
- Educación basada en WEB.
- Extended y Blended Learning.
- Uso de tecnología móvil en escenarios educativos (mobile learning)
- Uso de software educativo para el tratamiento de temas específicos de diversas disciplinas.
- Impacto de incorporación de TICs en experiencias de Educación no presencial en articulación/grado/postgrado y actualización profesional.
- Ambientes y Entornos Virtuales para Educación a Distancia. Extensiones de los Entornos Virtuales orientados a las asignaturas experimentales
- Entornos virtuales 3D para educación
- Herramientas y entornos virtuales para el desarrollo de trabajo colaborativo.
- Calidad en Educación Superior y Educación a Distancia. Modelos.
- Métricas de calidad para software educativo.
- Estándares para la gestión y desarrollo de material educativo.
- Laboratorios Virtuales.
- Laboratorios Remotos.
- Centro de Servicios Educativos centrado en portales WEB
- Uso de Classmates y OLPC en escuelas.
- Adaptación de interfaces de software educativo para el mejoramiento de la accesibilidad.

3. RESULTADOS ESPERADOS/OBTENIDOS

Se detallan a continuación los principales resultados obtenidos en este proyecto.

En el marco de la línea vinculada a desarrollo de materiales educativos hipermediales y uso de personajes, se realizó una tesis de Maestría en la que se implementó un material educativo hipermedial con la inclusión de personajes para el curso de ingreso a la Facultad de Informática. Se trabajó en un taller con este material, con una metodología propuesta y se obtuvieron resultados [8].

Se continuó con el uso del material hipermedial desarrollado para el curso de apoyo para rendir final en la Facultad [9, 10].

Se trabajó en el diseño de una metodología de trabajo colaborativo y se analizaron diferentes herramientas informáticas para acompañarlo [11].

Se abordó el trabajo en una escuela especial (Escuela 502 de Berisso), en el marco del proyecto: “Tendiendo puentes hacia el interior de la escuela y hacia la sociedad: un proyecto universitario para facilitar el acceso tecnológico en una escuela socialmente desfavorecida”. En este sentido, se realizó la capacitación de los docentes para uso de TICs en el aula. Se ha trabajado con un cuento multimedia para despertar motivación en los alumnos, acercándolos a la creatividad, a la música y otros temas de interés. También se implementó una experiencia de trabajo con classmates pc con alumnos y docentes de la escuela. Actualmente, se está avanzando hacia una propuesta de trabajo de grado para la realización de un software específico para trabajar con alumnos sobre las classmates.

Se continúa realizando un taller anual en el marco de la Maestría de Tecnología Informática Aplicada a Educación sobre los entornos virtuales más utilizados por el ámbito académico, comparando sus funcionalidades. Se desarrolla material educativo sobre estos temas.

También se extiende la investigación en el tema entornos virtuales de enseñanza y aprendizaje y como consecuencia se está avanzando en el estudio de entornos virtuales 3D para educación, esto se realiza conjuntamente con la Universidad de Zaragoza y la Universidad de Islas Baleares. Además se realizaron ajustes en el EVEA desarrollado en el III LIDI, como se mencionó, vinculándolo con funcionalidades de laboratorios remotos.

Por otra parte, se está dirigiendo una tesis de Maestría referida a usabilidad en entornos virtuales en el marco de la Maestría de Tecnología Informática Aplicada en Educación, y existe un proyecto con la Universidad Complutense de Madrid vinculado a la evaluación de calidad de entornos virtuales.

En cuanto a los proyectos vinculados con la temática, el III- LIDI participa en los siguientes:

- “Integración de la tecnología móvil a los entornos virtuales de enseñanza y de aprendizaje” apoyado por la CIC. Se trata de un proyecto desarrollado conjuntamente entre el III-LIDI y un equipo de trabajo de la Universidad Tecnológica Nacional Regional Avellaneda. Entregado el informe final a fines de 2007.
- “Laboratorio de Procesamiento Paralelo Multi-Cluster accesible vía WEB” subsidiado por la Facultad de Informática de la UNLP.
- “Experiencias de intervenciones docentes en espacios virtuales”. A desarrollar en conjunto con la Universidad Nacional de la Patagonia San Juan Bosco. Res. CAFI 198/05.
- “Acuerdo con la Escuela Especial 502”. Se han realizado diversas actividades durante todo el 2008 y continúan en 2009.
- Se ha firmado un acuerdo de colaboración en estos temas con la Universidad de Zaragoza. En este marco una docente de dicha Universidad ha realizado una visita a la Facultad y dictado la siguiente conferencia: “Las potencialidades de la informática gráfica en el ámbito educativo. El caso del motor de animaciones Maxine de la Universidad de Zaragoza” (Grupo de Informática Gráfica Avanzada).
- “Arquitecturas avanzadas, gestión del conocimiento y calidad: una respuesta coordinada a los retos de los campus virtuales de nueva generación”. Proyecto conjunto con la Universidad Complutense de Madrid y la UNED de España.

4. FORMACION DE RECURSOS HUMANOS

Dos integrantes de esta línea de investigación están desarrollando su Tesis de Magister en Tecnología Informática Aplicada en Educación. Uno de los integrantes obtuvo el año pasado el título de Magister

en la Maestría de Tecnología Informática Aplicada a Educación de la Facultad de Informática (UNLP).

Hay un becario de CIC, un becario de CONICET que es doctorando, y 3 Becarios alumnos del III-LIDI trabajando en el proyecto.

Se dirigen Tesis de Magister, Trabajos de Especialista y Tesinas de Licenciatura de la Facultad de Informática en temas relacionados con el Proyecto.

Entre 2007 y 2008, se han aprobado 7 trabajos de Especialización y tres de Maestría, en el área de Tecnología y educación, dirigidos por miembros de este proyecto.

Hay una serie de trabajos de Maestría y Especialización que están desarrollándose actualmente, algunos que tendrán exposición a la brevedad, y dos de doctorado que involucra en la dirección a miembros del proyecto.

Hay alrededor de 4 trabajos de grado que se están desarrollando y se vinculan con el proyecto.

En la siguiente sección se presenta: la bibliografía y trabajos referenciados aquí, textos de estudio que se utilizan en la investigación, y por otra parte, algunas publicaciones que forman parte de los resultados de este proyecto.

5. BIBLIOGRAFIA

- [1] Sanz C., Zangara A., Otero N. (2008). "El trabajo colaborativo como espacio de reflexión teórica y práctica. El proceso de negociación visto desde el punto de vista cognitivo y desde la lógica de cada disciplina." Conferencia Internacional ICDE 2008 (International Conference on Distance Education). Universidad del Caribe. Santo Domingo. República Dominicana
- [2] Pousa A. (2007). "Laboratorio de Procesamiento Paralelo Multi-Cluster accesible via WEB". II Congreso Tecnología en Educación y Educación en Tecnología. Teyet07. La Plata. Proceeding - CD Rom. Páginas 168-172. ISBN: 978-950-34-0445-4
- [3] Madoz C., Ibañez E., Pousa A., De Giusti A. (2008). "E-Learning: Extensiones de los entornos virtuales para el manejo de actividades experimentales en Informática". VII Workshop de Tecnología Informática Aplicada en Educación. CACIC 2008. Chilecito, La Rioja, Argentina. Proceedings - CD Rom. ISBN: 978-987-24611-0-2
- [4] Baldassarri, Cerezo, Seron. (2008). "Maxine: A Platform for Embodied Animated Agents". Computer & Graphics. Vol. 32(4), páginas 430 - 437.
- [5] Cerezo, Baldassarri, Seron. (2007). "Interactive agents for multimodal emotional user interaction" IADIS Multi Conferences on Computer Science and

Information Systems: Interfaces and HCI, páginas 35-42

[6] Seron, Baldassarri, Cerezo. (2006). "MAXINEPPT: Using 3D Virtual Characters for Natural Interaction" 2º International Workshop on Ubiquitous Computing & Ambient Intelligence, páginas 241-250.

[7] García Rojo M. "Microscopios Virtuales. Aspectos actuales y futuros de la digitalización de preparaciones histológicas y citológicas"- http://conganat.uninet.edu/ivcvhap/conferencias/MG_Rojo/MicroscopiosVirtuales.htm

[8] González A. (2008). "TICs en el proceso de articulación entre la Escuela Media y la Universidad. Personajes virtuales como herramientas de un entorno de aprendizaje multimedia". Tesis de Magister en Tecnología Informática Aplicada en Educación de la Facultad de Informática de la UNLP. Presentada en Diciembre 2008.

[9] Sanz C., Madoz C., Gorga G., González A. (2007). "La importancia de la modalidad blended learning. Análisis de una experiencia educativa". II Congreso en Tecnología en Educación y Educación en Tecnología. La Plata, Buenos Aires. Argentina. Páginas 148 - 156 - ISBN: 978-950-34-0445-4

[10] Pompeya Lopez V., Sanz C. (2007). "Desarrollo de un material educativo hipermedial para su incorporación en una propuesta de enseñanza y aprendizaje en modalidad blended learning". X Congreso Iberoamericano Edutec 2007. Buenos Aires, Argentina. Publicado en <http://www.utn.edu.ar/edutec2007/publicaciones.utn> - Artículo 44 y en CD del congreso

[11] Sanz C., Madoz C., Zangara A., Albanesi B. (2008). "El trabajo colaborativo y cooperativo mediado por TICs. Herramientas informáticas utilizadas en la mediación y experiencias realizadas". Congreso Argentino en Ciencias de la Computación. CACIC 2008. Organizado por la RedUNCI. Chilecito. La Rioja. Argentina. ISBN: 978-987-24611-0-2

[12] Sanz C., Cukierman U., Zangara A., Santángelo H. y otros. (2007). "Integración de la tecnología móvil a los entornos virtuales de enseñanza y de aprendizaje". II Congreso Tecnología en Educación y Educación en Tecnología. Teyet07. La Plata. Páginas 117-126. - ISBN: 978-950-34-0445-4

[13] Gorga G., Sanz C., Madoz C. (2007). "El tema de la calidad en la educación a distancia. Hacia una propuesta de evaluación en capas". CACIC 2007. Corrientes y Resistencia (Chaco), Argentina. Publicado en CD del Congreso. Páginas 855-866. ISBN: 978-950-656-109-3

[14] Ferreira Szpiniak A., Sanz C. (2007). "Hacia un modelo de evaluación de entornos virtuales de enseñanza y aprendizaje. La importancia de la usabilidad". Congreso Argentino de Ciencias de la

Computación. CACIC 07 - VI WTIAE. Corrientes y Resistencia, Argentina. Páginas 932-947. ISBN: 978-950-656-109-3

[15] González, A. (2007) “Propuesta metodológica para el desarrollo de materiales de estudio hipermediales para la articulación Escuela Media y Universidad”. Anales del XIII Congreso Argentino de Ciencias Informáticas y Computación CACIC 2007. Corrientes. Argentina. Páginas 1050 – 1061. ISBN: 978-950-656-109-3

[16] Rosenberg, Marc. (2001). “E-learning. Estrategias para transmitir conocimiento en la era digital”. Colombia: Mc Graw Hill.

[17] Brunner, José Joaquín. (2000). “Educación: escenarios de futuro. Nuevas Tecnologías y sociedad de la información.” PREAL, Santiago de Chile. Disponible en: <http://www.preal.org/16brunner.pdf>

[18] Bartolomé, A. (2000) “Innovaciones tecnológicas en la docencia universitaria”. Universidad de Barcelona.

[19] Quéau, Philippe. (1995) “Lo Virtual. Virtudes y vértigos”, Barcelona, Paidós.

[20] Dillenbourg P. ed. (1999) “Collaborative Learning: Cognitive and Computational Approaches”. Pergamon, Elsevier Science Ltd, Oxford, Amsterdam.

[21] Uwe M. Borghoff, Schlichter. (2000) “Computer-supported Cooperative Work”. ISBN: 9783540669845. Berlin. Springer -Velgas.

[22] Sanchez Montoya, R. (2002). “Ordenador y discapacidad”. ISBN: 8478694021. Madrid. Ciencias de la educación preescolar y especial.

[22] Rodríguez Vázquez J., Sánchez Montoya R. y Soto Pérez F. (2006). “Las tecnologías en la escuela inclusiva”. Murcia: Servicio de ordenación administrativa y publicaciones.

[22] Hurtado Montesinos D. y Soto Pérez F. (2005). “Tecnologías de ayuda en contextos escolares”. Murcia: Servicio de ordenación administrativa y publicaciones.

[23] Sancho Gil J., Woodward J., Navarro Sierra J., Escoin Homs J., Muñoz Sevilla J., Fonollosa Vives T., Gaitán Pacheco R., García-Camino Burgos M., Gil González S, y López Melero M. (2001). “Apoyos digitales para repensar la educación especial”. Ediciones Universitarias de Barcelona (EUB), ediciones Octaedro.

[24] Cisnado Torres Xiomara. (2007). “Virtualización de la Enseñanza-Aprendizaje de METAPLAN”. Recuperado Diciembre 27 del 2008 de

www.infodesarrollo.ec/component/docman/doc_download/132-virtualizacion-de-la-ensenanza-de-aprendizaje-de-metaplan.html

TICs y Estrategias de Formación de Recursos Humanos en Informática

De Giusti Armando⁽¹⁾, Pesado Patricia^(1,2), Madoz Cristina⁽¹⁾, Sanz Cecilia⁽¹⁾, Gorga Gladys⁽¹⁾, Gonzalez Alejandro⁽¹⁾, Ibáñez Eduardo⁽¹⁾, Martorelli Sabrina⁽¹⁾, Pousa Adrián⁽¹⁾, Albanesi Bernarda⁽¹⁾

⁽¹⁾Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

⁽²⁾Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

{degiusti, ppesado, cmadoz, csanz, ggorga, agonzalez, eibañez, smartorelli, apousa, balbanesi}
@lidi.info.unlp.edu.ar,

CONTEXTO

Esta línea de I/D se relaciona directamente con un proyecto acreditado en el III-LIDI, así como con varias iniciativas de la Facultad de Informática de la UNLP.

Es importante notar que en la Facultad, desde 1999 existe una Maestría y Especialización en Tecnología Informática aplicada en Educación, en la cual se pone especial énfasis en la formación de recursos humanos.

En 2008 la RedUNCI produce un documento de Formación de Recursos Humanos en Informática (desde las Universidades) que recorre los perfiles a los que se dirige este proyecto: desde la articulación con los ciclos primarios y secundarios, hasta la actualización de profesionales que ya están en el mercado. En todos los casos el objetivo es adecuar el empleo de Tecnología al objetivo pedagógico buscado.

Existen proyectos relacionados con esta línea de I/D tales como los del Ministerio de Educación (Apoyo a las carreras cortas en Informática, Becas TICs, PACENI) y de hecho la misma Facultad de Informática de la UNLP con unos 2500 alumnos de grado y unos 400 alumnos de posgrado es un escenario especialmente importante para el trabajo experimental de esta línea de I/D.

RESUMEN

Esta línea de I/D se enfoca en la investigación y aplicación de las TICs en la formación de recursos humanos en Informática. El enfoque trata de ser “integral”, partiendo de las herramientas aplicadas a la articulación con los niveles educativos primario y secundario y finalizando en el reciclaje y actualización de profesionales ya egresados empleando TICs.

El escenario central son las tareas de grado y posgrado en las carreras universitarias en Informática.

En este escenario se trabaja en el empleo de tecnología en los procesos clásicos de enseñanza y aprendizaje, así como en la virtualización de Laboratorios experimentales y la accesibilidad a Laboratorios remotos.

Asimismo se investigan metodologías y recursos asociados con las Tutorías, enfocadas tanto a los alumnos como a la capacitación y actualización de docentes y profesionales.

Keywords: *Tecnologías de la Información y la Comunicación. Educación mediada por Tecnología. WEB Based Learning. Formación de Recursos Humanos en Informática.*

1. INTRODUCCION

En la última década del siglo XX las Tecnologías de la Información y la Comunicación (TICs) han contribuido al crecimiento del PBI y de la productividad laboral en la mayoría de los países, en particular los que disponían de una base tecnológica y de recursos humanos formados. [1] [2] [3].

En este marco, el *conocimiento* ha pasado a ser el elemento esencial para el desarrollo de los países y su ubicación en un mundo globalizado. Consecuentemente, las empresas de base tecnológica, con eje en la innovación han superado ampliamente a las empresas industriales “clásicas”. [4].

Esta realidad obliga (y genera oportunidades) a competir en el mercado global, eligiendo escenarios donde el “valor agregado” sea aportado por los recursos humanos, a través de la innovación y el conocimiento. En particular la industria del software resulta extremadamente sensible a este elemento diferencial [5].

1.1-La Industria del Software en Argentina

En los últimos años, Argentina ha presentado un importante y sostenido desarrollo del sector de Software y Servicios Informáticos (SSI).

Un elemento para este desarrollo ha sido la disponibilidad de recursos humanos calificados. Factores asociados son los costos relativos, la recuperación del mercado interno, la infraestructura de comunicaciones y procesamiento disponible y los avances en la legislación que incentiva el desarrollo del sector.

En general la industria del Software tiene sus “puntos fuertes” en áreas geográficas donde existen Universidades que tienen recursos humanos calificados (tanto docentes como alumnos).

Una caracterización rápida de las empresas nos muestra:

- Un conjunto reducido de empresas con mayoría de capital nacional que proveen servicios y desarrollan o integran solu-

ciones de sistemas. Compiten fuertemente con empresas internacionales, sobre todo por los servicios al Estado.

- Un conjunto numeroso y heterogéneo de PYMES que desarrollan principalmente para el mercado interno y están tratando de incursionar en nichos del mercado internacional.
- Un número reducido de empresas internacionales que trabajan en el país, en general dando servicios al Estado y también produciendo software para el mercado global, aprovechando las ventajas competitivas (en particular las referidas a disponibilidad y costo de los recursos humanos).

Globalmente el sector ha crecido (2003-2008) *un 15% anual*, lo que es muy significativo, sobre todo por la capacidad de generación de empleo por las características del producto “SSI” que requiere mano de obra calificada.

1.2- Evolución de la demanda de recursos humanos en Informática en el país.

La información suministrada por las empresas del sector de Software y Servicios Informáticos muestra una demanda creciente y estratificada [6] [7].

Todos los indicadores señalan que la demanda (en todos los niveles) supera la producción de recursos humanos calificados [8].

Por otra parte, en un fenómeno realimentado, la misma demanda absorbe alumnos en proceso de formación, para atender los niveles más simples (y también más urgentes) que requieren las empresas.

Es interesante analizar el estado de madurez del sector SSI de Argentina, en comparación con los modelos de referencia habituales (Irlanda, India) e incluso con mercados emergentes como China o el mismo Brasil: Claramente estamos en una fase preliminar de la evolución, con un crecimiento desordenado y cierta competencia por la forma-

ción y gestión de recursos humanos que requiere *coordinación* entre los actores que deben cooperar (Estado, Empresas, Universidades).

1.3- El rol de la Universidad

Un punto fundamental que las Universidades desean reforzar es que el país tiene que formar recursos humanos con capacidad de innovación, especialmente para que las empresas puedan incorporar un alto valor agregado en sus productos y competir en el mercado globalizado.

En los últimos 15 años se han desarrollado notoriamente las carreras de grado en informática, centradas en titulaciones de Licenciatura e Ingeniería. [9] [10] [11].

La calidad de estas carreras de grado es heterogénea en las Universidades y responde fundamentalmente a la disponibilidad de recursos humanos formados, así como de grupos de I/D asociados con el dictado de los cursos de la currícula.

Un aspecto central para el proceso actual es considerar la relación entre vocaciones (inscripciones en las carreras de Informática) y titulaciones de grado concretadas. La gran deserción atenta contra las posibilidades de resolver los problemas de demanda y se debe a factores generales y propios de la disciplina [12][13].

En este contexto, la Universidad debe:

- Generar profesionales de calidad, con conocimientos actualizados y capacidad de innovación.
- Colaborar con otros niveles educativos, especialmente en la formación de docentes, en la actualización curricular y en la utilización de laboratorios y “know how” universitario.
- Ofrecer mecanismos de actualización profesional y Especializaciones orientadas a la demanda.
- Desarrollar carreras de Posgrado de nivel internacional que deben sustentarse en sus grupos de Investigación y Desa-

rollo, así como en la capacidad de transferencia y vinculación tecnológica.

Estos objetivos (que al mismo tiempo lo son para la línea de I/D que se expone) nos obligan a concentrar esfuerzos en el apoyo tecnológico a los procesos de Articulación, Ingreso, Deserción temprana, Desarrollo del grado, Deserción avanzada, Actualización Profesional y Posgrado, de modo de contribuir desde la tecnología a la generación de los recursos humanos de calidad que se busca. [14] [15].

2. LINEAS DE INVESTIGACION y DESARROLLO

- WEB Based Learning.
- Entornos Virtuales de Enseñanza y Aprendizaje.
- Modelos y experiencias de incorporación de TICs en la Escuela primaria.
- Articulación de contenidos con la Escuela Secundaria, utilizando recursos de EAD y Multimedia.
- Uso de tecnología móvil en escenarios educativos (mobile learning)
- Desarrollo de Laboratorios Virtuales para el aprendizaje de Ciencias.
- Incorporación de TICs para EAD postgrado y actualización profesional.
- Herramientas 3D para educación
- Trabajo colaborativo.
- Estándares y Métricas para la gestión y desarrollo de material educativo.
- Desarrollo de soporte tecnológico para Laboratorios Remotos utilizables en grado y posgrado.
- Blended y extended learning.
- Caracterización del modelo del alumno (en los diferentes niveles) y del modelo de docente y de tutor.
- Desarrollo de herramientas de autoaprendizaje basadas en TICs.
- Herramientas tecnológicas para integrar el trabajo profesional y la formación académica formal a distancia.

3. RESULTADOS OBTENIDOS /ESPERADOS

- ✓ Se han desarrollado herramientas interactivas para los procesos de articulación con la Escuela Media, enfocados a futuros ingresantes en Informática.
- ✓ Se ha extendido el aprendizaje presencial de los cursos de Ingreso, con herramientas de demostración y prueba de conceptos, basadas en tecnología.
- ✓ Se utiliza blended learning y tutorías semipresenciales (con una modificación del entorno de EAD Web-UNLP) para el proceso de retención de ingresantes.
- ✓ Se combinan herramientas de EAD y multimedia con los cursos “normales” presenciales en el grado, buscando atender perfiles específicos de alumnos (por ej. recursantes).
- ✓ Se utilizan las TICs para permitir el desarrollo de Trabajos Finales de carrera (Tesinas) directamente dentro del ámbito profesional del alumno (integrando mediante recursos colaborativos a los responsables técnicos de las empresas donde trabajan los alumnos).
- ✓ Se ha desarrollado tecnología para el acceso transparente a Laboratorios de cómputo, directamente desde una conexión de Internet, vía entorno de EAD.
- ✓ Se utiliza el modelo del entorno de EAD (con los diferentes recursos de comunicación) para realizar una mediación tecnológica y dar apoyo en la definición y desarrollo de Tesinas de grado y Tesis de Posgrado.
- ✓ Se utilizan recursos de EAD y multimedia, orientados a la formación de Tutores en diferentes niveles. Estos mismos recursos permiten el seguimiento de la actividad real de los Tutores con los alumnos.
- ✓ Se ha adaptado el entorno de EAD para diferentes procesos de actualización profesional (ej. Industria Química, Federación Médica de la Pcia. de Bs. As.)
- ✓ Se ha extendido el entorno de EAD para favorecer el trabajo colaborativo.

- ✓ Se integran actividades con los Centros Regionales y otras Universidades, con desarrollos basados en TICs.

4. FORMACION DE RECURSOS HUMANOS

Existe un equipo de trabajo que vincula la Secretaría Académica, la Dirección de Ingreso y la Dirección de Educación a Distancia de la Facultad con el Posgrado y el proyecto en curso en el III-LIDI.

Se completaron 3 Tesis de Magíster, hay otras 4 en curso y se está formando una Becaria CONICET cuya Tesis Doctoral está dentro de los temas del proyecto.

Hay 4 Tesinas de Grado relacionadas con el proyecto en desarrollo.

5. BIBLIOGRAFIA

1. Ake, Kevin, John Clemons, and Mark Cubine. 2004. *Information Technology for Manufacturing: Reducing Costs and Expanding Capabilities*. New York: St. Lucie Press.
2. National Academy of Sciences. 2003b. *Innovation in Information Technology*. Washington: National Academies Press.
3. Chesbrough, Henry. 2003b. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston: Harvard Business School Press.
4. Finkelievich S., Prince A. “Las Universidades Argentinas en la Sociedad del Conocimiento”. Bs. As. 2006.
5. *Las Tecnologías de la Información y la Comunicación en la Formación Docente*. UNESCO. 2004.
6. Plan Estratégico Nacional SSI, Industria del Software y Servicios Informáticos, 2004-2014.

7. Plan de Acción 2008-2011, Cámara de Empresas de Software y Servicios Informáticos de la República Argentina, CESSI
8. Agenda Digital de la Provincia de Buenos Aires, Plan de Acción 2008-2011, Software y Servicios Informáticos, Comisión de Investigaciones Científicas de la Pcia. de Bs. As.
9. Propuesta de Currícula RedUNCI, Carreras de Grado en Informática, Junio 2006, www.redunci.info.unlp.edu.ar
10. Proyecto Estratégico de Reforma Curricular de las Ingenierías 2005-2007. CONFEDI. www.confedi.org.ar
11. Documento de Acreditación de las Carreras de Informática, RedUNCI, Diciembre 2005. www.redunci.info.unlp.edu.ar
12. Anuario 2007 de Estadísticas Universitarias, Secretaría de Políticas Universitarias, Ministerio de Educación de la Nación. <http://www.me.gov.ar/spu/>
13. Anuario Estadístico 2008. Informe Anual comparado de Indicadores de la UNLP. EDULP – La Plata – Octubre 2008.
14. Formación de Recursos Humanos en Informática. El rol de las Universidades. Documento de la RedUNCI, Mayo 2008.
15. Programa de Mejoramiento Institucional de la Facultad de Informática de la UNLP. Secretaría de Autoevaluación Institucional, Julio 2008.

SOFTWARE EDUCATIVO PARA EL TRATAMIENTO DE ALGUNOS TEMAS DE CÁLCULO NUMÉRICO: PRIMERA ETAPA DE EVALUACIÓN

María E. Ascheri, Rubén A. Pizarro, Gustavo J. Astudillo, Pablo García, María E. Culla
Facultad de Ciencias Exactas y Naturales - Universidad Nacional de La Pampa

CONTEXTO

Nuestro Proyecto de Investigación, acreditado en la Facultad de Ciencias Exactas y Naturales, aborda líneas de investigación estrechamente vinculadas con varios de los Proyectos que se desarrollan en el Departamento de Matemática del cual formamos parte. Algunos de ellos se relacionan con la ingeniería didáctica, la inclusión de tecnologías en el proceso de enseñanza aprendizaje y la importancia de la visualización gráfica en la enseñanza de la matemática. Por tal motivo, el desarrollo del presente Proyecto demanda una continua interacción con los investigadores de los demás Proyectos de la Institución, actuando en beneficio de sus avances.

RESUMEN

En el presente Proyecto de Investigación, estamos desarrollando un software educativo para la enseñanza y el aprendizaje de los temas “Resolución Numérica de Ecuaciones no lineales” e “Interpolación y Aproximación Polinomial y Ajuste de Curvas por Mínimos Cuadrados”.

En una primera etapa, se hizo un relevamiento de herramientas para desarrollo Web libres que permitieran el diseño de este nuevo software.

Luego de seleccionar las herramientas para el diseño de la aplicación, se elaboró un primer prototipo que evolucionó hasta convertirse en una versión beta que permite la enseñanza y el aprendizaje del tema “Resolución Numérica de Ecuaciones no lineales”.

Actualmente, dicha versión esta siendo analizada y evaluada desde tres puntos de vista: el matemático, el pedagógico y de su usabilidad.

Concretado dicho análisis, este recurso pedagógico será utilizado y evaluado en el curso de Cálculo Numérico que se dicta para las carreras de Profesorado en Matemática (3°

Año), Licenciatura en Física (3° Año) e Ingeniería Civil (2° Año).

El objetivo final que se pretende alcanzar con la elaboración de este software es el de disponer de un material didáctico que permita facilitar el proceso de enseñanza – aprendizaje de las temáticas involucradas en Cálculo Numérico. Disponer de un sistema en línea amigable con el usuario ofrecerá un buen soporte al docente de esta asignatura y de otras vinculadas con los contenidos que aquí se abordan.

Palabras clave: software educativo, cálculo numérico, software libre

1. INTRODUCCIÓN

Habiendo finalizado exitosamente con el anterior Proyecto de Investigación, titulado *Software educativo para la enseñanza – aprendizaje de temas de Cálculo Numérico*, nos propusimos continuar trabajando dentro de la misma línea de investigación. En el Proyecto anterior desarrollamos un software educativo, utilizando el paquete MATLAB, para la enseñanza-aprendizaje del tema: “Resolución Numérica de Ecuaciones no Lineales”. Actualmente, estamos desarrollando un software educativo que incluya el tema mencionado anteriormente y además, “Interpolación y Aproximación Polinomial y Ajuste de Curvas por Mínimos Cuadrados”.

La diferencia fundamental entre el anterior Proyecto y el actual es que éste esta siendo desarrollado íntegramente con software libre. Se está utilizando PHP, HTML, CSS, la librería JGraph y GIMP para el diseño y edición de imágenes (en formato PNG).

En la primera etapa del Proyecto, navegamos la Web en busca de aplicaciones destinadas a la enseñanza y el aprendizaje de temas de Cálculo Numérico (particularmente, de aquellos que se abordarán en este software).

Los resultados de esta búsqueda fueron decepcionantes, ya que no se hallaron aplicaciones de acceso libre que abordaran los temas dinámicamente. Existen, en general, un gran número de sitios que ofrecen los algoritmos implementados en diversos lenguajes o applets que aplican los métodos para funciones fijas, o simplemente arrojan los resultados de la aplicación del método. Esto nos dio la pauta de que deberíamos desarrollar nuestro software educativo desde cero.

El desafío no estaba únicamente en desarrollar una aplicación Web, sino que además debería ser un software educativo. Debiéramos aclarar, entonces, qué entendemos por software educativo. Consideraremos software educativo a toda aplicación informática que se haya diseñado intencionalmente para impactar en el proceso de enseñanza-aprendizaje. Tomaremos la definición dada por Pere Marquès (1996, p.2): “*programas para ordenador creados con la finalidad específica de ser utilizados como medio didáctico, es decir, para facilitar los procesos de enseñanza y de aprendizaje*”.

Teniendo esto como referencia, avanzamos en la selección de herramientas de libre acceso que nos permitieran hacer el desarrollo de un software libre. Esto es, “*el software libre es aquél que puede ser distribuido, modificado, copiado y usado; por lo tanto, debe venir acompañado del código fuente para hacer efectivas las libertades que lo caracterizan*.” (Culebro Juárez, et al., 2006). Estas libertades, según la Free Software Foundation, son:

0. la libertad para ejecutar el programa sea cual sea nuestro propósito,
1. la libertad para estudiar el funcionamiento del programa y adaptarlo,
2. la libertad para redistribuir copias
3. la libertad para mejorar el programa y luego publicarlo para el bien de toda la comunidad.

Teniendo en vista los aspectos mencionados en los párrafos anteriores es que realizamos, durante el año 2008, el primer prototipo que evolucionó hasta convertirse en la primera versión beta de la aplicación (ver figura 1).

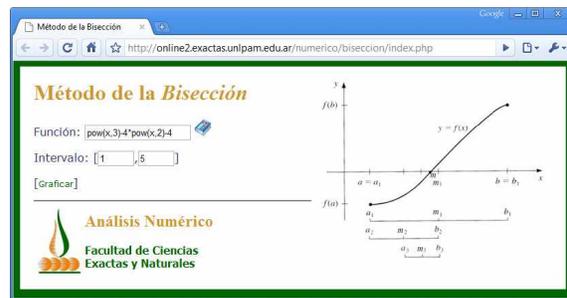


Figura 1. Captura de pantalla del primer prototipo

A finales del año 2008 y principios del 2009, se comenzó con la evaluación de la aplicación. Para esta primera etapa de la evaluación, se utilizó una técnica utilizada para evaluar sistemas denominada *caminata cognitiva*. En esta técnica “*un grupo de expertos simula la manera en como un usuario caminaría por la interfaz al enfrentarse a tareas particulares*.” (Baeza Yates & Rivera Loaiza, 2002, p.8). Esto nos permitió identificar un conjunto de fortalezas y debilidades en la aplicación desde los puntos de vista didáctico-matemático, del aprendizaje, del tratamiento del error y de la usabilidad.

Desde el punto de vista didáctico-matemático, se analizó cuál es el encuadre en el que se encuentra la aplicación acorde con la enseñanza de la matemática. También, se utilizó la clasificación propuesta por Kemmis et al. (1977) para identificar la fundamentación educativa de la aplicación, y clasificarla dentro del paradigma instruccional, revelatorio o conjetural. El primero, paradigma instruccional, deriva la instrucción programada (Skinner), y se caracteriza por el refuerzo inmediato y el avance en pequeños pasos debido a la atomización de tareas complejas. Los programas son del estilo drill-and-practice. Se le presenta al alumno actividades con el objetivo de avanzar en una dirección determinada. Por cada paso (en esa dirección) el alumno recibe un refuerzo en términos de correcto/incorrecto (retroalimentación). “*El rol de la computadora es la presentación de los contenidos, la prescripción de la tarea y la motivación del estudiante a través de una rápida retroalimentación*” (Kemmis et al., 1977, p.25). El segundo paradigma, el revelatorio, denominado así porque el conocimiento se le va revelando gradualmente

al alumno, se basa en las teorías de Bruner y Ausubel. Los programas del tipo simulación tienen la raíz en este paradigma. Las actividades que se proponen tienen la intención de reducir la brecha entre lo que sabe el alumno y el contenido (brecha cognitiva). Estas actividades deben estar diseñadas de manera que los alumnos accedan a los contenidos a partir del descubrimiento. “*La función de la computadora es la simulación o el manejo de información*” (Kemmis et al., 1977, p.26). El tercer paradigma, el conjetural, se basa en las teorías de Popper, Piaget y Paper. Propone la construcción del conocimiento a través de la experiencia. Las actividades deben propiciar el acceso al conocimiento a través de la generación y contratación de hipótesis. “*El rol de la computadora es manipular espacios, campos, lenguajes para la creación de modelos, programas, planes o estructuras conceptuales*” (Kemmis et al., 1977, p.27). Entrar en esta última clasificación los paquetes de modelización e inteligencia artificial.

Para identificar el tipo de aprendizajes que estimula el software se utilizó el análisis hecho por Squires y Mc Dougall (1994) de las relaciones entre los paradigmas antes mencionados y las habilidades cognitivas, también propuestas por Kemmis et al. La Tabla 1 muestra dichas relaciones.

Paradigma	Tipo de interacción				
	Reconocimiento	Recuerdo	Comprensión	Reconstrucción Global	Interpretación Constructiva
Instructivo	←————— ————→				
Revelador	←————— ————→				
De conjeturas	←————— ————→				

Tabla 1. Posibles interacciones entre paradigmas y tipos de interacción cognitiva propuesto por Squires y Mc Dougall (1994, p.74)

Para evaluar la usabilidad, es decir, la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario en condiciones específicas de uso (Wikipedia, 2009), utilizamos las heurísticas adaptadas por Instone (1997) y citadas por Baeza Yates et al. (2002) para evaluar la usabilidad:

- Visibilidad del estado del sistema.

- Similitud entre el sistema y el mundo real.
- Control por parte del usuario y libertad.
- Consistencia y cumplimiento de estándares.
- Prevención de errores.
- Preferencia al reconocimiento frente a la memorización.
- Flexibilidad y eficiencia de uso.
- Estética y diseño minimalista.
- Ayuda para que el usuario reconozca, diagnostique y se recupere de los errores.
- Ayuda y documentación.

El análisis dio como resultado un conjunto de recomendaciones que nos permitieron realizar varias mejoras al software, que implican desde una mejor navegabilidad y opciones más claras para facilitar la resolución de los ejercicios y la visualización de los resultados, hasta una visión global de los métodos en la asignatura e información de sus autores (ver figura 2).



Figura 2. Captura de pantalla de la versión mejorada

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación y desarrollo que seguimos son:

- Búsqueda y análisis de herramientas libres que permitan desarrollar software educativo con acceso Web.
- Desarrollo de software educativo para la enseñanza-aprendizaje de algunos de los métodos de resolución de ecuaciones no lineales.
- Elaboración de una planificación adecuada que permita la eficiente inserción del software en el desarrollo de las clases, optimizando de esta forma los resultados esperados.

3. RESULTADOS OBTENIDOS/ESPERADOS

Hasta el momento se ha logrado obtener una completa bibliografía sobre los temas que nos ocupan y sobre los software existentes. Es decir, se han podido detectar las herramientas a utilizar para el desarrollo del software y conocer además, los software educativos existentes destinados a temas de Cálculo Numérico.

También, se ha desarrollado el software para la implementación de algunos de los métodos de resolución de ecuaciones no lineales, tales como bisección, regula falsi y secante.

Actualmente, se está trabajando en el diseño de los métodos de Newton-Raphson, de von Mises e iterativo de punto fijo, y en la planificación para poder implementar estos métodos en las clases de Cálculo Numérico en el transcurso del segundo cuatrimestre del corriente año. Con ello, se espera poder llegar hacer los reajustes necesarios a efectos de optimizar el diseño y posterior uso del software educativo que estamos desarrollando.

4. FORMACION DE RECURSOS HUMANOS

Se han presentado y publicado en Reuniones Científicas de carácter nacional e internacional, varios trabajos de investigación que han surgido de este Proyecto. A partir de estas presentaciones, se ha podido hacer un relevamiento sobre los avances producidos en los temas involucrados en este Proyecto y se han intercambiado opiniones con otros grupos de investigación del área.

De esta forma, se han adquirido compromisos tendientes a realizar intercambios de las producciones que surjan de los Proyectos implicados, logrando así ampliar la población a la que estará dirigida la implementación del software educativo y que no se restrinja sólo a la cátedra de Cálculo Numérico, docentes y alumnos.

5. BIBLIOGRAFIA

Ausubel, D., Novak J., Hanesian H. (1997). *Psicología educativa. Un punto de vista cognitiva*. México. Trillas.

Baeza Yates, R. & Rivera Loaiza, C. (2002). *Ubicuidad y Usabilidad en la Web*. Consultado en Febrero, 11, 2009 en <http://www.dcc.uchile.cl/~rbaeza/inf/usabilidad.html>.

Bruner, J. (1960). *The Process of Education*. Cambridge: Harvard University Press.

Culebro Juárez, M., Gómez Herrera, W. y Torres Sánchez, S. (2006). *Software libre vs software propietario. Ventajas y desventajas*. México: (CC) Creative Commons. Consultado en Marzo, 16, 2009 en: <http://www.softwarelibre.cl/drupal/files/32693.pdf>.

Instone, K. (1997). *Site Usability Evaluation – Part. 1*. Consultado en Febrero, 12, 2009 en <http://instone.org/siteeval>.

Marqués, P. (1996). *El software educativo*. Universidad Autónoma de Barcelona. Consultado en Enero, 26, 2009 en: http://www.lmi.ub.es/te/any96/marques_software

Papert, S. (1987). *Desafío de la mente: Computadoras y educación*. Buenos Aires, Galápagos.

Piaget, J. (1985): *Psicología y Pedagogía*. Barcelona: Ariel.

Skinner, B. (1985). *Aprendizaje y comportamiento*. Barcelona. Martínez-Roca.

Squires D. y Mc Dougall A. (1994). *Cómo elegir y utilizar software educativo*. Morata. Barcelona.

Usabilidad. (2009, 28 de enero). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 22:41, febrero 11, 2009 from <http://es.wikipedia.org/w/index.php?title=Usabilidad&oldid=23617834>.

DIAGNOSTICO ADAPTATIVO DEL ESTUDIANTE EN SISTEMAS TUTORIALES INTELIGENTES

Laura Lanzarini¹
{laural@lidi.info.unlp.edu.ar}
III-LIDI²
Facultad de Informática. UNLP

Constanza R. Huapaya³
{huapaya@fi.mdp.edu.ar}
CIPBME⁴
Facultad Ingeniería-Psicología. UNMDP

CONTEXTO

La línea de I/D presentada corresponde a una Tesis de Posgrado de la UN Mar del Plata (en el área de Tecnología aplicada en Educación) que se desarrolla con la colaboración (y dirección) de la responsable del Subproyecto "Sistemas Inteligentes" del Instituto de Investigación en Informática LIDI.

RESUMEN

Esta línea de investigación analiza y desarrolla un componente del Modelo del Estudiante cuyo objetivo es mejorar la adaptación de un Sistema Tutorial Inteligente a sus usuarios, el estudiante y el docente. Este componente considera el *diagnóstico del estudiante*. A fin de tratar la ambigüedad propia de la evaluación humana del conocimiento de un alumno, se usará la lógica difusa para mejorar el Modelo del Estudiante y de este modo alcanzar un más alto nivel de adaptación.

Actualmente se está desarrollando un STI, llamado InfoSem [1] [2], en el Centro de Investigación en Procesos Básicos, Metodología y Educación (CIMEPB) de la Universidad Nacional de Mar del Plata. El dominio de InfoSem es la matemática propia de las carreras de Ingeniería (actualmente se usan tópicos de Análisis Numérico). El modelo del estudiante es un modelo de superposición con tres capas de adaptación. Se agregará el componente de diagnóstico basado en lógica difusa.

Palabras clave: *Sistemas Tutoriales Inteligentes, Modelo del estudiante, lógica difusa, diagnóstico adaptativo del estudiante*

1. INTRODUCCION

El objetivo de un Sistema Tutorial Inteligente (STI) es dar a una computadora la capacidad de involucrarse en el proceso de enseñanza/aprendizaje usando técnicas de Inteligencia Artificial. El estudiante que interactúa con el STI, está colocado en el centro de la concepción de estos sistemas y la computadora se transforma en un instructor dinámico en vez de un vehículo estático de información. El conocimiento que posea un STI le permite proveer asistencia a un aprendiz en forma parecida a un instructor humano. Un buen instructor tiene la habilidad de adaptar una lección a un estudiante individual a través del

¹ Profesor Titular. Facultad de Informática. UNLP.

² Instituto de Investigación en Informática LIDI. Calle 50 y 115 1er Piso, (1900) La Plata, Argentina, TE/Fax +(54) (221) 422-7707. <http://weblidi.info.unlp.edu.ar>

³ Profesor Titular. Facultad de Ingeniería. UNMDP

⁴ Centro de Investigación en Procesos Básicos, Metodologías y Educación. Complejo Universitario. Funes 3250. Cuerpo V Nivel III .(7600) Mar del Plata . Argentina, Tel: (0223) 4752266
<http://www.mdp.edu.ar/psicologia/grupo0.htm>

proceso de enseñanza. Un objetivo de un STI debe ser alcanzar este nivel de instrucción individualizada.

La investigación actual en Sistemas Tutoriales Inteligentes busca diseñar sistemas de aprendizaje fundamentados en principios propios y en nuevas teorías de aprendizaje así como su implementación con herramientas actuales. La evolución del área ha ido desde diseños generales hacia aplicaciones específicas con nuevas tecnologías (especialmente agentes) y nuevos estilos de sistemas (ambientes y sistemas colaborativos). Esta área de investigación se reconoce como Inteligencia Artificial en Educación (IAED)[3].

Los estudios sobre el *Modelado cognitivo* del estudiante involucran la creación de representaciones simbólicas plausibles de razonamiento necesario para resolver problemas en un dominio determinado [4].

En [5] se identifican los componentes de un STI y los denomina *módulo experto*, módulo de instrucción y currículo, y *módulo de diagnóstico del estudiante*. En 1988 [6] identificó, de manera similar, áreas como la representación del conocimiento y metodologías de enseñanza. Estos tres componentes son referidos a menudo como la ‘Trinidad tradicional’ de un Sistema Tutorial Inteligente y son presentados a continuación:

- Módulo experto (o dominio): es el conocimiento del dominio (o tópicos del tema que será enseñado) incrustado en el sistema y representa el conocimiento del experto en el tema y las características de la resolución de problemas asociados.
- Módulo del estudiante: busca capturar el entendimiento del aprendiz sobre el dominio.
- Módulo tutorial: contiene las estrategias tutoriales e instrucciones indispensables. Estas estrategias deben ajustarse a las necesidades del estudiante sin la intervención del tutor humano. El propósito principal de este módulo es reducir la diferencia del conocimiento entre el experto y el estudiante al mínimo (o a ninguna)

En particular, la línea de investigación a seguir se concentra en el modelado del estudiante. Este proceso puede ser definido como la recolección de información relevante del estudiante (habilidades de aprendizaje, sus fortalezas, estilos de aprendizaje, errores cometidos, etc.) a fin de identificar y representar su estado de conocimiento. Este modelo debe poseer la habilidad de suministrar la ayuda individualizada que el estudiante requiera.

Como la enseñanza es un proceso dinámico, un STI debe actualizar su Modelo del Estudiante a medida que el estudiante avanza en su aprendizaje. Dicho mantenimiento puede alcanzarse representando al estudiante de manera tal que sea fácilmente entendido y modificado por el STI. Se han presentado varios modelos [7] [8] a fin interpretar el progreso del estudiante, incluyendo el diagnóstico y la representación de los errores así como el seguimiento del estudiante a lo largo del tiempo. Es quizás, el área de investigación más activa actualmente y la que cuenta con importantes logros.

La calidad de la evaluación del conocimiento en un STI se determina por la cobertura y exactitud de la información construida en el modelo del estudiante y la habilidad de actualizarlo dinámicamente. El proceso de inferencia de las características internas del estudiante a partir de su comportamiento observado durante la interacción HC se llama *diagnóstico del estudiante* [9].

El punto central de los desarrollos en esta área es investigar un método que pueda analizar efectivamente las mediciones de la actividad de los estudiantes y hacer estimaciones sobre sus características internas actualizando dinámicamente el modelo del estudiante. Este modelo es usado luego para guiar el comportamiento del sistema, por ejemplo, aconsejando al evaluador o guiando al estudiante. Existen razones importantes para buscar nuevos métodos de evaluación del conocimiento del alumno, a saber:

- Puntuación imprecisa: los docentes asignan una nota, dentro de un esquema predeterminado, al trabajo de los estudiantes. Otra forma de evaluar sus aptitudes, competencias y habilidades es calificarlos con términos lingüísticos como excelente, regular, etc. Se considera que aptitud, competencia y habilidad son conceptos imprecisos, como también lo son los términos excelente y regular. Puede presuponerse que etiquetas lingüísticas de carácter difuso puedan ser capturadas mejor usando técnicas difusas.
- Control de la calificación: usualmente, los docentes evalúan colocando una nota al logro de su estudiante. Casi nunca se usa un método de evaluación alternativo para verificar esa nota final del curso, con la cual se da por aprobado o desaprobado. Un nuevo método eficiente de evaluación puede resultar de gran utilidad para confirmar o refutar la decisión tomada con métodos tradicionales.
- Uso de la lengua natural en la evaluación académica: las pruebas pueden ser evaluadas usando términos lingüísticos y con ellos lograr mayor flexibilidad en el juicio sobre el logro del estudiante.

Además, en un modelo de evaluación educativa, donde no hay comunicación directa entre el docente y el alumno, la información que adquiere el sistema difiere de la que se obtiene de la comunicación cara a cara, donde la interacción es más rica y sin interferencias.

Uno de los principales obstáculos en el proceso de diagnóstico en un sistema computacional es la *incertidumbre*. Esta incertidumbre se origina en la naturaleza abstracta de la cognición humana (p.e. la interpretación del docente sobre el rendimiento del estudiante medido generalmente con métodos estadísticos y aritméticos) y en el modo que el docente comunica sus preferencias y hábitos al sistema.

Los modelos difusos [10], pueden ‘imitar’ el modo de adquirir conocimiento por parte de los docentes cuando evalúan las características de aprendizaje de sus alumnos, tales como actitudes, nivel de conocimiento, competencias, motivación y estilo de aprendizaje. En particular, la lógica difusa [11], [12] es usada para acceder a un modo de razonamiento cualitativo cercano al razonamiento humano cuando toma decisiones, principalmente el manejo de la *imprecisión* y *vaguedad*. La forma de lograrlo es a través de la combinación de hechos difusos y relaciones difusas.

La evaluación del logro involucra, generalmente, términos lingüísticos como excelente, bueno, regular, etc. los cuales están fuertemente imbuidos de subjetividad. Estas etiquetas lingüísticas surgen a partir de diversos componentes de la evaluación como cuestionarios, planillas especializadas (como DistSem), etc. El evaluador propone estas etiquetas lingüísticas con las cuales el sistema debe llegar al diagnóstico.

En la literatura del área, existen varios desarrollos para la evaluación del estudiante basados en lógica difusa [13][14][15]. El enfoque aquí propuesto considera el tratamiento de la incertidumbre en el diagnóstico del estudiante con esta estrategia. Se prevé analizar dos alternativas a fin de alcanzar una posible integración entre ambas:

- Uso de variables lingüísticas asignadas a los componentes de las pruebas referidas a su importancia, complejidad y dificultad usando funciones miembro. Los mecanismos de inferencia usarán un sistema de reglas difusas.
- Uso de las pruebas evaluadas con un método tradicional con notas entre 1 y 10. A partir de estos valores numéricos, usar conjuntos difusos, como por ejemplo “habilidad más o menos competente para la comprensión del tópico x” o “habilidad competente para la comprensión del tópico x”. Para construir estos conjuntos difusos se puede considerar el conocimiento de los docentes experimentados cuando caracterizan cualitativamente el conocimiento del estudiante [16].

El diagnóstico del estudiante se basará en tres fuentes de evidencia mensurable:

1. Prueba DistSem. Esta prueba captura el conocimiento del estudiante sobre la distancia semántica entre conceptos. Posteriormente se mide la respuesta del estudiante con respecto a la solución de la misma prueba dada por un experto. Esta prueba es una aplicación sui generis del Análisis de Redes Sociales [17], donde los nodos a ser tratados son conceptos académicos propuestos por los docentes y las relaciones entre conceptos fueron capturadas en base a la similitud semántica estimada por los participantes. Este método permite capturar y colocar en una matriz, visualizar y comparar cuali y cuantitativamente, las redes semánticas de los participantes con un nivel de restricción aplicado a un número limitado de conceptos previamente definidos por el experto.
2. Pruebas tradicionales. Se usará la información brindada por dos asignaturas del plan de estudios de Ingeniería de la UNMDP.
3. Pruebas de opción múltiple.

2. LINEAS DE INVESTIGACION y DESARROLLO

- Estudio de un diagnóstico adaptativo del estudiante centrado en la aplicación de un modelo difuso. Esta línea utilizará los modelos planteados por Bitwas [13] y Nykanen [18]. Esto es, capturar tanto la imprecisión de clasificación inducida empíricamente como la imprecisión de las expresiones lingüísticas intuitivas vía el uso extensivo de los conjuntos difusos.
- Estudio y análisis de la aplicación de un modelo difuso, basado en la actividad del estudiante, a fin de clasificar su logro académico. La actividad del estudiante es medida con diversas pruebas, entre ellas DistSem y opción múltiple.
- Desarrollo de un modelo de reglas difusas para la aplicación, superador del modelo tradicional aritmético-estadístico, donde la inferencia para la clasificación sea más natural usando etiquetas lingüísticas en lugar de valores numéricos.
- Implementación de un prototipo para la aplicación donde, a partir de datos imprecisos, se emule el razonamiento humano.
- Comparación de los resultados obtenidos por el prototipo con los resultados de los métodos tradicionales de evaluación. Se utilizará datos de alumnos de la Facultad de Ingeniería de la UNMDP.

3. RESULTADOS OBTENIDOS/ESPERADOS

- Se han hecho pruebas DistSem con 200 alumnos de la Facultad de Psicología e Ingeniería usando InfoSem en su primera versión. Sus resultados están siendo analizados para la construcción del modelo difuso de evaluación.
- Desarrollar un STI con mayor capacidad de adaptación.
- Aplicar Infosem, con el nuevo componente difuso, en las carreras de Ingeniería y Psicología de la UNMDP.
- Ampliar el uso de InfoSem a otras áreas como su aplicación a la Neuropsicológica donde se estudiará la evolución del proceso de categorización en pacientes diagnosticados con la enfermedad de Alzheimer grado uno y dos y con ACV en el lóbulo temporal.

4. FORMACION DE RECURSOS HUMANOS

Dentro de los temas involucrados en esta línea de investigación se han finalizado dos (2) Tesis de Especialista y actualmente se está desarrollando una tesis de magister correspondiente a la Maestría en Tecnología Informática aplicada a Educación de la Facultad de Informática de la UNLP.

5. BIBLIOGRAFIA

- [1] Huapaya C.R., Arona G., Lizarralde F. y Vivas J. (2005). INFOSEM: Distancia Semántica entre Conceptos como base de la Evaluación Cognitiva. En el libro Las Ciencias del Comportamiento en los albores del Siglo XXI. UNMDP-CONICET. ISBN 987-544107-4. pp. 405-411.
- [2] Vivas J, Huapaya, C., Lizarralde, F., Arona, G., Comesaña, A, Vivas, L., García Coni, A. (2008). Distsem e Infosem: Instrumentos para la evaluación de la Memoria Semántica. Método y aplicaciones. En M. Concepción Rodríguez y V. Padilla Montemayor (Comp.) Cognición y memoria, sus representaciones y mediciones. Monterrey: Ed. Universidad Autónoma de Nuevo León.
- [3] Self J. (1999) The defining characteristics of intelligent tutoring systems research: ITSs care precisely. *International Journal of Artificial Intelligence in Education*. 10:350-364.
- [4] Anderson J.R., Corbett A.T., Koedinger K.R., y Pelletier R. (1995). The cognitive tutors: lessons learned. *Journal of the Learning Sciences*. 4(2):167-207.
- [5] Burns , Capps. (1988). *Foundation of Intelligent Tutoring Systems: An Introduction*. En M. Polson & Richardson (eds). *Foundation of Intelligent Tutoring Systems*. Lawrence Erlbaum Associates Publishers, USA.
- [6] Anderson John R. (1988). *The Expert Module*. In Martha C. Polson & J.J. Richardson. 1988. *Foundations of Intelligent Tutoring Systems*. LEA. Hillsdale, New Jersey.
- [7] Sison, R. y Shimura, M. (1998). Student modeling and machine learning. *International Journal of Artificial Intelligence in Education*. 9, 128-158.
- [8] Hartley R., Paiva A. , Self J.(1995). Externalizing Learner Models. *Proceedings de International Conference on Artificial Intelligence in Education*, editado por J. Greer, pp. 509-516. Washington. AACE.
- [9] VanLehn K.(1988). Student modeling. En *Foundations of Intelligent Tutoring Systems*, eds.: Polson M.C. y Richardson J.J. Lawrence Erlbaum.
- [10] Zadeh L.A (1965). Fuzzy sets. *Information and control* 8. p 338-353
- [11] Zadeh L.A.(1975). The concept of a linguistic variable and its application to approximate reasoning. I-III *Information Sci* 8, 199-250 y 301-357.
- [12] Dubois D. y Prade H. (1996). What are fuzzy rules and how to use them. *Fuzzy sets and systems*, 84:169-185.
- [13] Bitwas R. (1995). An application of fuzzy sets in students' evaluation. *Fuzzy sets and systems*, 74:187-194.
- [14] Law C. (1996) Using fuzzy numbers in educational grading systems. *Fuzzy sets and systems*. Pp. 311-323
- [15] Simizu S. y Yamashita H. (2000). *Educational Evaluation of Calligraphy: Applying Fuzzy Reasoning*.
- [16] Huapaya C., Lizarralde F., Vivas J., Arona G.(2007) Modelo de evaluación del conocimiento en un Sistema Tutorial Inteligente. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*. ISSN 1850-9959. 2: 20-29.
- [17] Wasserman S. y Faust K. *Social Network Analysis. Methods and Applications*. Cambridge University Press. 1998.

[18] Nykanen Ossi (2006). Inducing Fuzzy Models for Student Classification. *Educational Technology & Society*. 9(2), 223-234.

La telepresencia, la teleoperación y la generación de competencias en el Marco de Sistemas CIM (Computer Integrated Manufacturing).

Marcelo Estayno¹ (mestayno@fibertel.com.ar)
Jorge Bauer² (jbauer@ingenieria.unlz.edu.ar)
Corina Guardiola³ (corinaguardiola@gmail.com)
Diego Serra¹ (serradiego@ciudad.com.ar)

¹ Secretaría de Investigación /Facultad de Ingeniería/Universidad Nacional de Lomas de Zamora

² Laboratorio CIM/Sec. de Investigación /Facultad de Ingeniería/Universidad Nacional de Lomas de Zamora

³ Universidad CAECE

CONTEXTO

El Laboratorio CIM-Robótica de la FI-UNLZ trabaja hace años con modernos equipos de robótica industrial, en un ambiente CIM (Manufactura Integrada por Computadora). Preocupado por experimentar con conceptos organizativos y tecnologías asociadas al concepto "MFIF" (Multi-Function Intelligent Factory) [1] en todos los planos y direcciones posibles, investiga necesidades y posibles formas de telepresencia y teleoperación en la problemática de la manufactura moderna. Busca permanentemente interconectarse, lógica, física y organizativamente con industrias y laboratorios académicos colegas para potenciar el desarrollo conjunto en tiempo real de aplicaciones concretas. En ese camino desarrolla en forma modular herramientas y sistemas constructivos de manera de flexibilizar y agilizar los desarrollos para viabilizar diferentes experiencias.



Figura 1. Laboratorio CIM FI-UNLZ.

Las primeras construcciones específicas elaboradas operan generando en las articulaciones (en pocos grados de libertad) control sobre el ángulo direccional. Por ejemplo en una cámara de video estacionaria o manipuladores sencillos. Con esta construcción mecánica se trabaja en la experimentación y ampliación de los distintos módulos de software para generar variadas aplicaciones. Hoy estos elementos ya nos permiten

operar a distancia, vía Internet, servomecanismos en forma y tiempo compatible con la imagen de video y las necesidades específicas del proceso.

Siendo la telepresencia total [4] la utopía motora, los sistemas y desarrollos permiten, sobre la base de aplicaciones concretas, acumular experiencia tanto en aspectos fuertemente tecnológicos como metodológicos, ya sean estos organizativos o bien didáctico – pedagógicos.

RESUMEN

El presente trabajo se concentra en presentar desarrollos y, fundamentalmente, experiencias educativas realizadas en el marco de la principal línea de investigación del Laboratorio-Áulico-Experimental CIM-Robótica-FI-UNLZ, la cual busca integrar equipamiento y unidades de automatización robótica con un sistema de gestión no solo a nivel local, sino avanzando en un esquema de geografía global, por intermedio de Internet. En este marco se experimenta y se vuelven realidad las posibilidades de la telepresencia y la teleoperación ya sea en el entorno industrial productivo o en el ámbito académico, para la enseñanza de la robótica industrial y sus tecnologías asociadas. La posibilidad de compartir información e imagen en tiempo real adquiere con la operación total de robots a distancia, un nuevo impulso que genera un cambio cualitativo tanto en la organización y la gestión industrial en su conjunto, como en la actividad académica de los laboratorios de ingeniería, ayudando a transformar el paradigma de "dónde" se trabaja o se aprende. [6]

Independientemente de trabajos puntuales en tecnologías duras desarrollados por el equipo de investigación aplicada, necesarios para darle vida y entorno a las ideas organizativas, el presente escrito se centra en el aprovechamiento del concepto CIM y de las formas de organización modernas que apuntan al MFIF como modelo organizacional – productivo, con el objeto fundamental de aprovechar sus potencialidades para el desarrollo de programas para la formación de recursos humanos con competencias específicas en la temática.

Palabras clave: Teleoperación. Telepresencia. CIM. MFIF. Competencias. Aprendizaje a distancia.

1. INTRODUCCION

Las redes informáticas locales siempre han jugado un papel destacado en la construcción y evolución de sistemas CIM (Computer Integrated Manufacturing) e IMS (Intelligent Manufacturing Systems). Internet ha modificado radicalmente en forma cuantitativa y cualitativa las comunicaciones, permitiendo generar también profundas transformaciones en los sistemas productivos del tipo CIM-IMS. Los distintos componentes de la fabricación flexible y sus operadores humanos ya no necesariamente está concentrado físicamente en una planta productiva de geografía localizada actuando los trabajadores / operadores en forma directa sobre los equipos, sino que cada vez en mayor medida se incorporan formas de teletrabajo y telepresencia a las distintas tareas industriales. En este acelerado proceso, la teleoperación aplicada en los diferentes entornos se vuelve una poderosa herramienta para monitorear, controlar e intervenir activamente en los diversos procesos de diseño, fabricación, gestión o capacitación. El incremento de la velocidad de transferencia de datos y los adelantos en sistemas de procesamiento distribuido generan herramientas y aplicaciones que nos permiten participar activamente de procesos industriales que se desarrollan al otro lado del mundo.

Las soluciones de las dificultades tecnológicas evolucionan a ritmos vertiginosos abriendo nuevas posibilidades cuyo aprovechamiento real implica cambios radicales en las formas organizativas de los recursos humanos que integran el sistema productivo. Ello implica muchas veces cambios de paradigma en los esquemas organizativos y las formas en que los distintos trabajadores y actores se vinculan en las empresas. En especial para las pequeñas y medianas empresas PyMEs que incorporaron automatización flexible y tecnologías de la información en su proceso productivo se les abre una enorme oportunidad de compartir unidades y laboratorios de equipamiento costosos [3] que resultan cada día mas importantes en sus sistemas de calidad / metrológicos, avanzando paso a paso al concepto de empresa denominado MFIF (Multifunction Intelligent Factorys) [1].

MFIF (fig.2) es el concepto innovador cuyo modelo genera una colaboración entre las empresas con el objetivo diseñar y producir en forma cada día mas eficiente, haciendo uso intensivo y compartido de las tecnologías de la información, inteligencia artificial de telepresencia y teletrabajo.

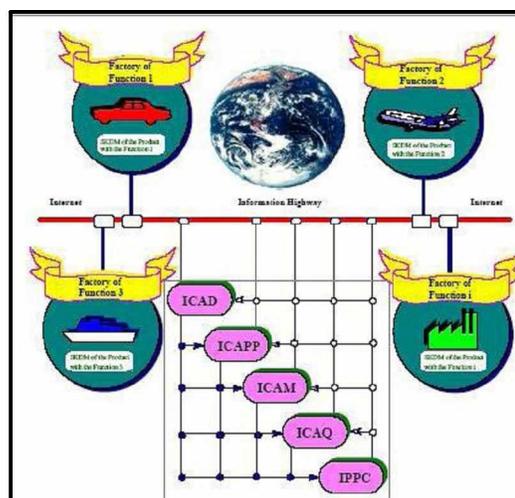


Figura 2. Multifunction Intelligent Factorys.

En base a lo expuesto, es obvio que el mundo académico no puede permanecer ajeno a este avance tecnológico, no sólo en lo que respecta al dictado de los contenidos de la temática, sino su utilización para maximizar el uso de los laboratorios específicos.

2. LÍNEAS DE INVESTIGACION y DESARROLLO

En el marco del concepto CIM-IMS, el grupo se encuentra actualmente abocado al diseño de una metodología de formación por competencias [11][12] que permita un máximo aprovechamiento de los recursos de hardware existentes en el laboratorio CIM para la formación de recursos humanos en las áreas específicas de robótica y robótica industrial.

Esta actividad implica, por un lado el desarrollo de una metodología pedagógico-didáctica basada en el concepto de formación por competencias y enmarcada dentro de la teoría constructivista [9] [10] y, por otro, el diseño de material y dispositivos didácticos que permitan su implementación efectiva en el laboratorio CIM, permitiendo el uso de éste de manera remota, y, por lo tanto, potenciando la utilización de un recurso escaso como éste a través del aprovechamiento de las herramientas de la telepresencia y teleoperación vía Internet. Sintéticamente, esto implica:

-El diseño y construcción de "pseudo grippers" mecánicos, cuyo objetivo será familiarizar al alumno con el uso de coordenadas espaciales y otros elementos básicos utilizados en robótica, como, por ejemplo, volumen de trabajo y volumen útil de trabajo.

-El desarrollo de un software simulación robótica plano o bidimensional.

-El desarrollo de un software de simulación robótica espacial o tridimensional capaz de simular

situaciones de uso reales idénticas a las de los equipos robóticos del laboratorio CIM de la FI-UNLZ. En dicho simulador, a diferencia de los simuladores de uso comercial existentes en el mercado, el alumno tendrá las mismas funciones y percepciones como si trabajara en el robot real ER IX del laboratorio CIM (fig.3).



Figura 3. Robot Antropomórfico ER9.

-El diseño y desarrollo de un mini-robot transportable, capaz de ser fácilmente trasladado, temporariamente, a los distintos centros de capacitación proyectados, siendo comandado desde una PC estándar, en cada centro de capacitación. Esta Herramienta, tiene como objeto lograr un fuerte contacto operacional real de los alumnos con un equipo robótico real.

- El desarrollo de un sistema de operación a distancia vía Internet mediante el uso de cámaras IP instaladas en el laboratorio CIM, en donde el alumno puede elegir no sólo entre tres posiciones de observación, sino también el ángulo y sector que desea observar en tiempo real, generando una percepción de participación a partir de las técnicas de tele-presencia y tele-operación.

-Desarrollar la posibilidad de operación del brazo robótico real (ER IX del CIM) vía tele-operación, en una primera etapa, y en una segunda, el resto de los recursos existentes en el laboratorio CIM como un todo.

3. RESULTADOS OBTENIDOS/ESPERADOS

Como resultado final se prevé el diseño de una metodología de capacitación eficaz, junto con el diseño de herramientas didácticas, físicas y virtuales, asociadas a la misma, a los fines de proyectar las actividades del laboratorio CIM – FI-UNLZ hacia alumnos de otras instituciones a través de un entorno

de telepresencia y teleoperación, capaz de generar competencias específicas en robótica y robótica industrial.

Se experimentó con diversas herramientas para los modelos de simulación, se realizaron actividades de personalización, adaptándolos a las características de los modelos ER IX y ER XIV. Se desarrolló un circuito electrónico para controlar a través del Puerto Paralelo de la PC los servomotores de las cámaras WEB y Robots.

Asimismo se ha desarrollado una investigación sobre como operar el software de base Visual Basic para generar software específico que manejen vía IP aplicaciones y comunicaciones, controlándose los servomotores asociados a las cámaras Web a través de comunicaciones vía Internet IP.

Por otra parte, se han desarrollado los primeros prototipos de simuladores bidimensionales: Prototipo Orientación Torre de Hanoi y Prototipo Orientación Brazo Antropomórfico, junto con el diseño preliminar del pseudogriper y el mecanizado del prototipo

4. FORMACIÓN DE RECURSOS HUMANOS

El Grupo del Laboratorio CIM FI-UNLZ es un grupo con fuertes raíces en la Facultad de Ingeniería de la Universidad de Lomas de Zamora, al que sin embargo adhieren investigadores de la Universidad Nacional de La Matanza y la Universidad CAECE.

El grupo no solo posee investigadores formados sino que recluta permanentemente y forma jóvenes investigadores los cuales colaboran como docentes en las diversas cátedras que integrantes del grupo conducen.

Por otra parte, se han desarrollado experiencias de capacitación, como así también trabajos conjuntos con otras instituciones del sistema universitario, como la Universidad Autónoma de Entre Ríos y las Regionales de Concepción del Uruguay, Buenos Aires y Avellaneda de la Universidad Tecnológica Nacional.

En el plano internacional, el laboratorio CIM trabaja en distintas actividades conjuntas junto con el Departamento de Metrología Aplicada, de la Universidad Técnica de Viena, Austria – (TU-Wien – AuM) y a través de este nexo con los colegas de todo el mundo nucleados alrededor de ICPM (International Congress of Precision Machining) y CIRAS International Conference (Computational Intelligence, Robotics and Autonomous Systems)

Entendiendo que la integración es mucho más que las suma de las partes, aún cuando el equipo cuenta para este proyecto inter- y trans-disciplinario con diferentes especialistas, la dirección del proyecto y todos sus integrantes manifiestan su interés y promueven aumentar cuantitativamente y

cualitativamente la red de especialistas para multiplicar y profundizar experiencias, tanto en los planos de tecnologías duras como mecatrónica (mecánica, electrónica e informática) como en los planos de gestión organizativa, de la calidad, de la logística y ,por supuesto, pedagógica.

5. BIBLIOGRAFIA

1. P.H. Osanna, L. Si. "Multi-Functions Integrated Factory MFIF – a Model of the Future Enterprise". ## 2000. Enterprise. Conference Proceedings: Internet Device Builder Conference, Sta. Clara, May 2000
2. J.M.Bauer; R. Schmit; J.U. da Silva Gillig. "Constructive systems of robotic and Telerobotic like contributions in inspection, control, monitoring or other industrial tasks, generating telepresence". IPCM2005- Austria. Available: <http://www.multiplo.com.ar>.
3. Global International, Cooperation of collaborating, Small and Medium size Enterprises to achieve Total Quality Management. P.H.Osanna, N.M.Durakbasa, J.M.Bauer, L.Kraeuter. Vienna University of Technologie – Tu-Wien, Vienna Austria
4. Tzyh-Jong Internet Based Manufacturing technologie: intelligent remote teleoperation.
5. Proceeding of the "000 IEEE/RSJ International conference on Intelligent Robots and Systems 2000" (IROS 2000)
6. M.Bauer "Sociedad del conocimiento, robótica, derechos humanos y transformaciones en el mundo del trabajo", Plenario: KCTOS - Viena 2007 Knowledge, Creativity and Transformations of Societies.
7. Schmit, Rubén; da Silva Gillig, Julián; del Rio, Hernán (2004) "Desarrollo de una arquitectura para prototipado de robots, XiOR, Grupo de Investigación en Robótica", Workshop CAFR2004
8. A. Afjehi-Sadat, N.M. Durakbasa, P.H. Osanna, J. M. Bauer , Modern Intelligent Production Under Special Consideration Of Metrology, Nanotechnology, Quality Management And Artificial Intelligence. Vienna University Of Technologie – Austria.
9. Carretero, M. "Constructivismo y educación" Aique. Buenos Aires. 2004
10. Solbes, J.; Vilches, Amparo. "El modelo constructivista y las relaciones ciencia/ técnica/ sociedad". Enseñanza de las Ciencias: revista de investigación y experiencias didácticas [en línea], 1992, Vol. 10 Núm. 2, p. 181-186. <http://www.raco.cat/index.php/Ensenanza/article/view/39819/0> [Consulta: 01/02/09].
11. Vargas, F. "Formación por competencias". Notas de capacitación. OIT/Turín. 1996.
12. Albergucci, Roberto. "La transformación de la educación técnica en la República Argentina". En el Boletín 141. Cinterfor/OIT. Montevideo. 1997.
13. Estayno, M. y Grinsztajn, F. "Hacia un nuevo paradigma en la formación de profesionales de informática y TIC'S.". JEITICS 2005 - Primeras Jornadas de Educación en Informática y TICS en Argentina.
14. Santángelo, Horacio Nestor. "Un modelo de enseñanza no presencial, basado en nuevas tecnologías y redes de comunicación para la Universidad Tecnológica Nacional". Centro de Planeamiento Tecnológico y Teleeducación "El Centro". Rectorado U.T.N. Año 2000.

Calidad y Semántica de Objetos de Aprendizaje

Francisco Ibáñez
fibanez@iinfo.unsj.edu.ar

Emilio Ormeño*
eormeno@iinfo.unsj.edu.ar

Susana Ruíz*
sbruiz@iinfo.unsj.edu.ar

Sergio Ochoa†
pzarate@iinfo.unsj.edu.ar

César Collazos‡
sflores@iinfo.unsj.edu.ar

CONTEXTO

La línea de I+D aquí expuesta, se relaciona con los proyectos: “Medición de la Reutilización de Objetos de Aprendizaje” (E-827), “Especificación de Semántica en Metadatos de Objetos de Aprendizaje” (E-828), ambos aprobados y financiados por CICITCA. Además, posee un fuerte vínculo con el programa Proceis en donde participan la Universidad de San Juan, la Universidad de Chile, y la Universidad de Cauca, Colombia.

RESUMEN

Los Objetos de Aprendizaje (OA) son piezas de contenido didáctico digital que apoyan el logro de un único objetivo instruccional; con el fin de que un profesor, ensamblando dichas piezas, pueda construir su propio software educativo con mínimo esfuerzo. Esas piezas deberían estar diseñadas de forma que se puedan reutilizar en otras instituciones, por otros profesores, y ojalá en otros cursos. Sin embargo, lograr OA reutilizables no es trivial, en parte por la inexistencia de modelos matemáticos que midan (1) ¿cuánto depende un OA de otros?, (2) ¿en qué medida apoya un único objetivo?, y (3) ¿qué semántica requieren los metadatos de los OA para optimizar su accesibilidad? Esta línea de I+D tiene como objetivo, por un lado, obtener un modelo matemático para predecir el potencial de reutilización de un OA en base a sus características textuales, y por el otro, proponer contextos formales (ontologías) que se ensamblen con los metadatos estándares, a fin de optimizar su acceso desde repositorios.

Palabras clave: Objetos de Aprendizaje, Reutilización, Calidad, Métricas, Ontologías Web, Web Semántica.

1. INTRODUCCION

Los Objetos de Aprendizaje [1] (OA) son piezas de material didáctico digital (imágenes, documentos, ejemplos, etc.) que apoyan el logro de un único objetivo instruccional [8]; con el fin de que un profesor, al ensamblarlos tal como si de un juego de lego [16] se tratase, pueda construir su propio software educativo con mínimo esfuerzo.

Físicamente, un OA se encuentra empaquetado (por ejemplo en un archivo .ZIP) junto a texto, imágenes, enlaces, etc.; y junto a una serie de documentos que describen su naturaleza, alcance, nivel, palabras claves para su búsqueda; llamados metadatos, bajo algún estándar (por ejemplo, IMS+LOM, SCORM, etc.) [11, 12], para ser reutilizados en cualquier entorno que soporte dicho estándar.

Estos elementos tienen como fin, lograr que esas piezas sean efectivas en su uso [2, 3], es decir, procurar darle sustento a su principal atributo de calidad: su capacidad de ser reutilizados [4].

Sin embargo, construir o encontrar OAs que sean efectivamente reutilizables [5, 6], no es una tarea trivial debido, en parte, a que no se cuenta con modelos matemáticos de medición que respondan a cuestiones tales como: (1) ¿cuánto depende un OA de otros?, y (2) ¿en qué medida apoya un único objetivo instruccional?. Ambas cuestiones, que se relacionan con los conceptos de acoplamiento y cohesión de la teoría de Orientación a Objetos de la Ingeniería de Software [7, 9, 10], definen la capacidad de reutilización de un componente.

Aquí se podría decir que existen dos líneas que probablemente conduzcan a OA reutilizables. Por un lado, está la cuestión de las características textuales del OA en sí; y por el otro, el contexto para el cual la reutilización de un OA específico tiene sentido. Las secciones siguientes describen ambas problemáticas.

1.1. Naturaleza textual

Si bien esta perspectiva es comprehensiva, en el estado actual del arte, los únicos elementos de esta naturaleza que pueden ser reutilizados son cursos completos (máxima granularidad). Esto se debe a que construir piezas de menor granularidad (capítulos, secciones, etc.) es una tarea compleja. Tómese el caso, por ejemplo, de tratar de reutilizar una actividad práctica para un curso cualquiera:

- en primer lugar tenemos que el OA debe estar en el mismo idioma,
- debe tener un formato (tipo de letra, colores, etc.) similares a los del curso completo,
- debe tener un estilo de escritura similar (primera o tercera persona);

· Universidad Nacional de San Juan, Facultad Ciencias Exactas, Físicas, y Naturales, Instituto de Informática, Tel: +54 264 4265101, San Juan, Argentina

†Universidad de Chile, Facultad de Exactas, Departamento de Ciencias de la Computación, Tel: +56 2 978 4965

‡Universidad del Cauca, Tel: +57 2 8209900

- debe adaptarse a términos culturales o a una idiosincrasia específica;
- no debe tener referencias a otros OA (por ejemplo, supóngase un texto que comience diciendo: "... como se vió en el capítulo anterior..."), claramente no es reutilizable sin tener que modificarse.

Por otro lado, una granularidad intermedia, permite flexibilidad al profesor a la hora de construir un curso. Dado que un curso completo posee inmersa una estrategia ya armada. Esto sería el equivalente entre la decisión en utilizar un "software a medida" versus un "software general".

Desde la perspectiva de los OA, se podría decir que un OA es altamente reutilizable si puede ser utilizado en múltiples contextos diferentes para el cual fue concebido [13], por ejemplo, en otro país, en otra institución, por otro profesor, y en el mejor de los casos, en otro curso.

Debido a que la teoría de OAs proviene del paradigma de Orientación a Objetos de la Ingeniería de Software [7], se podría llegar a extrapolar diciendo que un OA hereda los principales atributos de calidad que todo objeto debería tener: una alta cohesión, un bajo acoplamiento, y la adhesión a una interfaz estándar con el fin de potenciar su capacidad de reutilización. Para ello, la Ingeniería de Software posee desde hace tiempo técnicas y métricas que sirven para medir tales atributos de calidad.

1.2. La Perspectiva Semántica

Los repositorios actuales de objetos de aprendizaje, tales como MERLOT [14] o CAREO [15], generalmente describen los distintos recursos didácticos existentes en la Web, almacenando registros de metadatos asociados a los objetos descritos. De este modo se garantiza una búsqueda mucho más estructurada del conocimiento existente. Al proporcionar un soporte para albergar los metadatos, estos repositorios desempeñan un importante papel de cara al futuro. No sólo los humanos pueden consultar y buscar información, sino también agentes software o sistemas LMS (Learning Management Systems) externos.

Sin embargo, para procesar la información existente en los metadatos, es necesaria la presencia de metadatos de calidad, entendiéndose por ello que cumplan unos mínimos de completión y que los datos aportados se correspondan con un esquema de metadatos preestablecido, uniforme y, a ser posible, universal. El principal problema de estos repositorios es el de carecer de un modelo conceptual que establezca qué es un objeto de aprendizaje y qué descriptores de metadatos hay asociados a cada una de las diferentes conceptualizaciones. Sin acuerdo universal sobre el modelo de metadatos a utilizar, ni certeza sobre la completión de los mismos por parte del creador del registro de metadatos, nos encontramos una ante una

grave carencia que dificulta la automatización en estos repositorios.

Los recursos de la llamada "Web Semántica", se presentan como una de las opciones más viables para agregar elementos que le brinden semántica a los OA. De esta forma, no solo los humanos podrán buscar el OA más adecuado, sino que además, al depender de una Ontología, los LMS, podrán recuperar y/o sugerir otros OA relacionados a la temática.

Sobre la Web Semántica, Hedler (et. al.) dicen [17]: "*The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. It is based on the idea of having data on the Web defined and linked such that it can be used for more effective discovery, automation, integration, and reuse across various applications*".

Quizás las frases más relevantes de la definición sean el de "significado bien definido" y el de "información en la Web definida y enlazada", es decir, "anotada", lo cual se sustancia bajo la forma de metadatos.

2. LINEAS DE INVESTIGACION y DESARROLLO

A continuación se realiza una breve revisión de los antecedentes que conforman la línea de I+D y de cómo ésta fue cambiando a lo largo del tiempo, coincidiendo con tendencias internacionales y con el nivel de comprensión que el equipo de trabajo fue adquiriendo sobre la naturaleza real de la problemática abordada.

1. La línea se inicia con el trabajo conjunto entre el Instituto de Informática y el Departamento de Ciencias de la Computación de Chile. Dicho equipo, liderado por el ahora PhD Sergio Ochoa, estaba trabajando en una técnica de desarrollo de software educativo basado en componentes.
2. La experiencia adquirida por el equipo de trabajo, se sustanció con el desarrollo del software PREDIST que se trataba de una herramienta para realizar presentaciones distribuidas. Con esta herramienta, un profesor podía exponer un tema desde un navegador a un conjunto de alumnos que podían seguir su clase a través de mensajes de chat. La herramienta fue publicada en el congreso CACIC 2000 que tuvo lugar en Ushuaia bajo el nombre "**Predist: una herramienta para presentaciones distribuidas**".

Estas nuevas ideas originaron la necesidad de un nuevo proyecto de investigación orientado específicamente al área educativa denominado "**Diseño e implementación de herramientas informáticas basadas en tecnología Web de apoyo a la docencia universitaria (HediWeb)**" el cual se realizó entre los años 2001 a 2003. En este proyecto se planteaba la necesidad de orientar los estudios tecnológicos a aquellos requeridos por el área educativa. Las siguientes actividades y publicaciones tuvieron lugar durante este proyecto.

3. Si bien la herramienta **PREDIST** dio buenos resultados, y de hecho fue utilizada para dictar dos cursos, el hecho de haber surgido desde una perspectiva puramente tecnológica, es decir, sin que medie un aspecto pedagógico-instruccional, ni mucho menos una perspectiva que tratara de reutilizar material existente; hizo que muy pronto dejara de ser utilizada por ser complicado para el profesor la preparación del material. Esto, dio nuevas ideas respecto a un planteo basado en tres principios:
 - Diseño instruccional
 - Desarrollo de software basado en componentes
 - Objetos de Aprendizaje
4. Estas nuevas ideas, basadas en la técnica **CMT (Course Development System)** desarrollada como parte de la tesis doctoral del profesor Sergio Ochoa, produjo entre los años 2000 y 2001 la primera versión de la herramienta **CDS (Course Development System)** que se trataba de una herramienta para construir cursos.
5. Debido a limitaciones tecnológicas. La herramienta CDS sufrió una reingeniería, y bajo esta nueva perspectiva, se publicó bajo el nombre "**Reusing Courseware Components**", en SEKE'02 patrocinado por ACM.

Si bien los resultados fueron más que satisfactorios, con el fin de colaborar en forma más estrecha con el equipo del Departamento de Ciencias de la Computación de la Universidad de Chile, se planteó la necesidad de que la plataforma sobre la que estaba montada **CDS**, agregara aspectos de apoyo a la colaboración. Esto dio pie al surgimiento del proyecto "**Plataformas de Trabajo Colaborativo Aplicadas a Ambientes Empresariales y de Educación**"; dirigido por el Dr. Raymundo Q. Forradellas, que tuvo lugar entre los años 2003 a 2005.

6. La nueva versión de CDS se basó en una nueva perspectiva arquitectónica basada en componentes denominada que se sustanció en una plataforma denominada **EBCDS (Enterprise Business Components Development System)**.
7. Debido a que muchos de los entornos colaborativos requieren de elementos gráficos que brinden percepción en forma sincrónica, se amplió la plataforma a través de nuevos componentes colaborativos, los cuales se implementaron bajo el nombre de **DCFrame**.
8. Algunos de esos aspectos colaborativos se orientaron hacia el área de juegos colaborativos para el área educativas, en ese ámbito, se colaboró con los doctores Sergio Ochoa, y José Pino del Departamento de Ciencias de la Computación de la Universidad de Chile en el póster "**Using games to model and evaluate collaborative learning**", el cual esta publicado en el "International Conference on Learning Sciences - Proceedings of the 6th international conference on Learning sciences" - Santa Monica, California.

Hasta ese momento se contaba con piezas de software probadas y robustas, con una arquitectura claramente definida, en la cual la herramienta CDS se ejecutaba; abarcando inclusive los aspectos colaborativos. De hecho, la introducción de la nueva plataforma planteó la necesidad de cambios en la herramienta CDS por lo que pasó a llamarse "**CDS 2.5**" (versión 2.5).

Se comenzó con la puesta en producción de la herramienta CDS, y principalmente con la medición de la usabilidad a través de cuestionarios fiables a sus usuarios. Esto dio lugar al planteo del actual proyecto: "**Medición de la Satisfacción de Usuarios de la Herramienta CDS (Course Development System) en cursos Universitarios**" dirigido por el Dr. Francisco S. Ibañez; el cual se llevó a cabo entre los años 2005 a 2007.

9. Posteriormente, parte de la plataforma se publicó bajo el nombre "**Una Plataforma de Componentes Heterogéneos para Entornos de Diseño con soporte J2EE**". En IDEAS 2006, en La Plata, Argentina.
10. Durante este período, se estrecharon aún más los lazos con el Departamento de Ciencias de la Computación de la Universidad de Chile, al trabajar en forma conjunta en el proyecto internacional "**e-Learning**" (2004-2006) financiado por el Chile-Korea IT Cooperation Center.
11. Sin dejar de lado los aspectos colaborativos de la herramienta CDS, se continuó con esta línea en forma paralela, lo cual dio lugar a la publicación de un capítulo con el nombre "**MOCET: a MOBILE Collaborative Examination Tool**" en el HCI Internacional 2007. Beijing, China. El cual se encuentra publicado en libro "**Human Interface and the Management of Information. Interacting in Information Environments**" editado por "**Springer Berlin / Heidelberg**" en 2007; ISBN N° 978-3-540-73353-9. En este trabajo, se mostró una experiencia de evaluación colaborativa utilizando dispositivos móviles, realizada con alumnos del Departamento de Ciencias de la Computación de la Universidad de Chile.
12. Lo novedoso de la filosofía planteada por la plataforma EBCDS completa (y en producción) con su documentación y especificación formal, dio lugar al "**Trabajo de tesis de maestría en informática de la Universidad Nacional de La Matanza del Magister Emilio G. Ormeño**", bajo el título "**Una Especificación de Componentes de Software Visuales y Persistentes**".

Es en este punto donde se produce una inflexión en la línea de investigación. Por lo que, si bien entre las actividades planificadas para este proyecto requerían la puesta en producción de la herramienta CDS, y de la medición de su usabilidad por parte de sus usuarios, el equipo de investigadores del instituto de informática debió replantear los objetivos debido a que no era oportuna aún su puesta en producción a la luz de una perspectiva de Objetos de Aprendizaje que hace efectivamente usable la herramienta. Es decir, se consideró que cualquier medición de la usabilidad en el estado actual de la herramienta estaría muy influenciada por una filosofía demasiado estructurada y una visión demasiado tecnológica. Este replanteo produjo dos hechos de importancia:

- En primer lugar, la creación de un nuevo programa marco para los proyectos de investigación denominado "Tecnologías de Software de Apoyo a la Educación Superior" del cual el Mg. Emilio G. Ormeño es director.
- En segundo lugar, se decidió de dejar de lado, por un momento, la producción de la herramienta CDS y

centrarse en los componentes didácticos en si, de cuya naturaleza poco se sabe.

Por ello, a principios del año 2007, se comenzó a trazar una línea de investigación, orientada principalmente a obtener información (medir objetivamente) las piezas de material didáctico denominadas Objetos de Aprendizaje.

13. En primer lugar, se planteó la necesidad de un formato estándar para el contenido. Para ello se propuso como formato estándar de los OA, el formato OpenOffice. Por motivos que se explican en la publicación de revista: **“Reutilización de Componentes de Courseware en Base a un Formato Estándar”**, Sergio F. Ochoa, Emilio G. Ormeño, en los “Anales del Instituto de Ingenieros de Chile”. Vol. 119 N°1 – Abril 2007. ISSN 0716-2340. Revista Chilena de Ingeniería.
14. Este estudio se consolidó con la creación de un repositorio de Objetos de Aprendizaje en formato OpenOffice en la Web con una interfaz WebService para su interoperabilidad. El software fue desarrollado como Trabajo Final de la carrera de programador universitario por la programadora María Laura Oviedo denominado **“Repositorio de Componentes Didácticos Reutilizables”**, el cual fue dirigido por el Mg. Emilio G. Ormeño.
15. Los OA almacenados en el repositorio antes comentado, carecían de una adhesión a algún estándar, por lo que se encaró otro trabajo final denominado **“Implementación de Metadatos en Objetos de Aprendizaje”**; de la carrera de Licenciatura en Ciencias de la Información. Desarrollado por los alumnos Víctor Rosales y Daniel Rosselot y dirigido por el Mg. Emilio G. Ormeño.
16. Posteriormente se encaró el problema de reconocer OA que tengan la menor cantidad de referencias, es decir, desde el punto de vista de la ingeniería de software, tratar de obtener piezas de contenido didáctico con bajo acoplamiento y alta cohesión. Esta problemática y posible solución quedó publicada bajo el nombre **“LORT: Una Técnica para Obtener Objetos de Aprendizaje”**, Emilio G. Ormeño, Cintia Ferrarini Oliver, Antonio Aranda, María Inés Lund, Lorena Vanesa Parra, en el Congreso de Informática del Nuevo Cuyo 2007.

3. RESULTADOS OBTENIDOS

“A CMT-based Modeling Language for Courseware Design”. Proc. of the 13th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Santiago, Chile, IEEE CS Press, pp. 565-570, 2009.

“Medición del Acoplamiento de Objetos de Aprendizaje”. CLEI 2008. XXXIV Conferencia Latinoamericana de Informática (CLEI 2008), pp. 400-408, 2008.

“Una Métrica para Determinar el Nivel de Acoplamiento entre Objetos de Aprendizaje a Partir del Análisis Semántico de su Contenido”. Revista Chilena de Ingeniería, Anales del Instituto de Ingenieros de Chile. Vol. 120 N° 1 – Abril 2008. ISSN 0716-2340.

“Medición del acoplamiento entre Objetos de Aprendizaje a partir del análisis lingüístico de su contenido”. X International Conference On Engineering and Technology on Education, 2008, Perúibe. Santos: COPEC, 2008. ISSN/ISBN: 8589120546. p. 273-278.

“Un Lenguaje para Obtener la Semántica Instruccional de Cursos”. Francisco Ibáñez, Pedro Zárate, Fabián Berón, Sergio Flores. (Abstract enviado).

“Experiencia Preliminar Incubación de Empresas Locales desarrolladoras de Software”, trabajo aceptado en el III Congreso Internacional: "El Aporte de la Universidad al Desarrollo Sostenible" – Univ. Católica de Cuyo – Instituto de Desarrollo Sostenible. San Juan. Autores: Pedro Zárate, Sergio Zapata, María Inés Lund, Emilio Ormeño.

“Una técnica de evaluación colaborativa soportada por computador para escenarios de educación superior”. Ochoa, S., Collazos, C., Bravo, G., Neyem, G., Ormeño, E., Guerrero, L., IX Congreso Internacional Interacción, Albacete (España), 9-11 Junio 2008, pp. 71-80, ISBN: 978-84-691-3871-7.

“Indicadores para la Evaluación de Diseños de Sistemas a través del Modelo de Casos de Uso en la Educación Superior”. Trabajo aceptado en el ICECE'2009 - International Conference on Engineering and Computer Education. Autores: María Inés Lund, Cintia Ferrarini Oliver, Emilio Ormeño, Sergio Zapata

4. FORMACION DE RECURSOS HUMANOS

Tesis de Grado

“Repositorio de Componentes Didácticos Reutilizables”. Desarrollado por Laura Oviedo como Trabajo final de la carrera de Programador Universitario.

“Implementación de Metadatos en Objetos de Aprendizaje”. Desarrollado por los Licenciados Víctor Rosales y Daniel Rosselot como trabajo final de la Carrera de Licenciatura en Ciencias de la Información.

“Una Ontología para clasificar Objetos de Aprendizaje del área Estadística”. En desarrollo por los alumnos Juan Chire, y Claudio Videla.

Tesis de Maestría

“Encriptación en Contextos de Colaboración”. Ing. Sergio R. Flores. Asesor: Francisco Ibáñez.

“Modelo para Datos sobre Microarreglos”. Lic. Susana B. Ruíz. Asesor: Francisco Ibáñez. Maestría en Estadística Aplicada. Universidad Nacional de Córdoba.

“Metodología para Modelar Procesos de Educación a Distancia”. Lic. María Rosalía Gema Romagnano. Asesor: Emilio Ormeño. Maestría en Informática. Universidad Nacional de La Matanza.

“Indicadores de Calidad de Modelos de Casos de Uso”. Asesor: Emilio Ormeño. Maestría en Informática de la Universidad Nacional de La Matanza.

5. BIBLIOGRAFIA

- [1] LTSC. (2000a). Learning technology standards committee website [On-line]. Available: <http://ltsc.ieee.org/>
- [2] Gibbons, A. S., Nelson, J., & Richards, R. (2000). The nature and origin of instructional objects. In D. A. Wiley (Ed.), The instructional use of learning objects. Bloomington, IN: Association for Educational Communications and Technology.
- [3] Hodgins, Wayne. (2000). Into the future [On-line]. Available: <http://www.learnativity.com/download/MP7.PDF>

- [4] Bloom, B. S. (1956). Taxonomy of educational objectives, handbook 1: Cognitive domain. New York: Longmans Green.
- [5] Wiley, D. A., South, J. B., Bassett, J., Nelson, L. M., Seawright, L. L., Peterson, T., & Monson, D. W. (1999). Three common properties of efficient online instructional support systems. The ALN Magazine, 3(2), [On-line]. Available: http://www.aln.org/alnweb/magazine/Vol3_issue2/wiley.htm
- [6] Wiley, D. A. (2000). Learning object design and sequencing theory. Unpublished doctoral dissertation, Brigham Young University. Available: <http://davidwiley.com/papers/dissertation/dissertation.pdf>
- [7] Pressman, R. S. Ph.D (1982). Software Engineering - A Practitioner's Approach - Fourth Edition.
- [8] Reigeluth, C. M. & Nelson, L. M. (1997). A new paradigm of ISD? In R. C. Branch & B. B. Minor (Eds.), Educational media and technology yearbook (Vol. 22, pp. 24-35). Englewood, CO: Libraries Unlimited.
- [9] ADL. (2000). Advanced distributed learning network website [On-line]. Available: <http://www.adlnet.org/>
- [10] ARIADNE. (2000). Alliance of remote instructional authoring and distribution networks for Europe website [On-line]. Available: <http://ariadne.unil.ch/>
- [11] IMS. (2000a). Instructional management systems project website [On-line]. Available: <http://imsproject.org/>
- [12] LOM (2000). LOM working draft v4.1 [On-line]. Available: <http://ltsc.ieee.org/doc/wg12/LOMv4.1.htm>
- [13] Gamma, E. (1995). Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Professional; 1st edition.
- [14] CAREO. Campus Alberta Repository of Educational Objects. Consultado en: (<http://www.careo.org/>) setiembre-2007.
- [15] MERLOT Multimedia Educational Resources for Learning and Online Teaching. Consultado en: (<http://www.merlot.org/merlot/index.htm>) setiembre-2007.
- [16] LEGO, Wikipedia. Consultado en: (<http://es.wikipedia.org/wiki/LEGO>) abril-2009.
- [17] Hendler, J., Berners-Lee, T., and Miller, E. Integrating Applications on the Semantic Web, 2002, <http://www.w3.org/2002/07/swint.html>

DESARROLLO DE COMPETENCIAS PARA LA GESTIÓN DE INFORMACIÓN Y CONSTRUCCIÓN DE CONOCIMIENTOS: *TICs y Nuevos Ambientes Educativos*

Mg. GONZÁLEZ DE DOÑA, Mónica Gilda

Mg. AGUADO, Laura Irene

Lic. SALVIOLI, María Carolina

Lic. CASTRO ALANIZ, Gabriel Alejandro

Departamento de Informática/ Programa Permanente de EAD/ Facultad de Ciencias Exactas, Físicas y Naturales/ Universidad Nacional de San Juan

CONTEXTO

Este proyecto se desarrolla en el marco del Programa Permanente de Educación a Distancia, en la línea de investigación Diseño Didáctico-Formación de competencias para la gestión de información en docentes y estudiantes universitarios.

El Programa Permanente de Educación a Distancia creado por Resolución 08/05 de la Facultad de Ciencias Exactas, Físicas y Naturales, constituye un espacio académico y de investigación interdisciplinario, cuyo propósito fundamental es generar experiencia, conocimientos, soporte tecnológico-comunicacional y organizativo a propuestas pedagógicas en la modalidad a distancia, como así también formar recursos humanos capacitados para implementar experiencias educativas en esa modalidad.

RESUMEN

Las TICs han modificado todas las actividades de la vida social provocando cambios sustantivos en los modos de pensar, sentir y actuar, transformando los estilos de interacción social, sobre todo en lo referente a la comunicación y, en consecuencia, a los procesos de enseñar y aprender. Este será el recorrido que asumiremos para explicar la influencia de las TICs en la educación superior, puesto que la llegada de los formatos electrónicos ha puesto en debate la articulación fundamental que supone la comunicación entre los distintos agentes participantes del sistema educativo universitario.

En el marco de la sociedad actual, denominada de la Información o Sociedad del Conocimiento, nos encontramos invadidos por un exceso de información en diferentes formatos y por una gran variedad de opciones de acceso. Se hace

imprescindible el dominio de determinadas competencias para participar activamente en este nuevo contexto de interacción social.

El fin primordial de nuestro trabajo de investigación supone diseñar estrategias que promuevan, en los aspirantes a ingresar y estudiantes de los primeros años de las carreras de la FCEFYN, la formación de competencias para la gestión de información y construcción de conocimientos, en los nuevos ambientes educativos mediados por las TICs.

Palabras claves: Competencias-TICs-Gestión de Información-Construcción de Conocimientos.

1. INTRODUCCION

Las TICs han modificado todas las actividades de la vida social provocando cambios sustantivos en los modos de pensar, sentir y actuar, transformando los estilos de interacción social, sobre todo en lo referente a la comunicación y, en consecuencia, los procesos de enseñar y aprender. Este será el recorrido que asumiremos para explicar la influencia de las TICs en la educación superior.

En el marco de la sociedad actual, denominada de la Información o Sociedad del Conocimiento, nos encontramos invadidos por un exceso de información en diferentes formatos y por una gran variedad de opciones de acceso. Se hace imprescindible el dominio de determinadas competencias para participar activamente en este nuevo contexto.

Gómez J. A. (2000) indica que más que enseñar habilidades informáticas debemos enseñar *“a valorar la información, apreciarla y comunicarla*

adecuadamente”¹. No se trata sólo de buscar información, sino de hacerlo organizadamente para facilitar su selección y análisis crítico y así lograr la generación sistemática de nuevos conocimientos.

Al respecto, en los estándares de competencias en TICs propuestos por UNESCO, se afirma que: *“Para vivir, aprender y trabajar con éxito en una sociedad cada vez más compleja, rica en información y basada en el conocimiento, los estudiantes y los docentes deben utilizar la tecnología digital con eficacia. En un contexto educativo sólido, las Tecnologías de la Información y la Comunicación (TICs) pueden ayudar a los estudiantes a adquirir las capacidades necesarias.”*²

Por su parte María Pinto³ sostiene que las exigencias de la Sociedad de la Información y del Conocimiento son, entre otras: adquirir nuevas capacidades, habilidades y competencias transversales y específicas a lo largo de los estudios de grado y postgrado; adoptar un modelo proactivo de aprendizaje significativo que favorezca el autoaprendizaje, la autonomía y la creatividad de los estudiantes; rediseñar los métodos de docencia universitaria; y concientizar a los docentes y estudiantes universitarios acerca de los nuevos retos que supone adaptar estos procesos a los nuevos ambientes educativos.

Por lo anteriormente expuesto nos preguntamos ¿cuál es el rol de la institución, del docente y del alumno en el contexto universitario actual teniendo en cuenta las exigencias de los nuevos ambientes educativos?

¿Cómo gestionan la información y construyen conocimientos los estudiantes de las carreras de la Facultad de Ciencias

Exactas, Físicas y Naturales?¿ Cómo aprenden?

¿Qué estrategias son necesarias para atender al desarrollo de competencias y aprender en los nuevos ambientes educativos?

¿Cuáles son los diseños didácticos que se requieren? ¿Qué criterios deben tenerse en cuenta en su elaboración?

En la comunidad universitaria siempre se ha valorado de manera especial la formación científica de los docentes, en cambio la formación didáctica- pedagógica ha sido un campo poco legitimado. A fin de reparar esta disociación de aspectos de implicancia mutua, nos preocupan, en estos momentos, dos problemas estrechamente relacionados: por una parte, el reconocimiento de cuestiones didácticas en el nivel universitario y por otra, una *búsqueda de reconceptualización de las cuestiones didácticas.*⁴

En este marco, cobran relevancia las acciones de mediación pedagógica y comunicacional de los contenidos, la forma de los materiales, las prácticas de aprendizaje y los modos de evaluación, acordes a las exigencias de los nuevos ambientes educativos.

Es por eso que en el marco de este proyecto nos proponemos diseñar estrategias que promuevan, en los estudiantes de los primeros años de las carreras de la FCFyN, el desarrollo de competencias para la gestión de información y construcción de conocimientos, en los nuevos ambientes educativos mediados por las TICs; como así también, la generación de lineamientos que orienten a los docentes en sus diseños didácticos a fin de asegurar el desarrollo de estas competencias.

Según el informe de PISA 2003⁵, *“el éxito de los estudiantes está ligado – entre otros factores – a su competencia para comprender lo que leen, al uso de estrategias de aprendizaje en forma amplia,*

¹ GÓMEZ J. A. (2000). *La alfabetización informacional y la biblioteca universitaria. Estrategias y modelos para enseñar a usar la información*, Murcia.

² UNESCO. *Estándares de competencias en TICs para docentes*, Londres, 2008

³ PROYECTO ALFINEEES, PINTO María, disponible en:

http://www.mariapinto.es/alfineeess/que_es.htm

⁴ Litwin, Edith, *Las configuraciones didácticas*. Editorial Paidós, Bs As 2000

⁵ OCDE, PISA Cadre d' evaluation PISA 2003- Connaissances et compétences en mathématiques.

a su capacidad para: reflexionar en forma profunda sobre sus procesos de aprendizaje, poner en juego estrategias generales de resolución de problemas y trabajar en grupo” [SIC]. En coincidencia, María Pinto propone las siguientes competencias: “Aprender a aprender, Aprender a buscar y evaluar información, Aprender a analizar, sintetizar y comunicar, Aprender a generar conocimiento, Aprender a trabajar juntos y Usar la tecnología para aprender.”⁶

Sin embargo, en este vasto pero cabal conjunto de capacidades podría reorganizarse en dos competencias dominantes: “aprender a aprender y aprender a hacer”. Así, el buen desempeño académico de los estudiantes universitarios estará relacionado significativamente con el desarrollo de competencias cognitivas y pragmáticas, respectivamente. Aunque los estudiantes posean una facilidad casi innata e intuitiva de la dimensión operativa de las TICs., no sucede lo mismo con la adquisición de las competencias para comprender, procesar y asimilar la información que las TICs han puesto a su alcance. Contrariamente, se advierte que la mayor dificultad de los docentes se presenta en la dimensión operativa de las TICs. De manera que el desencuentro entre ambas generaciones se hace patente en la educación.

De la anterior situación deviene la necesidad de realizar ajustes en lo que se refiere a los roles clave asignados a las instituciones educativas, en nuestro caso, en la Universidad. Como respuesta a esta realidad, se advierte un esfuerzo sostenido de la institución universitaria para adaptarse gradualmente al cambiante entorno social. Esto supone un replanteo de su propia razón de ser, de sus objetivos y servicios, de los sistemas de organización, de las metodologías de trabajo, de los planes de estudio y de sus programas de formación de recursos humanos.

La realidad educativa actual, pone en evidencia que todavía prevalece la supremacía de un conocimiento fragmentado según las diferentes disciplinas, y como afirma Edgar Morin, propulsor del paradigma de la complejidad, esto “... impide a menudo operar el vínculo entre las partes y las totalidades”. Por tanto es deseable que se dé paso a “un modo de conocimiento capaz de aprehender los objetos en sus contextos, sus complejidades, sus conjuntos. [...] Es necesario enseñar los métodos que permiten aprehender las relaciones mutuas y las influencias recíprocas entre las partes y el todo en un mundo complejo.”⁷

Esto significa que se debe dar paso a otras formas de aprender y de utilizar métodos que promuevan en los estudiantes el aprendizaje activo, es decir, que los motiven y logren de esta manera ser conscientes de qué y por qué necesitan adquirir ciertos conocimientos, saber qué información necesitan, localizarla, organizarla y comunicarla adecuadamente, cumpliendo así con la adquisición de las competencias comunicativas expuestas por todos los autores consignados en esta exposición.

El paradigma de la complejidad de Morin representa una ruptura epistémica y por ende una transformación fundamental en el modo de pensar, percibir, aprehender la realidad en un contexto global, en el que circulan y se interconectan (a través de las TICs) ideas, pensamientos, modos de hacer y de ser. Se trata de una realidad globalizada en la cual los sucesos, fenómenos y procesos están en una relación de interdependencia.

En consecuencia, la meta educativa será formar personas que sepan desenvolverse crítica e inteligentemente a través de redes de información, de modo que no estén indefensas intelectual y culturalmente ante las mismas. En palabras de Edgar Morin, se trata de “crear cabezas bien puestas más

⁶ PINTO María, *Op. Cit.*

⁷ Morin, E. *Los siete saberes necesarios para la educación del futuro*. Nueva Visión. Buenos Aires. 2001.

que bien llenas". Y, agregamos, se trata de formar individuos competentes en el dominio de lo hipertextual.

Esta hipertextualidad, como característica de lo complejo, no es otra cosa que el reflejo externo del proceso de aprendizaje. Ya las Teorías del Aprendizaje han hecho lo propio para explicar que las conexiones entre conceptos evidencian la existencia del desarrollo cognitivo de un sistema capaz de contenerlos. Lo cual sugiere, probablemente una creatividad potenciada en tanto que permita acceder a una mejor comprensión de la realidad.

Pero sin duda no es sino hasta la llegada de los formatos electrónicos que la hipertextualidad adquiere una presencia notable entre las formas de comunicación. La posibilidad de establecer vínculos temáticos sin limitaciones genera nuevas posibilidades expresivas en tanto permite una mayor adecuación al modo multidireccional en el que naturalmente fluye nuestro pensamiento.

No importa aquí si esta desestructuración responde al espíritu de las teorías críticas o a las problemáticas de la posmodernidad. Insistiremos sí en el aspecto pragmático de esta característica discursiva: la hipertextualidad emula nuestra forma espontánea de asociar ideas. Desarrollar ideas dentro de una red conceptual hipervinculada permite adecuar el discurso a la realidad en que las ideas se producen, con una fidelidad mucho mayor. Efectivamente, el hipertexto propicia las asociaciones conceptuales tanto por parte de autor del hipertexto como su lector, ambos, en un plano de igualdad.

En conclusión, la aplicación de la cultura de las competencias transversales supone una participación activa del estudiante en su propio proceso de aprendizaje. Deberá familiarizarse con todo lo que atañe a la INFORMACIÓN y el CONOCIMIENTO, desde su generación, organización, análisis y síntesis, hasta su evaluación, gestión y utilización, de manera que sea capaz de integrar, transformar, utilizar y transferir ambos: la información y el conocimiento

2. LINEAS DE INVESTIGACION y DESARROLLO

Nos proponemos definir el perfil de los estudiantes (ingresantes y alumnos de los primeros años de las carreras de FCEFyN.) en función del acceso, evaluación y utilización de la información, como así también, objetivar la práctica docente para redireccionar esta actividad en los nuevos entornos educativos soportados por las TICs. A partir de esta definición y de la objetivación proponemos construir y desarrollar estrategias que sugieran líneas de superación de las problemáticas corroboradas en la actualidad, y que garanticen el aprendizaje de competencias para la gestión de información que formarán parte de una propuesta didáctica pedagógica para ser utilizada en el Curso de Ingreso y que quedará a disposición de los docentes de primer año de todas las carreras.

Por ello, contextualizamos nuestro proyecto en la siguiente línea de Investigación:

-Diseño Didáctico

Sin dejar de reconocer que la formación de competencias en el docente supone un aporte a otras líneas de Investigación, entre ellas:

-Los medios, materiales o recursos de enseñanza

-Gestión Educativa

-Enseñanza y Aprendizaje en los nuevos entornos soportados por las TICs

3. RESULTADOS OBTENIDOS/ESPERADOS

Está en ejecución el procesamiento de información generada a partir de los Talleres, de las entrevistas focalizadas y de las encuestas administradas a los docentes de primer año de las carreras de la FCEFyN.

No obstante, la producción del equipo se ha difundido en los Eventos Científicos tales como:

INTERTECH'2008 International Conference on Engineering and Technology Education

Diseño de estrategias colaborativas: descripción de una experiencia de aprendizaje colaborativo con estudiantes de carreras de Informática. Ferrarini Oliver-

González de Doña. Perúibe, Brazil, March 02 - 05, 2008

ICECE'2009 VI Conferencia en Educación en Ingeniería y Computación Metodología LuCoA Descripción de una experiencia de aprendizaje colaborativo con estudiantes de las carreras de Informática de UNSJ Aprob Nov 2008
Ferrari Oliver González de Doña.
Buenos Aires, Argentina

Jornadas de Investigación Universidad Nacional de Villa María Nov 2008
Córdoba. *Pautas para orientar el proceso de mediación de prácticas pedagógicas en la modalidad EaD* González de Doña y otros.

En Publicaciones:

Tipología de Estrategias de Aprendizaje para la Comprensión Lectora en Inglés en la Modalidad EaD RED Revista de Educación a Distancia Murcia España Año VIII N° 20 30 Setiembre 2008
<http://www.um.es/ead/red/20/>

ISSN 1578-7680. González de Doña y otros
Problemática de la mediación de prácticas pedagógicas en la modalidad no presencial Propuesta de Solución.

González de Doña y otros. *Enviada a Revista Cognición 30 Octubre 2008*

4. FORMACION DE RECURSOS HUMANOS

Este proyecto contribuirá a la profundización de conocimientos en las temáticas bajo estudio, de los integrantes del proyecto y de los investigadores en formación.

Se está codirigiendo una tesis en el marco de la Maestría en Informática cuyo tema es Metodología LuCoA para el Aprendizaje Lúdico-Colaborativo en Educación Superior. Autora: Ferrari Oliver- Facultad de Ciencias Exactas Físicas y Naturales- UNSJ .

Además se prevé el asesoramiento de trabajos finales de Licenciatura en Sistemas de la Información, como así también la Capacitación de Egresados de la carrera mencionada y de las carreras Licenciatura en Letras y Lic. en Psicopedagogía, en el uso y aplicación de tecnologías Web 2.0 y

en tecnología informática como tecnología educativa.

Formación de los integrantes del Equipo de Investigación:

Doctorado en Educación : Aspirante al Título de Doctor en Educación, Proyecto de Tesis Presentado (15/12/08) y Aprobado (25/03/09) para su ejecución González de Doña. Universidad Católica de Cuyo- Universidad Católica Argentina.

Maestría en Informática: Tesis en elaboración Ferrari Oliver FCEfyN- Universidad Nacional de San Juan

Formación de Docentes-Investigadores en

Cursos de Postgrado

AN APPROACH TO READING AND WRITING SCIENTIFIC ARTICLES
Aguado Laura. FCEfyN. Universidad Nacional de San Juan. 2008

TECNOLOGÍAS WEB Y EDUCACIÓN FORMACIÓN DE COMPETENCIAS PARA LA GESTIÓN DE INFORMACIÓN González de Doña Mónica Gilda. FCEfyN. Universidad Nacional de San Juan. 2008

DISEÑO DE PRÁCTICAS PEDAGÓGICAS CON TECNOLOGÍAS WEB 2.0 PARA EAD González de Doña Mónica Gilda. FCEfyN. Universidad Nacional de San Juan. 2008

5. BIBLIOGRAFIA

-BALLESTEROS, F. *La brecha digital: El riesgo de exclusión en la Sociedad de la Información.* Madrid: Fundación Retevisión, 2002

-FAINHOLC. Beatriz. *Aportes sobre el concepto de mediación en educación a distancia,* artículo de la revista N°3 de RUEDA, Bs As 1997, pag 53

-GARCÍA ARETIO, Lorenzo. *Algunos Modelos de Educación a Distancia.* Editorial del Bened, noviembre de 2004.

-GARCÍA ARETIO, Lorenzo ; *Tipos de Ambiente en EAD,* Editorial BENED, febrero 2007 ; en <http://www.uned.es/catedraunesco-ead/editorial/p7-2-2007.pdf>

-GÓMEZ J. A. *La alfabetización informacional y la biblioteca universitaria. Estrategias y modelos para enseñar a usar la información,* Murcia. 2000

- GONZÁLEZ DE DOÑA, Mónica, Llarena, Myriam y Villodre, Silvia. *Problemática de la mediación de prácticas pedagógicas en la modalidad no presencial Propuesta de Solución*. Octubre 2008
- JOHNSON, D. y JOHNSON, R. (1999). *Aprender juntos y solos*. Aprendizaje cooperativo, competitivo e-learning. In N. Spada and P. Reimann (eds.), *Learning in Humans and Machines*, 1995.
- LITWIN, Edith. *Las configuraciones didácticas*. Editorial Paidós, Bs As 2000
- MORIN, E. *Los siete saberes necesarios para la educación del futuro*. Nueva Visión. Buenos Aires. 2001.
- OCDE, PISA Cadre d' evaluation PISA - Connaissances et compétences en mathématiques. 2003
- PROYECTO ALFINEEES, PINTO María, disponible en: http://www.mariapinto.es/alfineees/que_es.htm
- UNESCO. Estándares de competencias en TICs para docentes, Londres, 2008

**Líneas de investigación Programa
Permanente de Educación a Distancia de
la Facultad de Ciencias Exactas, Físicas y
Naturales de la UNSJ**

Mg. De Luca Adriana; Esp. Posito Rosa

Programa Permanente de Educación a
Distancia

Facultad de Ciencias Exactas, Físicas y
Naturales

Universidad Nacional de San Juan

CONTEXTO

El Programa Permanente de Educación a Distancia fue creado por Resolución 08/05 de la Facultad de Ciencias Exactas, Físicas y Naturales.

Se constituye en un espacio académico y de investigación, que se sustenta en la conformación de equipos de trabajo interdisciplinarios abocados a la tarea de generar experiencia, conocimientos, soporte tecnológico-comunicacional y organizativo a las propuestas académicas bajo la modalidad a distancia en nuestra Facultad.

El propósito:

- Construir un marco de referencia que permita sistematizar los conocimientos multidisciplinares que involucra la modalidad Educación a Distancia
- Potenciar la capacidad tecnológica de la FCEFyn manifiesta en recursos humanos y equipamiento orientados a la Educación a Distancia

Sus objetivos son:

- Diseñar un marco referencial normativo para actividades de Educación a distancia
- Identificar modelos educativos sobre los que ha de sustentarse la oferta Educativa bajo la modalidad a distancia
- Realizar estudios sobre la factibilidad y la demanda de ofertas a distancia
- Realizar estudios sobre el perfil de los alumnos de un sistema EAD y de los graduados potenciales
- Relevar y analizar las características y particularidades de los diferentes

soportes tecnológicos disponibles para Educación a Distancia

- Diseñar, producir y evaluar materiales específicos para la modalidad Educación a Distancia tanto desde lo pedagógico como desde lo tecnológico
- Identificar modelos de gestión pertinentes a la modalidad EAD
- Realizar asesoramiento científico-tecnológico en la problemática de la Educación a Distancia
- Formar recursos humanos capacitados para llevar adelante la modalidad EAD experiencias en la modalidad a distancia desarrolladas en nuestra institución.

El Programa Permanente de Educación a Distancia cristaliza y sistematiza investigaciones y procesos realizados en torno a la Educación a Distancia en el ámbito de esta institución que se vienen realizando a partir del año 2000.

RESUMEN

Esta ponencia se centra en la descripción de las acciones del Programa Permanente de Educación a Distancia de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan. Este centro de investigación y producción fue creado a partir de una resolución del año 2005 y aglutina el trabajo de investigadores de esta institución en torno a la Educación a Distancia, que se reconoce como un objeto de conocimiento complejo y multidimensional.

Se presentan las variadas líneas de investigación que dicho objeto supone, los resultados obtenidos y las acciones de transferencia iniciadas y previstas.

Palabras clave:

Educación a Distancia- Objeto de estudio multidisciplinar- Tecnología Educativa

1. INTRODUCCION

La modalidad de Educación a Distancia se ha desarrollado en los últimos años de una manera particularmente acelerada, especialmente en función de las posibilidades que ofrecen las TIC. Sin

embargo es una modalidad a la que se le reconocen al menos 150 años de historia a lo largo de los cuales ha ido tomando nuevas configuraciones y ofreciendo diferentes posibilidades, pero en cada una de ellas el rol que cumplió la Tecnología fue clave para posibilitar su desarrollo.

Para Adell, las nuevas tecnologías no sólo van a incorporarse a la formación como contenidos a aprender o como destrezas a adquirir. Serán utilizadas de modo creciente como medio de comunicación al servicio de la formación, es decir, como entornos a través de los cuales tendrán lugar procesos de enseñanza/aprendizaje. Como señala Martínez (1996, pág 111), "en los procesos de enseñanza/aprendizaje, como prácticamente en la totalidad de los procesos de comunicación, pueden darse diferentes situaciones espacio-temporales, tanto en la relación profesor-alumno, como en relación a los contenidos". Las aulas virtuales, la educación en línea, a través de redes informáticas, es una forma emergente de proporcionar conocimientos y habilidades a amplios sectores de la población. Los sistemas asíncronos de comunicación mediada por ordenador proporcionarán la flexibilidad temporal necesaria a las actividades para que puedan acceder a la formación aquellas personas con dificultades para asistir regularmente a las instituciones educativas presenciales debido a sus obligaciones laborales, familiares o personales¹.

La educación a distancia es un objeto de conocimiento complejo y multidimensional que requiere un abordaje multidisciplinar que permita develar su naturaleza. Es dable observar esta naturaleza compleja en cualquiera de las conceptualizaciones que se hace para precisar el término.

¹Jordi Adell; Tendencias en educación en la sociedad de las tecnologías de la información; Educec Revista de Tecnología Educativa:
<http://www.uib.es/depart/gte/edutec-revelec7/revelec7.html>; consultado noviembre 2007.

En la Normativa vigente en nuestro país(Decreto 1717, MCYE, 2004), se denomina Educación a Distancia" el proceso de enseñanza aprendizaje que no requiere la presencia física de; alumno en aulas u otras dependencias en las que se brindan servicios educativos, salvo para trámites administrativos, reuniones informativas, prácticas sujetas a supervisión, consultas tutoriales y exámenes parciales o finales de acreditación, siempre que se empleen materiales y recursos tecnológicos especialmente desarrollados para obviar dicha presencia y se cuente con una organización académica y un sistema de gestión y evaluación específico diseñado para tal fin.(1716 MCE)

En todas las conceptualizaciones aparecen dimensiones que recortan la Educación a Distancia como objeto de estudio y la identifican:

- No coincidencia espacio-temporal de los actores
- Como correlato de la anterior: mediatización de la relación a través de tecnologías de diferente generación
- Rol jerarquizado de los materiales de estudio
- Soporte tutorial al alumno que autogestiona su aprendizaje
- Preocupación por el diseño didáctico
- Organización institucional que sustente los procesos involucrados

La terminología que remite al concepto de educación a distancia es amplia y variada: se la reconoce también como "educación no presencial", "educación semipresencial" "educación abierta", "educación flexible", "educación asistida", "educación virtual", "educación on line", "e-learning"...

El proceso de Enseñanza y de Aprendizaje se encuentra mediado por tecnología de diferentes características. La mediación tecnológica – mediación en el sentido de tender puentes- tiene como objetivo "reducir la distancia", ha ido variando a lo largo del tiempo en función de

las innovaciones tecnológicas. Así es que se reconocen tres etapas claramente definidas o “generaciones” tal como las denominan Roger Batro y Garrison teniendo en cuenta el soporte tecnológico.

La **primera generación** corresponde al periodo en que el medio que se utiliza es la “correspondencia” y el material que realiza la propuesta educativa es impreso; la imprenta es la tecnología que permite la generación de los materiales, y el correo postal el que facilita el acceso a los mismos. Esta generación de la EAD comienza a fines del siglo XIX, y principios del XX.

La **Segunda generación** de EAD comienza a hacer uso de los medios masivos de comunicación, complementados por material impreso. (García Aretio 2002) denomina a esta etapa “Enseñanza Multimedia” por la utilización de diversos medios, recursos escritos, visuales, audiovisuales, la creación de la “Open University” en Gran Bretaña puede ser considerado el paradigma de esta generación

La Educación a Distancia de **3^o generación**, es aquella que se desarrolla a partir de la aparición de las nuevas tecnologías de comunicación e información (TIC) que en muchos casos desplazan y en otros complementan los soportes impresos y analógicos. Para García Aretio, el inicio real de esta etapa se concreta a mediados de 1980 y se caracteriza por la Educación Telemática.” La integración de las telecomunicaciones con otros medios educativos, mediante la informática definen esta tercera etapa. Se apoya en el uso cada vez más generalizado del ordenador personal y de las acciones realizadas en programas flexibles de Enseñanza Asistida por Ordenador (EAO) y de sistemas multimedia (hipertexto, hipermedia, etc). Se potencian en esta generación las emisiones de radio y televisión, la audioteleconferencia y la videoconferencia. Se establece algo así como un anillo o malla de comunicaciones al que cada actor del hecho educativo accede desde su propio lugar al resto de sectores con los que debe

relacionarse. La inmediatez y la agilidad, la verticalidad y la horizontalidad se hacen presentes en el tráfico de comunicaciones” (García Aretio,2002)

García Aretio(2002) afirma que la relación docente/alumno en un proceso de aprendizaje a distancia es una “conversación didáctica guiada”, donde los materiales de estudio favorecen la interacción del alumno con el conocimiento; “A través de un **diálogo simulado y asíncrono** se establece la primera comunicación de doble vía entre la institución que enseña a través de los materiales y el estudiante que pretende aprender.

La participación del docente es fundamental para favorecer instancias de interacción con los alumnos de tal manera de acompañarlos en su camino de aprendizaje.

El rol que cumple el **docente**, así como el del **tutor** en el seguimiento y orientación de los aprendizajes se considera vital en este tipo de relaciones ya que puede acarrear una merma importante en la motivación del alumno y por lo tanto orientarlo al abandono o por el contrario su accionar puede incentivar en el alumno una actitud pro-activa que impulse sus aprendizajes y que por lo tanto aumente sus posibilidades de permanencia en el sistema.

El rol que cumple el estudiante en esta modalidad es clave para que el aprendizaje se concrete. De él depende el momento, la estrategia y el tiempo que va a dedicar a interactuar con el objeto de estudio. Se requiere de él una participación activa, con alto grado de responsabilidad y compromiso. Deberá disponer de cualificaciones referidas al *aprender a aprender*.

En la “conversación didáctica guiada” los **materiales** son fundamentales para que se produzca de manera exitosa el aprendizaje. El material de esta modalidad tiene características particulares como “orientadores” del aprendizaje y por tanto deben ser realizados con ese fin. No se trata tan sólo de “digitalizar” o “compensar” bibliografía, sino que se trata de brindar al

alumno las mejores posibilidades para que pueda por sí solo ir recorriendo el camino del conocimiento. En principio, unos materiales ideados para la enseñanza a distancia deben representar algún tipo de diálogo con el destinatario de los mensajes. Pero será un diálogo simulado, nunca real, y por la propia naturaleza del soporte, de carácter asíncrono, es decir, sin que coincidan en el tiempo la producción o emisión del mensaje, con la respuesta del receptor..."Pues bien, estos materiales una vez **distribuidos** son estudiados por el alumno de forma relativamente **independiente** y autónoma. Es el estudiante el que decide *si desea estudiar o no y cuándo hacerlo*. Decidirá también sobre el *ritmo y pasos que habrá de seguir, los objetivos y medios, la forma de cómo estudiar*. Independencia, por tanto, no sólo con respecto al espacio y al tiempo, sino también en su potencialidad de independencia en el control y dirección del aprendizaje. Un estudiante de enseñanza a distancia tiene un amplio campo de **independencia**, contrariamente a lo que sucede con la dependencia o supeditación al profesor, propia de un estudiante presencial." (García Aretio, Lorenzo, 2002; pp.111)

El material adecuado para esta modalidad debe estar preparado de tal manera de propender a generar aprendizajes endo-dirigidos, desde una visión que supere lo meramente disciplinar y que favorezca la formación de sujetos con pensamiento sistémico.

En esa comunicación didáctica cumplen también un rol muy importante las tecnologías de comunicación como por ejemplo el teléfono, el chat, el correo electrónico

Otra dimensión que necesariamente ha de considerarse cuando se refiere a la modalidad a distancia es que requiere para su correcto funcionamiento de una organización que asegure mecanismos de una gestión, de administración específicos, particulares a la modalidad que son

notablemente diferentes respecto de la modalidad presencial.

2. LINEAS DE INVESTIGACION y DESARROLLO

El Programa Permanente de Educación a Distancia ha definido las siguientes líneas de investigación que no se consideran exhaustivas para el abordaje de este objeto de estudio complejo:

- Tecnología Educativa
- Los medios, materiales o recursos de enseñanza
- Modelos pedagógicos que sustentan la Educación a Distancia
- Diseño Didáctico
- Enseñanza y Aprendizaje en la Educación a Distancia
- Calidad en la oferta Académica en modalidad a distancia
- Gestión tecnológica
- Gestión Educativa.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados se pueden agrupar en torno a cuatro ejes. Por razones de espacio, en cada uno de ellos se hace una presentación general de los mismos:

• Investigación:

Publicaciones. Formación de Recursos Humanos

Los equipos de investigación que conforman este Programa han contribuido a la comunidad científica internacional con variadas publicaciones en

Revista de Educación Superior ANUIES, México

Revista Tecnología y Comunicación Educativa, ILCE, México

Revista Formación Universitaria, Chile

Revista de Educación a Distancia (RED) N° 20. Murcia, España.

Revista Iberoamericana de Educación (RIE) Anales de Congresos de Educación a Distancia Costa Rica

Anales de Congreso de México, RUEDA, TE y ET

• Desarrollo:

Adaptación de la distribución de GNU/Linux "Adrenalinux" como herramienta de trabajo para el taller de

capacitación en Moodle. La distribución puede descargarse desde <http://adrenalinix.org/iso/adremoodle/adrena-moodle061207.iso>

Instalación y configuración de la plataforma educativa Moodle en el dominio <http://www.exactasad.org/>

Plataforma Moodle en el dominio <http://www.adrenalinix.org/adrenavirtual/>
Desarrollo de materiales didácticos en diferentes formatos y utilizando soportes diferentes: Webquest, wiki, Blog de cátedras, Podcast, Multimedia,

Animaciones y video

Tecnología Educativa:

MACCAD: Metodología para la Evaluación de la Calidad de Estrategias Didácticas de Cursos a Distancias

SICED: Software para evaluar calidad de cursos a distancia en <http://www.gc.unsj-cuim.edu.ar/calidad/>-

Manual de Procesos

Manual de Organización y Funciones

- **Capacitación:** Cursos, Talleres y Seminarios

El Programa Permanente de Educación a Distancia ofrece a la comunidad universitaria Cursos de Capacitación y Actualización: Diseño de Prácticas Pedagógicas para Educación a Distancia con Tecnologías Web: Tecnologías Web y Educación: Formación de Competencias para la Gestión de Información.

Asimismo realiza periódicamente Seminarios y Talleres internos de difusión y actualización.

- **Asesoramiento:**

Se brinda asesoramiento a las cátedras que desean implementar la modalidad a distancia.

Asesoramiento Metodológico a Recursos Humanos que cursan carreras de postgrado y /o en los procesos de producción de comunicaciones científicas .

- **Producción:**

Desarrollo tecnológico transferido 23 cátedras y cursos de Posgrado y Grado se han diseñado e implementado.

4. FORMACION DE RECURSOS HUMANOS

- Miembros de este equipo dirigen/codirigen 3 tesis de la Maestría tecnología Informática Aplicada a la educación de la UNLP Estado: Proyecto de Tesis presentado y aprobado
- Miembros de este equipo dirigen 2 tesis de Maestría en Informática, Universidad Nacional de La Matanza Estado Ante Proyecto de Tesis presentado
- Miembros de este equipo han codirigido 3 tesis de Maestría en Educación SicoInformática de la Universidad Nacional de Lomas de Zamora ya aprobadas.
- Miembros de este equipo han dirigido 2 Tesis Maestría en Informática, Universidad Nacional de La Matanza que ya han sido aprobadas.
- Tesis de grado dirigidas: 5
- Tesis de grado en proceso: 1
- Títulos de Maestría obtenidos: 3
- Dirección de Becarios nacionales: 5
- Dirección de Becarios internacionales: 1

Se prevé fortalecer los vínculos con otras Universidades, como el que hoy sostenemos a través de Miembro de este equipo que es la Coordinadora Adjunta de la Maestría en Informática, sede San Juan, para la Maestría en Informática, Universidad Nacional de La Matanza.

5. BIBLIOGRAFIA

Assad Claudio, El lenguaje audiovisual, Material de estudio. Curso Comunicación y Lenguaje, PROED Cba, 2007

Bartolomé Antonio; Lenguaje Audiovisual-Mundo Audiovisual, Universitat de Barcelona

Fernández Diez Federico y Martínez Abadía José. "Manual básico de lenguaje y narrativa audiovisual". Cap. 2

Gubern Roman. "Del bisonte a la realidad virtual. La escena y el laberinto". Ed. Anagrama, Barcelona

Marqués Graells Pere. Introducción al lenguaje Audiovisual, 1995 (última revisión 2004)

García Aretio, Lorenzo; La Educación a Distancia, De la teoría a la práctica, Barcelona, Ariel Educación 2002,328pp

Litwin Edith; Tecnologías Educativas en tiempos de Internet; Buenos Aires, Amorrortu Editores, 1º Ed. 2005, 256 pp.

Burbules N y Callister T (h); Educación: riesgos y promesas de las nuevas tecnologías de la información; Barcelona, Granica,2001, 304 pp

Casas Armengol, Miguel; Tendencias Actuales e innovaciones en la educación superior a distancia. Cabero Julio, Bases Pedagógicas del E-learning; Revista Universidad y Sociedad del Conocimiento, vol 3, nº 1 abril 2006. UOC; (consultado febrero 2007).

Potencialidad y restricciones en Latinoamérica, Agenda Académica, volumen 7, año 2000.

García Aretio, Lorenzo; La Educación a Distancia, De la teoría a la práctica, Barcelona, Ariel Educación 2002,328pp

García Aretio,L (2001) Historia de la Educación a Distancia; Universidad Nacional de Educación a Distancia (UNED) <http://www.utpl.edu.ed/ried/images/pdfs/vol2-1/historia.pdf>, (consultado octubre 2004)

Garmendia, E; Juric,J,;Malvassi,S (1999), La Educación a distancia, ¿Una modalidad educativa innovadora?; Documento Base. Curso de postgrado y de extensión: “Bases conceptuales y Modelos organizativos a distancia” dictado con modalidad virtual. Sistema de Educación Abierta y a distancia-SEAD, Universidad Nacional de Mar del Plata, <http://www.mdp.edu.ar/uabierta> (consultado marzo 2008)

Jordi Adell; Tendencias en educación en la sociedad de las tecnologías de la información; Edutec Revista de Tecnología Educativa: <http://www.uib.es/depart/gte/edutec-revelec7/revelec7.html>; consultado noviembre 2007.

Perkins D.N; La persona - más: una visión distribuida del pensamiento y el

aprendizaje; en Cogniciones distribuidas, Consideraciones psicológicas y educativas, Buenos Aires, Amorrortu Editores.

Salomon Salomón, Gabriel (2001) Cogniciones distribuidas. Amorrortu ediciones. Bs. As.

Area Moreira, M. (1990). Los medios de enseñanza: Conceptualización y tipología. Documento inédito elaborado para la Asignatura Tecnología Educativa, Universidad de La Laguna. Web de tecnología educativa. Disponible en Internet:

http://www.uclm.es/PROFESORADO/RICARDO/Clasificaciones_medios/doc_ConceptMed.html

Avances en el Proyecto Software para Aprendizaje y Trabajo Colaborativo

Laura Sánchez, Lidia López, Jorge Rodríguez, Jorge Sznek, Ana Carolina Alonso de Armiño,
Eduardo Grosclaude, Javier Balladini, Claudio Zanellato, Claudia Rozas
{lmsanchez, lidiamlopez, jrodrig05, jsznez, anacarolinaalonsode, eduardo.grosclaude,
javier.balladini, cipotano, claurozaz}@gmail.com

Departamento de Ciencia de la Computación – Facultad de Economía y Administración
Universidad Nacional del Comahue

CONTEXTO

En el proyecto de investigación Software para Aprendizaje y Trabajo Colaborativo que se desarrolla en el Departamento de Ciencias de la Computación de la Universidad Nacional del Comahue, concurren varias líneas de investigación cuyos objetivos incluyen desde los enfoques teóricos hasta las aplicaciones que soportan mediante computación y comunicaciones, el trabajo y el aprendizaje colaborativos. Cada línea desarrolla su tarea en tres espacios: modelos y abstracciones, desarrollo de prototipos e implantaciones de tecnología útil en el medio.

RESUMEN

En el presente trabajo se presenta el tema actual de investigación y/o desarrollo de cada línea con referencias a los resultados obtenidos y publicados con anterioridad.

Palabras clave: *Sistemas colaborativos, TICs, visualización, seguridad*

1. INTRODUCCION

Se presentan a continuación cinco de las líneas de investigación que continúan y expanden líneas presentadas en años anteriores o que inician nuevas investigaciones. Estas son:

- Soporte TIC para la elaboración de ponencias en los EPIEM ¹
- Incorporación de nuevas herramientas a las prácticas docentes mediante PEDCO ²
- Visualización y análisis de las consultas realizadas por los usuarios de PEDCO.
- Seguridad en ambientes virtuales
- Video bajo demanda

¹ EPIEM: Los Encuentros de Profesores de Informática de Enseñanza Media se realizan en la Universidad del Comahue desde 2002 y brindan, entre otros, un espacio para la socialización de experiencias de los docentes participantes.

² PEDCO: Plataforma Moodle de Educación a Distancia de la Universidad del Comahue para e-learning y b-learning.

2. Líneas de Investigación y Desarrollo

2.1 Soporte TIC para la elaboración de ponencias en los EPIEM

Los Proyectos Educativos Colaborativos Soportados por Computadora – PECSCs – [1] se presentan como enfoque didáctico general que tiene la posibilidad de aportar lineamientos metodológicos para el desarrollo de experiencias educativas mediadas por TICs.

Este enfoque se estructura metodológica-mente a partir de los conceptos de aprendizaje colaborativo como organizador de la interacción social, abordaje por proyectos como estructurante del proceso de aprendizaje y construcción colectiva que enriquece las construcciones y posibilita la integración de múltiples puntos de vista en construcciones de mayor envergadura y calidad, aumentando la apropiación, posibilitando la trascendencia del espacio áulico y maximizando los procesos dialógicos.

En lo que respecta al esquema de colaboración se identifican tres espacios:

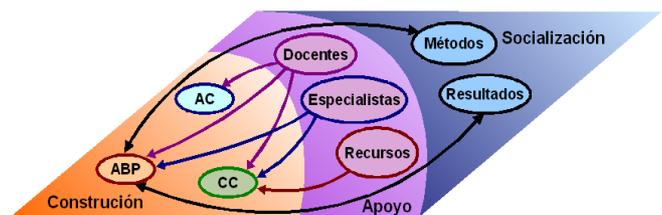


Figura 1 – Espacios e interacciones

El espacio de construcción - colaboración primaria - en el que interactúan grupos de docentes; el espacio de apoyo - colaboración cercana - donde participan otros docentes, especialistas y asesores externos correspondientes a la temática del proyecto, que aportan solidez disciplinar; y el espacio de socialización - colaboración externa - donde las producciones son socializadas hacia el espacio social al que son relevantes posibilitando la concreción de instancias meta-cognitivas, el reconocimiento del docente de escuela media como sujeto socialmente activo y autor y favoreciendo la colaboración externa.

Las tecnologías de Groupware soportan la colaboración, comunicación y coordinación y los métodos de CSCL³ aportan las técnicas para el planeamiento de la interacción.

Proveen:

- Herramientas tecnológicas para colaboración, comunicación y coordinación.
- Esquemas de colaboración para planear la interacción.

Posibilitan:

- La colaboración temprana con el grupo de apoyo fortaleciendo el seguimiento del proceso.
- La documentación de la interacción, lo que abre la posibilidad de análisis con fines científicos.

Fortalecen:

- Los procesos comunicacionales, en particular los relacionados a la expresión escrita.
- Los procesos de formalización de conocimientos i.e. al tener que escribir para que otro lo entienda se presenta la necesidad de depurar el discurso y depurando el discurso se explicitan debilidades, se articulan ideas, se fortalecen saberes y cierran círculos.

Las interacciones mostradas en la Figura 1, deben ser soportadas por TICs. En este trabajo se presentan algunas de esas interacciones sobre las que se investiga y desarrolla.

Año a año, a los docentes de informática de las escuelas medias de la región se les ofrece un espacio en el que pueden socializar sus experiencias e innovaciones y participen en la formulación de producciones teóricas en los EPIEM. Se los invita a socializar métodos y resultados: Espacio de Socialización de la Figura 1.

Es grande la dificultad que encuentran para estas presentaciones, principalmente debido a que no se reconocen como autores, a la hora de escribir un texto de divulgación referido a su experiencia en la formulación de un PECSC.

En este contexto se propone la definición de una herramienta que actúe como andamiaje para los procesos de producción de textos de divulgación y la formulación de proyectos educativos.

Esta definición está orientada por el enfoque propuesto para el desarrollo de PECSCs, es decir que la producción de un texto de divulgación, referido a las conclusiones desprendidas de la ejecución de un PECSC, o la formulación de un proyecto educativo es considerado como:

³ CSCL: Computer Supported Cooperative Learning

Un proyecto: construir un texto de divulgación.

Un trabajo colaborativo: un grupo de docentes podrán participar de la producción.

Una construcción colectiva: el grupo de docentes en diálogo con especialistas diversos construyen un texto.

Esta producción es apoyada por especialistas en diversas disciplinas para que finalmente luego de varios procesos de reescritura, el texto logrado participe de espacios de divulgación, en particular pensamos en EPIEM.

La herramienta que estamos desarrollando comprende:

Un espacio para los docentes autores y co-autores: con formatos adecuados a cada tipología de texto i.e. relato de experiencia, formulación de proyecto, etc.

Un espacio para colaboradores: Revisores de la presentación: Comité de publicaciones, revisores, otros docentes, etc.

Un espacio público: para la publicación en Internet de las producciones finales que habiendo pasado por el proceso de re-escritura son finalmente aceptadas en EPIEM.

2.2 Incorporación de nuevas herramientas a las prácticas docentes soportadas por PEDCO

La plataforma PEDCO [2] brinda soporte a las carreras presenciales que dictamos en la Universidad Nacional del Comahue, proveyendo la posibilidad de experiencias e-learning y b-learnig.

Surgen así nuevas interacciones entre docentes y estudiantes al tener acceso a nuevos soportes para la interacción, la experimentación, el trabajo colaborativo, el trabajo autónomo, el trabajo a distancia, favoreciendo la producción colectiva de conocimiento.

A través del uso de esta plataforma ha sido posible el diseño y desarrollo de diversas experiencias educativas, con el objetivo brindar a los estudiantes mejores y alternativos espacios de aprendizaje i.e. Tutorías, foros, entrega de trabajos prácticos, chat, etc.

Estos son algunos ejemplos de las tareas que se han ido desarrollando y que se siguen proponiendo y evaluando en el marco de nuestro proyecto de investigación.

Así como las plataformas educativas dan soporte y promueven otra forma de trabajo con los estudiantes, hay un gran número de nuevas tecnologías que pueden ser incluidas tanto en el ámbito del aula en educación presencial como en la modalidad a distancia. En la actualidad nos enfocamos en la articulación de PEDCO con la implementación de laboratorios remotos [3], articulación con el sistema SIU WEB [4] e incorporación de semántica [5], lo cual permitirá que los estudiantes accedan con más flexibilidad a la información.

2.3 Visualización y análisis de las consultas realizadas por los usuarios de PEDCO

Las técnicas de visualización multidimensional de datos proveen de una representación que ayuda en el análisis del comportamiento de los usuarios de la plataforma PEDCO.

En el contexto de la plataforma educativa diseñada en Moodle que tiene como principio permitir el desarrollo de experiencias educativas con una filosofía solidaria, se han analizado las representaciones gráficas estándar que proporciona la herramienta.

La base de datos interna de Moodle almacena un gran archivo log de datos de las actividades de los alumnos en un curso y brinda un reporte de actividad para monitorear al alumno que habilita al instructor visualizar datos estadísticos tal como el número de accesos realizados por el estudiante a cada recurso, la historia de las páginas visitadas, el número de hits para cada día del curso. El objetivo de esta herramienta es que los instructores puedan utilizar esta información para monitoreo y para identificar problemas potenciales [5].

Además las técnicas de visualización de datos se pueden aplicar a las consultas que realizan los usuarios. En este sentido la Web Semántica puede proveer ciertas ventajas, orientadas principalmente a la clasificación del conocimiento en ambientes de aprendizaje para brindar a los docentes y estudiantes otro tipo de información que puede ser de gran utilidad [6].

Para ello, se ha comenzado a estudiar la tecnología peer-to-peer (P2P) que se refiere a un esquema de comunicaciones que permite que una computadora en red pueda interactuar con otra sin necesidad de la intervención de un servidor central. Bajo este modelo, cada parte involucrada tiene las mismas capacidades y puede iniciar una sesión de comunicación en cualquier momento. Las arquitecturas P2P se usan para una variedad de aplicaciones diferentes, entre las cuales se tienen:

- Comunicación y Colaboración. Esta categoría incluye sistemas que proveen la infraestructura para facilitar la comunicación y la colaboración directa, normalmente en tiempo real, entre las computadoras a la par. En esta categoría se incluyen los chat y las aplicaciones de la mensajería instantáneas.

– Distribución de contenido. La mayoría de los sistemas actuales P2P cae dentro de esta categoría, ya que incluyen sistemas e infraestructuras diseñadas para compartir de medios digitales y otros datos entre los usuarios. El rango de sistemas de distribución de contenidos va desde aplicaciones para la simple distribución directa de archivos a sistemas más sofisticados que crean un medio del almacenamiento distribuido para publicar, buscar, actualizar y recuperar datos de manera segura, eficaz, organizada e indexada. Algunos ejemplos son: Ares, Gnutella, Kazaa.[7]

Uno de los inconvenientes de los sistemas P2P actuales son sus capacidades de consulta. Muchos de los sistemas sólo soportan consultas de palabras claves planas que sufren la pérdida de convenciones de nombre y la ambigüedad del lenguaje natural.

El uso de la Web Semántica permite consultas más expresivas y detalladas donde los metadatos – datos de los datos - juegan un rol central en el esfuerzo de proveer técnicas de búsqueda que van más allá del macheo de cadenas. Los metadatos basados en ontología –representación de un conjunto de conceptos dentro de un dominio y las relaciones entre ellos- facilitan el acceso al conocimiento del dominio.

Las técnicas de Visualización de la Información pueden ayudar a gestionar la complejidad que proviene del uso de conocimiento conceptual en compartir la información en los sistemas P2P [8].

2.4 Seguridad en ambientes virtuales

Honeynets en ambientes virtuales:

Las honeynets [9] surgen como una solución de seguridad diseñada para ser sondeada, atacada y comprometida por un hipotético intruso. Es una herramienta compuesta por: entornos de redes; conjuntos de aplicaciones de monitoreo y almacenamiento de eventos; etc. que está dispuesta a recibir ataques logrando mantener un ambiente controlado y propicio para el estudio de los eventos ocurridos. Luego, mediante el análisis posterior de dichos sucesos, es posible comprender los objetivos, tácticas e intereses que tienen los atacantes para el entorno propuesto.

Las honeynets virtuales [10] son un concepto relativamente nuevo. La idea consiste en combinar todos los elementos físicos de una honeynet dentro de una única computadora, utilizando para ello software de virtualización.

Se han implementado dos honeynets virtuales con sus correspondientes herramientas de seguridad sobre diferentes topologías de red, a los efectos de estudiar en cada caso un conjunto de ataques previamente seleccionado.

El software utilizado en la implementación es:

- QEMU: para crear máquinas virtuales [11].
- VDE-SWITCH: (Virtual Distributed Ethernet) para efectuar la intercomunicación entre las máquinas virtuales y la máquina física.
- SNORT: IDS de red [12].
- OSSEC: IDS de host [13]
- MYSQL: motor de base de datos donde se almacena toda la actividad registrada por los IDS [14].
- APACHE: servidor web

2.5 Video bajo demanda

Esta línea de investigación tiene por objetivo realizar aplicaciones educativas de video bajo demanda. Se continúa con investigaciones previamente realizadas [15] [16] al haber podido reincorporar recientemente a nuestro proyecto de investigación al Dr. Javier Balladini.

3. FORMACION DE RECURSOS HUMANOS

Uno de los integrantes del proyecto se encuentra en etapa de tesis de magíster y tres de sus integrantes en etapa de tesis o créditos finales para doctorados.

4. BIBLIOGRAFIA

- [1] L. Sánchez, J. Rodríguez – Modelo de Colaboración para PECSCs – Latineduca, 2005 <http://www.latineduca2005.com>
- [2] C.Fracchia, A.Alonso de Armiño. “PEDCO (Plataforma de Educación a Distancia Universidad Nacional del Comahue)”. Workshop de Tecnología Informática aplicada en Educación (WTIE). Congreso Argentino de Cs. de la Computación. Universidad Nacional de La Matanza , San Justo - Buenos Aires. Octubre 2004.

[3] L.Bertogna, et.al. “Infraestructura para laboratorios de acceso remoto”. CACiC 2007

[4] <http://siufaea.uncoma.edu.ar/>

[5] L.M.López, A.Alonso de Armiño. “Incorporación de Semántica en plataformas para e-learning”. Teyet 2008

[6] L.M.López “La Web Semántica como herramienta para e-learning” WICC 2008

[7] A Metadata Model for Semantics-Based Peer-to-Peer Systems - Jeen Broekstra, Marc Ehrig, Peter Haase, Frank van Harmelen, Arjohn Kampman, MartaSabou, Ronny Siebes, Steffen Staab, Heiner Stuckenschmidt, Christoph Tempich: Workshop on Semantics in Peer-to-Peer and Grid Computing at the 12th International World Wide Web Conference in Budapest, 2003.

[8] Visualizing RDF Data for P2P Information Sharing – Ouwerkerk A., Stuckenschmidt – AI Department – Vrije Universiteit Amsterdam

[9] Addison-Wesley. Know Your Enemy – Learning about security threats -. The Honeynet Project.

[10] Gallego Eduardo, López de Vergara Jorge E. Aprendiendo del atacante. Dpto. de Ingeniería de Sistemas Telemáticos, Universidad Politécnica de Madrid

[11] www.qemu.org

[12] www.snort.org

[13] www.ossec.net

[14] www.mysql.com

[15] J. Balladini, L. Souza, R Suppi. “A network scheduler for an adaptive VoD Server”. In E-Business and Telecommunication Networks, Communications in Computer and Information Science, volume 9, pages 237-251. Springer-Verlag Berlin Heidelberg, 2008.

[16] J. Balladini, L. Souza, R. Suppi. “Un planificador de canales lógicos para un servidor de VoD en internet. CACiC 2006.

ESTUDIO DE MUNDOS VIRTUALES PARA MEJORAR EL APRENDIZAJE

Pérez Cota, Manuel

**Área de Lenguajes y Sistemas Informáticos/Departamento de Informática
Escuela Universitaria de Ingeniería técnica Industrial
Universidad de Vigo**

García Neder, Héctor

**Departamento de Informática
Facultad Regional Río Grande
Universidad Tecnológica Nacional**

Maldonado, Calixto

Grosso, Mario

Barrera, Diego

Marciszack, Marcelo

**Laboratorio de Investigación de Software/Dep. Ing. en Sistemas de Información/
Facultad Regional Córdoba
Universidad Tecnológica Nacional**

Santos, Virginia

Campazzo, Eduardo

Universidad Nacional de La Rioja

CONTEXTO

Este trabajo es resultado de la elaboración de Tesis de Doctorado dentro del programa “Ingeniería de Software Basado en componentes reutilizables con aplicación a interfaz hombre computador” de la Universidad de Vigo España, desarrollada en el ámbito del Laboratorio de Investigación de Software de la Fac. Reg. Cba. de la UTN Argentina

RESUMEN

Este trabajo de investigación está basado en el tema de tesis de doctorado de Calixto Maldonado, dirigida por el Dr. Manuel Pérez Cota de la Universidad de Vigo, que tiene como uno de sus objetivos el estudiar y desarrollar pautas, estrategias y materiales que permitan aprovechar las ventajas del uso de Mundos Virtuales (de aquí en adelante MV) sobre todo aquellos mundos virtuales masivos de juegos de rol o en su sigla en Inglés MMORPG (por Massive Multiuser Online Role Play Games), para impartir educación mejorando la aceptación del mensaje por venir de un canal novedoso, atractivo y familiar para la nueva generación de educandos.

Palabras clave:

MMORPG, Mundos virtuales, aprendizaje, videojuegos

1. INTRODUCCION

Es una percepción habitual, al observar a alumnos en el colegio medio y de prestar atención al discurso de profesionales de educación, en el que los docentes afirman “A los alumnos no les interesa nada”. Es una de las explicaciones de los docentes sobre el fracaso de un alto porcentaje de sus alumnos. Esta falta de comunicación es uno de los factores de conflicto en la educación media y hace pensar en la necesidad de restablecer el canal que permita mejorar el aprendizaje al estudiante, al encontrar un canal común de comunicación.

Para este estudio se han tomado las ideas expuestas por Marc Prensky [Pre01] que promueve el uso de los juegos como estrategia de apoyo para transmitir a éstos nuevos estudiantes, a los que llama “nativos digitales”. Estos están habituados a usar herramientas tecnológicas para jugar y entretenerse en largas horas de ocio. También consideraremos posturas de expertos en educación como Edgar Morin [Mor99], Jean Piaget [Pia72], Ernesto Sabato [Sab96].

Para lograr el objetivo de obtener una mayor atención y, por ende, mayor recordación, de algunos contenidos formales de la educación, que deben ser aprendidos para cumplir los objetivos primordiales,

se propone usar una estrategia integral que tenga la tecnología de MV como soporte de los métodos pedagógicos estudiados durante la elaboración de la Tesis, como aprendizaje Colaborativo, Constructivismo y Aprendizaje Vicario. Se ha realizado un relevamiento de Second Life y Hipihi, mundos virtuales en funcionamiento en la actualidad y se propone utilizar la tecnología de Project Darkstar de Sun como base tecnológica de un futuro aplicativo específico para educación dada su característica de ser Abierto, no propietario como las versiones comerciales nombradas.

Edgar Morin [Mor99], afirmaba en su trabajo de 1999 “Los siete saberes necesarios a la educación del futuro” que *“En esta evolución hacia los cambios fundamentales de nuestros estilos de vida y nuestro comportamiento, la educación, en su sentido mas amplio, juega un papel preponderante. La educación es uno de los instrumentos más poderosos para realizar el cambio. Uno de los desafíos más difíciles será el de modificar nuestro pensamiento de manera que enfrente la complejidad creciente, la rapidez de los cambios y lo imprevisible que caracterizan nuestro mundo”*.

Jean Piaget [Pia72] en “A donde va la educación” afirmaba que ésta *“...debe orientarse hacia una reducción general de las barreras o hacia la apertura de múltiples puertas laterales que permitan a los alumnos (universitarios y secundarios) el libre paso de una sección a otra con la posibilidad de elegir múltiples combinaciones”*, lo que viene a confirmar que se debe trabajar en múltiples dimensiones y que estas estén conectadas. La brecha de comunicación está dada porque las generaciones crecen en compartimientos estancos. En esta misma obra, Piaget, enuncia que uno de los problemas *“... es la preparación de los maestros, y que esta es una cuestión previa a cualquier reforma pedagógica, ya que sin esta preparación los mejores programas y atractivas teorías no será posible llevar a cabo los cambios necesarios”*

Este trabajo explora herramientas pedagógicas y que los profesores aprendan como usarlas en un entorno de MV.

Ernesto Sabato [Sab96] afirmaba que *“El libro es una magnífica ayuda, cuando no se convierte en un estorbo. Si Galileo se hubiese limitado a repetir los textos aristotélicos (como uno de esos muchachos que ciertos profesores consideran “buenos alumnos”), no habría averiguado que el maestro se equivocaba sobre la caída de los cuerpos. Esto que yo digo para los libros también vale para el maestro, que es bueno cuando no es un obstáculo, lo que parece un broma, pero es una de las calamidades mas frecuentes. En el sentido etimológico, educar significa desarrollar, llevar*

hacia fuera lo que aún está en germen, realizar lo que sólo existe en potencia. Esta labor de partero del maestro muy pocas veces se lleva a cabo, y talvez es el centro de todos los males de cualquier sistema educativo. No sé que profesores tenía Galileo en el momento que se le ocurrió subir a la torre para tirar abajo dos piedras y a la vez, la teoría de Aristóteles; si eran malos, se habrán irritado por aquel crimen; si eran maestros de verdad, se habrán alegrado de aquella sagrada rebelión.” Uniendo aquella necesidad de Morin y Piaget con la caracterización de los maestros de Sabato, podemos inferir que el cambio es necesario y nos debemos preparar para no aislarnos de los alumnos y no obstaculizarlos con nuestras limitaciones.

Marc Prensky describe en [Pre01] otro motivo para este trabajo, que es *“... el niño o joven nacido desde los años 89 y 90 a esta parte. Ellos tienen el privilegio de ser la primera generación digital, y han disfrutado del aprendizaje de misiones arriesgadas y exigente en sus habilidades de motricidad fina, para conseguir algún trofeo que les brinda reconocimiento entre sus pares en los videos juegos. Los juguetes influyen en la infancia y está es una etapa fundamental en el aprendizaje y éste influye en sus estructuras mentales de la persona en su desarrollo. Este desarrollo no se inicia con la escolaridad inicial a los 6 años, puede iniciarse antes existen numerosos intentos de la industria del entretenimiento para llegar a los niños a través de música, Bach, Mozart, Beethoven para niños, juegos en consolas como la Wii de Nintendo y PlayStation de Sony, para lograr atrapar su atención y desarrollo de alguna habilidad, o iniciar la fase de ir a la cama inducidos por elementos visuales y auditivos”*.

En el artículo ya referenciado de Prensky [Pre01] afirma que *“Nuestros estudiantes han cambiado radicalmente, el estudiantes actual no es la persona para la que nuestro sistema educacional fue diseñado para enseñarle”*. Siguiendo con sus fundamentos, *“el estudiante de hoy representa la primera generación en crecer con esta nueva tecnología. Ellos han pasado su vida rodeados y usando computadoras, videojuegos, reproductores digitales de música, cámaras de video, celulares y todos los otros juguetes y herramientas de la era digital. Un graduado promedio de la educación media, ha pasado menos de 5 mil horas de su vida leyendo, y mas de 10 mil horas jugando video juegos, sin mencionar las 20.000 horas mirando TV. Los juegos de computadoras, correo electrónico, Internet, mensajes instantáneos y celulares son parte de su vida”*. Afirma también en su artículo que el problema planteado por esta nueva situación es que, *“... en general nuestros los instructores, son Inmigrantes digitales que hablan un lenguaje*

desactualizado, de la era pre-digital que están luchando para enseñar a una población que habla un lenguaje completamente nuevo”

Según lo anterior los profesores y padres son los Inmigrantes digitales, y es posible relacionarlo con nuestra historia reciente de inmigración, los mayores nacimos en Europa, el ‘viejo mundo’, el mundo pre-digital, y en un caso en el extremo, no hablamos una palabra del idioma, y en el otro, mas cercano, no se puede disimular que el lenguaje digital no es nuestro idioma madre y lo hablamos con alguna dificultad. Somos por lo tanto inmigrantes digitales en un mundo digital, imperio de la Informática.

Sigue Prensky marcando algunos problemas adicionales. *Los inmigrantes digitales no creen que sus estudiantes puedan aprender exitosamente mientras miran TV o escuchan música, debido a que ellos mismos como inmigrantes, no pueden hacerlo. También, muchos de ellos piensan También, muchos de ellos piensan que el aprendizaje no es divertido, o no debiera serlo. Los nativos digitales tienen poca paciencia para las lecturas y la lógica paso a paso. Los inmigrantes suponen que los aprendices son los mismos de siempre y que los mismos métodos que funcionaron para sus maestros cuando eran estudiantes, van a funcionar para sus estudiantes. Hay profesores que piensan que no merece la pena prestarle atención la educación, comparado con todo lo que experimentan a diario con la tecnología. Por todo esto, a menos que queramos olvidar de educar a los nativos digitales hasta que ellos crezcan y lo hagan por sí solos, tenemos que enfrentar este problema y reconsiderar nuestra metodología y nuestros contenidos.”*

Este cuadro de situación aporta algunos de los fundamentos que ayudaran a instrumentar las pautas metodológicas que aporten contenido formal al uso de mundos virtuales con el fin de mejorar los resultados en la educación, aportando un canal de comunicación conocido y atractivo para esta generación de estudiantes digitales.

Los avances en la tecnología de los Mundos Virtuales

En el estudio de las herramientas que la tecnología ofrece para mejorar la experiencia de aprendizaje, resalta la posibilidad que tiene la realidad virtual, y mas precisamente los mundos artificiales producidos en un computador y al que los usuarios acceden con una sensación de inmersión parcial, ya que no se utilizan cascos y trajes que harían mas completa esta sensación [Cam07]. En los MV en los que trabajaremos no usaremos ningún artefacto como casco y traje, será una inmersión parcial, pero que permitirá a la persona tener un estímulo suficiente como para poder situarse y tomar una clase grupal, con compañeros a los que escuchará, a través de

auriculares comunes que permitirá ubicar desde que punto viene el sonido, si de adelante, atrás, izquierda, derecha, arriba y abajo. Además de estas posibilidades como la de tomar clase en un aula, desde un computador hogareño, como si estuviera dentro del aula real, simultáneamente con instructor y demás participantes. Esto sería un inicio porque se pueden construir representaciones de lugares famosos, pudiendo realizar una visita virtual a escenarios que pudiendo constituir una experiencia vívida de aprendizaje.

Definamos a los MV con la siguiente analogía [Bar04] lo real es lo que ES, lo imaginario es lo que NO ES, y lo virtual NO ES, pero tiene la forma o efecto de lo que ES. Donde también los MV son lugares donde lo imaginario puede transformarse en algo similar a lo real.

Los MV tienen aplicaciones más allá del mero entretenimiento. Ellos empezaron como juegos de computadora pues la industria relacionada a los juegos es el motor que, hasta el momento, impulsa el desarrollo de los MV. Actualmente, por el uso corporativo dado para realizar sus reuniones de empleados de todas sus áreas distribuidas en todo el mundo la empresa Sun ha desarrollado su plataforma llamada Proyecto Darkstar y propone distintas aplicaciones como Wonderland que es un ejemplo de desarrollo de un MV para interactuar con otros usuarios, en un lugar para reuniones, exposición de posters o diapositivas, con sonido inmersivo que configura lo necesario para tener una primera aproximación a los MV [Sun08].

Una sala de Chat no es un MV por que no tiene física, un juego de guerra no define un carácter a través del cual el jugador actúe. Un juego vía e-mail, no es en tiempo real y un juego con un solo jugador no es compartido y un tirador en primera persona no es persistente [Bar04].

El uso de la informática en la educación está aún en crecimiento, Pérez Cota et Al. [Pér01] profundizaba en las aplicaciones de la informática nombrando “La posibilidad de realizar simulaciones, permitiendo mostrar al alumnado como funcionan las cosas en la realidad” y este artículo retoma la propuesta de este trabajo para agregarle la posibilidad de mostrarlo en un MV estas simulaciones para lograr mayor participación en este proceso.

Los mundos virtuales de la actualidad

Hoy en día, tenemos cerca de una docena de Mundos Virtuales, consolidados y en aumento constante, pero por cuestiones de madurez y perspectivas se han elegido dos, de los EE. UU. Second Life (www.secondlife.com), consolidado y en madurez plena y un mundo virtual de China, Hipihi (www.hipihi.com), por el potencial de

visitantes que podría reunir dado que tiene especial dedicación al público asiático. Se decidió estudiar estos dos por que Second Life ha alcanzado una mayor madurez por ser el primero en llegar a la cantidad de usuario que demuestra, y Hipihi por la potencial cantidad de habitantes que podrían acceder a este MV. En ellos se desarrollarían instalaciones, artefactos y cursos para hacer las pruebas de campo requeridas para respaldar la tesis.

Second Life

Investigadores de la Universidad de Barcelona [Gra07], lo definen con esta descripción “Second Life (en adelante SL) es un mundo virtual en 3D creado por la empresa Linden Lab en 2002 y permite interacciones entre usuarios que simulan comportamientos humanos: hablar, bailar, cambiarse de ropa o incluso intimar con otros usuarios, ver documentos y compartirlos, ver vídeos, escuchar audio, visitar lugares imaginarios o recreaciones de localizaciones reales, coger y manipular objetos, crear nuevos objetos, comprar y vender servicios, crear una casa y decorarla, recibir o impartir formación, asistir a obras de teatro, manifestarse, visitar museos. La idea de Linden Lab es demostrar la viabilidad de un modelo de economía virtual o sociedad virtual, algo que Philip Rosendale [Ter04] el creador de la empresa, definió: *«no estoy construyendo un juego, estoy construyendo un nuevo continente»*. Hoy en día, el continente virtual generado por la compañía posee cerca de diez millones de habitantes censados, los cuales, como mínimo, se han descargado el programa cliente a sus ordenadores y han llegado a entrar en el juego alguna vez. Este mundo mide unos 500 millones de kilómetros cuadrados y genera lo que podríamos llamar un «producto interior bruto» (PIB) de unos 4.500 millones de unidades económicas denominadas Linden Dollars (L\$), que cotizan a 266 L\$/US\$. Unos 17 millones de dólares” [Gra07] Con la ayuda del autor catalán, agregamos más detalles sobre Second Life [Gra07] “SL permite a los usuarios del entorno utilizarlo de manera diferenciada, individual y acorde a sus expectativas. Si analizamos las acciones que son posibles en SL veremos que no se puede considerar un juego sino una realidad paralela:

- Crear un personaje y diseñar su apariencia (incluyendo rasgos físicos y ropa)
- Conversar con otras personas (utilizando el teclado o la voz), establecer redes sociales y grupos.
- Interactuar con otras personas (a través de gestos, de movimientos corporales, intercambiando objetos...)
- Visitar lugares o recreaciones virtuales de ciudades existentes, entrar en Edificios, crear un lugar propio.

- Interactuar con los objetos que ya existen (sentarse en un sillón, mover una silla, encender la radio, lanzar un objeto, mover una pieza de ajedrez...)
- Crear nuevos objetos (desde una esfera a un edificio).
- Ejecutar vídeos, audio o archivos de presentaciones llamados Machinimia.
- Vender o comprar objetos y obtener un rendimiento económico real.
- Vender o comprar servicios.
- Asistir a conciertos, eventos de presentación de películas, sesiones de moda, obras de teatro.
- Impartir o recibir cursos de formación, asesoramiento
- Realizar simulaciones para aprendizaje (una simulación de accidentes, una Simulación de una oficina bancaria y así)
- Organizar manifestaciones, montar partidos políticos, convocar huelgas.”

Conclusión

Es muy estimulante la posibilidad de estudiar la aplicación de este tipo de tecnologías en el proceso educativo como una contribución para revertir la crisis social en la que estamos inmersos. En este artículo se nombraron antecedentes, herramientas y se enunciaron algunos principios en los que basarse para realizar las propuestas aplicables con las nuevas tecnologías informáticas, en especial la de los mundos virtuales. El objetivo del trabajo es mejorar el aprendizaje con el uso de esta tecnología, concientes que la crisis educativa actual lo hace urgente porque los nativos digitales no se sienten integrados al sistema educativo, o los estamos empujando a un fracaso con reglas impuestas hace mucho tiempo. Otro tema que hemos visto es la brecha de comunicación junto con la brecha de conocimiento tecnológico de los “inmigrantes digitales” y ha sido considerado como un problema a resolver. Se esta estudiando entre otras formas de estrategias el construir instalaciones, por ejemplo, museos para que sirvan como visitas al pasado y hechos históricos, recorridos que permitan revivir situaciones pasadas para lograr una mejor experiencia educativa a los nativos digitales reacios a leer libros .

2. LINEAS DE INVESTIGACION y DESARROLLO

Educación en entornos virtuales, estudio de teorías epistemológicas y pedagógicas para su aplicación en MV. La propuesta es desarrollar contenido educativo que pueda aplicarse dentro de Second Life y desarrollar con la Tecnología del Proyecto Darkstar un entorno virtual completo que permita desarrollar las actividades planteadas también para Second Life.

3. RESULTADOS OBTENIDOS/ESPERADOS

A lo largo del período de trabajo para la tesis doctoral se realizaron varias acciones, estudio de antecedentes, asistencia a cursos y congresos de temáticas relacionadas, entrevistas con directivos de escuelas medias y universidades, encuestas a alumnos y trabajos de práctica en el uso de Second Life. Falta aún la sistematización de los conocimientos adquiridos, respuestas encontradas, respuestas aún no dadas y la enunciación de técnicas, procesos y modalidades a aplicar la tecnología de un mundo virtual

En la actualidad se ha elaborado un marco teórico sobre educación, buscando la definición de la estrategia educativa que se adaptaría mejor a los objetivos y a la herramienta. Al respecto nos basaremos en el aprendizaje Colaborativo, Constructivismo y se esta considerando también otros tipos como el Vicario para poder construir los elementos a utilizar. Se estan construyendo las instalaciones virtuales y configurando el uso de un sistema de administración de aprendizaje en la plataforma Moodle, ver el sitio desarrollado www.temuestrocomo.com.ar (dentro de él seguir los links “cursos” y “aula” para llegar a la portada de Moodle de este sitio).

Está previsto utilizar lo desarrollado en la escuela de verano en una universidad de Finlandia, a cargo del profesor Dr. Manuel Pérez Cota. Se espera realizar algunas pruebas en un escenario preparado para una actividad práctica tal como construir el mapa conceptual en tres dimensiones con bolas y palos, al estilo de los modelos tridimensionales de las moléculas en química.

4. FORMACION DE RECURSOS HUMANOS

Este trabajo esta en la mitad de su desarrollo y durante este período se ha trabajado realizando encuestas entre alumnos de la escuela media de la ciudad de Córdoba, junto con la revisión bibliográfica y el desarrollo un sitio web que permita la interacción de educadores y educandos en practicas de uso de MV. También se ha trabajado para mejorar la comunicación a través de MV entre los integrantes del curso de doctorado y los tutores de tesis. Los avances en cuanto a la relación con otros grupos, están dados porque los integrantes del programa de Doctorado localizados en Argentina realizamos reuniones y lectura crítica de los distintos trabajos de publicación, cargamos la bibliografía leída, con opiniones sobre el contenido, posibles aplicaciones, análisis de contradicciones entre autores que es mantenido en un repositorio común del grupo y se esta trabajando en la unificación para su publicación en un solo sitio web que permita la discusión y difusión de los avances entre todos los integrantes del programa, radicados en La Rioja, Catamarca, Córdoba, Tierra del Fuego, Santa Cruz y

Santa Fe en Argentina, en Santa Cruz de la Sierra Bolivia, Castelo Branco y Porto en Portugal y Vigo, Pontevedra y Ourense en Vigo España. La dirección del sitio es:

<http://www.frrea.utn.edu.ar/vigo/inicio.htm>

5. BIBLIOGRAFIA

[Bar04] Bartle, Richard A. Designing Virtual Worlds. New Riders-California-Pearson Education ISBN 0-1310-1816-7

[Cam07] Técnicas de Interacción y Dispositivos Utilizados en Mundos Virtuales. TÉCNICAS DE ESPECIFICACIÓN Y MODELOS DE SISTEMAS INTERACTIVOS UNIVERSIDAD DE VIGO 08T151A031 – Prof. Javier Rodeiro Iglesias.

[Gra07] Mariona Grané, Joan Frigola, Miguel Angel Muras Second Life: Avatares para aprender - Universitat de Barcelona -2007

[Lop04] «Por una Educación Literaria a partir de la LIJ», LÓPEZ MOLINA, JUAN en CLIJ, nº 167. Enero.

[Mal04] Horacio Maldonado “Escritos sobre psicología y educación” Ed. Espartaco Córdoba 2004

[Mor99] Edgar Morin, “Los siete saberes necesario a la educación del futuro” UNESCO. París, 1999

[Pér01] Manuel Pérez Cota; Ana Diez Sánchez; Luis Vilán Crespo ;José Paulo Machado Da Costa LA APLICACIÓN DE LAS NUEVAS TECNOLOGÍAS EN LA ACTIVIDAD DOCENTE: La creación de una “nueva” teoría educativa. Vigo 2001

[Pia72] Piaget, Jean “A donde va la educación” Ed. Teide Barcelona 1978

[Pre01] Marc Prensky Digital Natives, Digital Immigrants - On the Horizon (NCB University Press, Vol. 9 No. 5, October 2001)

[Sab96] Sabato, Ernesto, “Sobre algunos males de la educación” Apologías y Rechazos. Ed. Seix Barral, Buenos Aires 1996.

[Ter04] Terdiman, Daniel (May 8, 2004). "[Fun in Following the Money](http://www.wired.com/news/games/0,2101,63363,00.html)". [Wired Magazine](http://www.wired.com/news/games/0,2101,63363,00.html). Retrieved on 2006-12-05.

Calidad en el Diseño y Producción de Materiales de Educación a Distancia

Mg. De luca, Adriana J. Facultad de Ciencias Exactas Universidad Nacional de San Juan

Mg. Llarena, Myriam. Facultad de Ciencias Exactas Universidad Nacional de San Juan

CONTEXTO

El Proyecto “Calidad en el Diseño y Producción de materiales para Educación a Distancia” se desarrolla en el marco del Programa Permanente de Educación a Distancia que fue creado por Resolución 08/05 de la Facultad de Ciencias Exactas, Físicas y Naturales. En este espacio académico y de investigación se conforman equipos de trabajo interdisciplinarios abocados a la tarea de generar experiencia, conocimientos, soporte tecnológico-comunicacional y organizativo a las propuestas académicas bajo la modalidad a distancia en nuestra Facultad.

Su propósito es:

- Construir un marco de referencia que permita sistematizar los conocimientos multidisciplinares que involucra la modalidad Educación a Distancia
- Potenciar la capacidad tecnológica de la FCEFYN manifiesta en recursos humanos y equipamiento orientados a la Educación a Distancia.

En este contexto, esta línea de investigación encuentra canales de interacción con otras líneas tales como *Intervención Didáctica para Educación a Distancia* y *Estrategias Gestión de Calidad de Cátedras*.

Interacción que se manifiesta en recursos compartidos en la ejecución de proyectos conjuntos tales como: *Proyecto de Implementación de Oferta Académica de Grado en la Modalidad a Distancia* (Disposición 7/08) y en los Proyectos de Implementación de *Cursos de Posgrado de Gravimetría* realizados en forma conjunta entre investigadores de ésta y de otras instituciones universitarias.

RESUMEN

El Programa Permanente de Educación a Distancia ha definido variadas líneas de

investigación que no se consideran exhaustivas para el abordaje del objeto de estudio : Educación a Distancia.

Este proyecto se encuadra en una de ellas: *Calidad en la oferta Académica en modalidad a distancia*.

La problemática de la determinación de criterios y estándares para evaluar la calidad se analiza desde diferentes enfoques. En la línea de investigación que desarrollamos se prioriza la necesidad de generar sistemas de evaluación contextualizados, pero que a la vez atiendan los estándares exigidos a nivel internacional y “las buenas prácticas” de las instituciones relevantes en la modalidad de educación no presencial.

Palabras clave:

Calidad- Educación a Distancia- Materiales didácticos

1. INTRODUCCION

El diseño y producción de materiales para la modalidad a distancia es considerado un factor crítico por su incidencia en la calidad de las propuestas bajo esta modalidad. Supone el trabajo mancomunado de docentes, diseñadores, programadores, y tecnólogos. Esta investigación toma como problemáticas: ¿qué metodología de seguimiento y evaluación de calidad será adecuada para las etapas de diseño y producción de materiales adecuados a la modalidad no presencial? ¿qué instrumentos se podrán utilizar para asegurar la calidad de los materiales?

“...la nueva Educación Superior deberá poder ofrecer una formación básica de calidad que favorezca la adquisición de competencias para seguir aprendiendo toda la vida y para operar en la Sociedad, al mismo tiempo que promueva continuas adaptaciones a nuevos contextos con el plus de la adquisición de las Nuevas

Tecnologías...Lo que estamos señalando, es que los nuevos escenarios demandan nuevas configuraciones en las instituciones de nivel superior. Ya no resulta posible, por lo tanto pensar solamente en una institución estática y limitada a un edificio. Las nuevas configuraciones que nacen a la luz de las NTICs, han roto la idea de "agrupamientos académicos fijos" y la han reconfigurado hacia la creación de "agrupamientos académicos diversificados", cuya característica fundamental reside en su alto poder expansivo, que como en el caso de la virtualidad, permiten la conformación de espacios de trabajo académicos capaces de extenderse a través de miles y miles de kilómetros, desarrollarse en variables y múltiples contextos y viabilizarse a partir de diferentes recursos materiales y pedagógicos. La cultura institucional, en consecuencia, requiere una profunda reformulación que le permita responder adecuadamente a nuevos tipos de demandas, ofreciendo propuestas de formación que respondan eficaz y eficientemente a diferentes necesidades sociales y educativas, generando a la vez una apertura institucional hacia nuevos espacios formativos abstractos, y, por sobre todo configurando nuevos "tipos" institucionales capaces de hacer frente a estas exigencias y características"¹

Las instituciones universitarias argentinas que tradicionalmente han ofrecido su oferta académica bajo la modalidad presencial, están ampliando sus propuestas de estudios en la modalidad a distancia. La Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan acorde con estos procesos de cambio, ha iniciado acciones tendientes a desarrollar actividades bajo la modalidad no presencial, proponiéndose como desafío continuar realizando propuestas de calidad a la comunidad.

A pesar del incremento de las propuestas de

¹ Lugo, Vera y Rossi. Situación presente y perspectivas de desarrollo de Programas de Educación Superior Virtual en Argentina.

educación a distancia en América Latina y a nivel internacional, es muy poco lo avanzado respecto de la evaluación de su calidad."En la mayoría de los países se señala una significativa carencia de profesionales formados para tareas de evaluación y acreditación... Hay, además, un importante vacío de criterios y metodologías para la evaluación y acreditación de la educación a distancia virtual...hay un limitado desarrollo conceptual y metodológico para la evaluación y acreditación de los aspectos pedagógicos"²

En este contexto, la conveniencia de medir y evaluar la calidad de sus propuestas educativas, ya no es tema de debate, está instalado en el ámbito de las universidades. El motivo de análisis está centrado ahora en determinar aspectos, características e instrumentos de evaluación.

La modalidad no presencial requiere proponer prácticas pedagógicas que potencien la interactividad alumno/material e interacción entre distintos actores del proceso educativo. Las universidades argentinas, en su mayoría, no disponen de equipos interdisciplinarios para la generación de materiales adecuados a esta modalidad. Los docentes se transforman en planificadores, expertos en contenidos, tutores, generalmente interactúan con pedagogos y tecnólogos, roles que necesariamente deben ser desempeñados por expertos en dichas temáticas.

Desde nuestra experiencia como responsables de cursos, dictados en el marco del Programa Permanente de Educación a Distancia(EaD) de la FCEfyN, advertimos que el docente manifiesta dificultades a la hora de diseñar el material mediado para esta modalidad.

Por ello, el Proyecto *Calidad en el Diseño y Producción de Materiales de Educación a Distancia*, como parte del mencionado

² Fernández Lamarra N. La Evaluación de la calidad de la Educación Superior en América latina y El caribe. Situación, Tendencias y Perspectivas.

Programa plantea los siguientes objetivos específicos:

1. Definir estrategias de seguimiento y evaluación que aseguren la calidad de los procesos de diseño y producción de materiales mediados para E a D.

2. Generar instrumentos de seguimiento y evaluación de la calidad de los procesos de diseño y producción de materiales mediados para Ea D

2. LINEAS DE INVESTIGACION y DESARROLLO.

La línea de investigación se centra en torno a la *Calidad en la Educación a Distancia*.

El estudio de la calidad en esta modalidad puede ser abordado en todas las dimensiones que implica llevar adelante cursos de Educación a Distancia, administrativa, pedagógica, entre otras. Esta línea de investigación se centra en los procesos de Enseñanza/Aprendizaje y dentro de ellos, se acota al estudio de la calidad del material. Específicamente se centra en la evaluación de la interactividad alumno /material e interacción entre los distintos actores del proceso educativo. Las fases que se abordan son las de Diseño y Uso del material.

En la determinación del modelo de evaluación se prioriza la necesidad de generar sistemas de evaluación contextualizados, pero que a la vez atiendan los estándares exigidos a nivel internacional (Modelo de Calidad Total) y "las buenas prácticas" de las instituciones relevantes en la modalidad de educación no presencial.

Desde el enfoque de la Calidad Total se requiere no solo atender a la calidad del servicio, sino también la satisfacción del usuario. Por ello en el marco del prototipo de metodología desarrollada MECMAD (Metodología para Evaluación de la Calidad de Materiales a Distancia), se han generado instrumentos para evaluar la calidad de los procesos mencionados, desde el punto de vista de los alumnos (usuarios) y de los docentes (diseñadores).

3. RESULTADOS OBTENIDOS

Referentes en la modalidad no presencial indican: "se ignora en buena medida, al menos en la práctica, la diferencia entre diseño y uso, entre lo que el diseño tecnopedagógico del proceso de enseñanza y aprendizaje virtual plantea, y lo que los participantes realmente acaban haciendo en el proceso de desarrollo de ese diseño y en el uso concreto de las herramientas, materiales y recursos que el diseño provee"³ Por ello, en el marco del proyecto, se generó la metodología **MECMAD** para guiar al docente en la etapa de diseño de materiales y en la evaluación de la calidad del uso de la propuesta generada.

Conforman esta metodología las fases de *Formación del docente en el diseño de materiales*, *Revisión y Puesta a Punto del Diseño* y *Evaluación de la Calidad en el uso del material* propuesto en relación a la **interactividad alumno /material e interacción entre los distintos actores del proceso educativo**

Primera fase: Formación del Docente

Fase de concientización: Dada la complejidad del proceso de mediación y teniendo en cuenta que no todos los profesores tienen formación docente, la metodología propone como primera fase la Formación docente en aspectos didáctico / pedagógicos relacionados con la modalidad no presencial. Esta fase concluye con el diseño del material pertinente para desarrollar una unidad didáctica

Segunda Fase: Revisión y Puesta a Punto del Diseño realizado

Para esta fase se utiliza como instrumento, un documento que permite al docente analizar y controlar si el diseño realizado cumple con requerimientos mínimos de calidad en los procesos de interactividad e interacción.

³ Onrubia, J. (2005). Aprender y Enseñar en Entornos Virtuales

Tercera Fase: Evaluación de la Calidad del uso de la propuesta

El diseño tecno-pedagógico que establece actividades a ser desarrolladas para que el alumno pueda acceder al conocimiento, es el mismo para todos los alumnos. Indica Onrubia (2005): “La estructura lógica de un contenido remite a la organización interna del material de aprendizaje en sí mismo, y puede **considerarse estable entre contextos, situaciones y aprendices**. La estructura psicológica del contenido, en cambio, remite a la organización de ese material para un alumno concreto, y depende de lo que, en cada momento, el alumno aporta al proceso de aprendizaje”. Para que el alumno atribuya significado al contenido no sólo debe darse la significatividad lógica relacionada con la estructura y organización lógica de los contenidos en el entorno, sino también la significatividad psicológica. Mientras la primera se puede garantizar con un buen diseño de los materiales, la psicológica sólo puede asegurarse si el proceso de aprendizaje de ese material se adapta a los alumnos que deben aprenderlos. El alumno puede realizar prácticas no previstas ni promovidas o no realizarlas tal como se pretendían en las planificadas en el diseño. Por ello, tal como expresa Onrubia(2005), es necesario reconocer “ la diferencia entre lo que el diseño tecno-pedagógico del proceso de E-A virtual plantea, y lo que los participantes realmente acaban haciendo en el proceso de desarrollo de ese diseño”.

El modelo adoptado para evaluar la calidad del uso de la propuesta está sustentado en dos enfoques que se utilizan para determinar criterios e indicadores de calidad: los Sistemas Benchmarking y los Sistemas de evaluación de la calidad centrados en Modelos de Calidad Estándar. El primer enfoque consiste en atender las buenas prácticas realizadas por instituciones líderes en la temática. En este sentido se consideró la experiencia y trayectoria de instituciones como la ANECA (Agencia Nacional de Evaluación de la Calidad y Acreditación de España) y

la Universidad de La Loja, unidad ejecutora del proyecto “Centro Virtual para el Desarrollo de Estándares de Calidad para la Educación Superior a Distancia en América Latina y el Caribe”, cuyas propuestas han sido el resultado de la comparación de varios modelos vigentes en universidades de punta en la modalidad a distancia internacionales.

Una metodología de evaluación debe estar sustentada en una concepción de calidad. La metodología generada se sustenta en la Calidad Total, que requiere considerar no sólo la calidad del proceso sino también la satisfacción del usuario para lo cual propone *Gestionar y Garantizar la Calidad*. Basados en los principios del Modelo de Calidad Total, la metodología propone dividir la fase de evaluación en sub fases:

- 1- Planificación de la evaluación** , en la que se gestiona la calidad
- 2- Ejecución**, en la que se comprueba la calidad
- 3- Elaboración de informe con propuestas de mejoras**, destinado a mejorar la calidad

Para *evaluar la calidad de la interacción e interactividad en la etapa de desarrollo (uso)* de la propuesta formativa se han construido instrumentos para recabar información desde el punto de vista de los alumnos (usuarios) y del profesor (diseñador).

La información es tratada en forma cuantitativa y cualitativa atendiendo a la característica de las variables consideradas. En ambos casos los resultados obtenidos permiten reorientar los aspectos necesarios y potenciar otros.

El informe que recibe el docente contiene las fortalezas y debilidades de los criterios analizados, como así también propuestas de mejora.

La metodología ha sido aplicada para evaluar la calidad de los procesos de Enseñanza/Aprendizaje en Cursos de Posgrado y de Grado.

4. FORMACION DE RECURSOS HUMANOS

Se ha participado en el Diseño, Producción y Evaluación de Calidad de los siguientes cursos de posgrado:

- Diseño de Prácticas Pedagógicas con Tecnologías web 2.0 para EAD (Primera Edición)
- Diseño de Prácticas Pedagógicas con Tecnologías web 2.0 para EAD (Segunda edición)
- Tecnologías Web y Educación: Formación de Competencias para la Gestión de Información.
- Proyecto Implementación oferta académica de grado en modalidad semi-presencial.

Se han dirigido 5 Tesis de Posgrado (dos Títulos obtenidos) , 2 Tesis de Grado y 1 Becario.

Los cursos mencionados fueron el espacio de testeo del prototipo de la Metodología MECMAD-.

Se espera continuar la validación de este prototipo, para realizar ajustes que permitan mejorarlo y definir mecanismos que permitan agilizar el procesamiento de la información y estrategias para mejorar los procesos de retroalimentación.

5. BIBLIOGRAFIA

- Lugo, Vera y Rossi(2003). Situación presente y perspectivas de desarrollo de Programas de Educación Superior Virtual en Argentina.
http://www.iesalc.unesco.org.ve/programas/internac/univ_virtuales/
- Fernández Lamarra N (2004). "La Evaluación de la calidad de la Educación Superior en América latina y El caribe. Situación, Tendencias y Perspectivas".
- Onrubia, J. (2005). Aprender y Enseñar en Entornos Virtuales: actividad conjunta, ayuda pedagógica y construcción del conocimiento. RED. Revista de Educación

a Distancia, número monográfico II. Consultado (febrero2009) en

<http://www.um.es/ead/red/M2/>.

- Barberá, E et al. (2004). *Pautas para el análisis de la intervención en entornos de aprendizaje virtual: dimensiones relevantes e instrumentos de evaluación* [documento de proyecto en línea]. IN3:UOC. [consultadoNoviembre2008].
<<http://www.uoc.edu/in3/dt/esp/barbera0704.html>>
- Garcia Aretio (2002) *La Educación a Distancia, de la teoría a la Práctica*. Segunda Edición Barcelona. Ed. Ariel S.A Cap.7: Comunicación a través de medios.
- Lima Montenegro C. (2007) *Educación a Distancia y Preparación de Materiales Educativos con uso de La Hipermedia*. Instituto Pedagógico Latinoamericano y Caribeño. <http://revista.iplac.rimed.cu/>
- Litwin E(2000), *La educación a distancia Temas para el debate en una nueva agenda educativa* Amorrortu Editores S.A. Buenos Aires.
- Mauri T, Onrubia J, Coll C y Colomina R.(2005) *La Calidad de los Contenidos Educativos reutilizables: diseño, usabilidad y prácticas de uso* Universidad de Barcelona. Departamento de Psicología Evolutiva y de la Educación. Passeig de la Vall d'Hebron, 171. 08035 Barcelona <http://www.um.es/ead/red/M2/>.
- Prieto Castillo(2000), *Tratamiento del Contenido, La Enseñanza en la Universidad*, Mendoza EDIUNC

ESTRATEGIA PARA UNA GESTION DE CALIDAD DE CATEDRAS UNIVERSITARIAS.

Rosa M Pósito – Alfredo Leiva
Nestor Ferrari – Flavia Millán - Sonia Pelayes
Dpto de Informática
Facultad de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de San Juan

CONTEXTO

El Proyecto *Estrategia para una gestión de calidad de cátedras universitarias* se desarrolla en el marco del Programa Permanente de Educación a Distancia creado por Resolución 08/05 de la Facultad de Ciencias Exactas, Físicas y Naturales.

El mencionado programa constituye un espacio académico y de investigación, que se sustenta en la conformación de equipos de trabajo interdisciplinarios abocados a la tarea de generar experiencia, conocimientos, soporte tecnológico-comunicacional y organizativo a las propuestas académicas bajo la modalidad a distancia en nuestra Facultad. Tiene como propósitos:

- Construir un marco de referencia que permita sistematizar los conocimientos multidisciplinares que involucra la modalidad Educación a Distancia
- Potenciar la capacidad tecnológica de la FCEfyN manifiesta en recursos humanos y equipamiento orientados a la Educación a Distancia

RESUMEN

Las instituciones educativas se encuentran inmersas en procesos de cambios, enmarcados en un conjunto de transformaciones sociales propiciadas por la innovación tecnológica. Estas transformaciones suponen cambios en los modelos educativos, en los usuarios de la formación y en los escenarios donde ocurre el aprendizaje.

Las cátedras universitarias concebidas como una unidad constituida por un equipo de profesionales que desarrollan prácticas educativas en un área temática curricular, no están al margen de esas transformaciones.

Para cumplir su cometido, las cátedras deben abandonar sus prácticas perimidas, para dar paso a una verdadera "organización

inteligente". El éxito de su gestión condiciona fuertemente la formación integral del alumno.

Los modelos de gestión educativa, la filosofía emergente de la Norma Internacional para Gestión de Calidad ISO 9000:2000 y la estructura metodológica de la Norma Internacional Global-Gap (Eurep-Gap) ofrecen un marco de referencia para diseñar una estrategia que permita la gestión de la calidad educativa de las cátedras.

La expectativa de alcanzar estándares de calidad superiores en una cátedra aumenta, en la medida que esté soportada por un "sistema de gestión de la calidad". El diseño de una herramienta tecnológica que permita gestionar dicho sistema representará un aporte importante para los responsables de cátedras.

Palabras clave: *Cátedra universitaria – Gestión - Calidad*

1. INTRODUCCION

Las **cátedras universitarias** son consideradas "estructuras primordiales", "unidades operativas básicas" (Burton Clark, 1981) que se remontan a la primigenia universidad medieval organizada como federación gremial de maestros. Etimológicamente el concepto Cátedra, deriva de latín *cátedra* (silla) y ésta del griego *Asiento, trono, silla*, "asiento"¹ como "artefacto cultural" que aparece connotado como símbolo del poder y autoridad. La asociación poder-asiento deviene en principio fuera del ámbito educativo, en el eclesial, se llama *cátedra* al sillón en que se sienta el obispo en los oficios litúrgicos.

Este concepto se ha extendido al ámbito universitario y tiene que ver con ese "asiento elevado desde donde el profesor explica", "asiento elevado desde donde el maestro da lección a sus discípulos, el que

¹ Diccionario Crítico Etimológico. J. Corominas. Madrid, Gredos, 1954

sabe explica, supuestamente al que no sabe”.² En un principio la cátedra fue un lugar de habla, de palabra. El uso de la palabra “hablada” es uno de los atributos del poder. La cátedra parece construirse a partir de una oralidad que se ejerce, o de oralidades que se ejercen, en relaciones cara a cara. Y que en su acontecer fue transformándose y complejizándose, convirtiéndose en espacio de producción y de organización simbólica para diferentes interlocutores y a través de códigos diversos. (Cañabate, Jorge Orlando, 2004).

A través del tiempo las cátedras han tenido que acompañar continuos procesos de cambios dados por los avances en su área de conocimiento específica, como en las áreas Didáctica, Gestión y Tecnología. Afrontar problemáticas como la incorporación en los planes de estudio de los últimos desarrollos teóricos, la obsolescencia del conocimiento originada por la vertiginosidad en que se producen los cambios a raíz del desarrollo científico-tecnológico, la necesidad de actualización permanente de egresados y de personas insertas en el mundo productivo, la incorporación de personas adultas a estudios superiores, entre otros. Dar respuesta a aprendizajes continuos, capacitación y educación permanente, reconversión y reorientación de la formación. (Litwin Edith, 2000).

Por lo tanto, atender a las nuevas necesidades educativas que la evolución de la sociedad y la evolución misma de las TIC generan, y la anticipación de las necesidades educativas que la evolución futura planteará, requiere una transformación de los ambientes educativos. Dentro de este contexto los docentes que integran los equipos de cátedras deben afrontar la realidad que les imponen estos nuevos ambientes educativos conformados por las condiciones en las que se vive en esa institución; los espacios físicos de la misma; las relaciones sociales presenciales que allí se desarrollan; el objeto de esas relaciones, etc., que en definitiva van configurando el clima de ese contexto que, influye de manera directa en el desarrollo y adquisición de conocimientos,

² Diccionario Ideológico de la Lengua Española. Casares. Barcelona, Gili, 1959

competencias, habilidades, valores y, en definitiva, de conductas. (García Aretio, 2007) Es la oportunidad para detenerse, reflexionar sobre la misión y visión de las cátedras universitarias y realizar la transformación necesaria en su estructura, procesos, actores para CRE- SER, Crecer + Ser, que tiene que ver con un crecimiento cualitativo, actitudinal, a nivel personal, grupal, institucional y social. (Pósito, R, Pelayes, S, 2009)

El crecimiento cualitativo implica en primer lugar un autoconocimiento, una mirada al interior de la cátedra, y preguntarse: *quiénes somos?, a qué respondemos? y cómo lo hacemos?*, “*Quiénes*” remite a la identidad, “*Qué*” al mandato social y “*Cómo*” a la forma de organizarse y a partir de esta reflexión determinar *qué cambiar?, cuánto cambiar? y cómo hacerlo?*. (Liliana Farjat, 1999).

Estos interrogantes responden a la necesidad de “*gestionar la cátedra*” a fin de transformarse en una “organización inteligente” (Liliana Farjat, 1999), concentrada en lo que sabe y debe hacer para concretar sus fines, sin malgastar energía en actividades inútiles y en permanente aprendizaje para no repetir problemas ni errores.

Gestionar la cátedra, al igual que cualquier organización/ Institución en el marco de la calidad educativa implica necesariamente trabajar sobre un **Modelos de gestión** y un **Sistema de gestión de la calidad**.

El proceso de gestión es “el proceso mediante el cual el directivo o equipo directivo determinan las acciones a seguir (planificación), según los objetivos institucionales, necesidades detectadas, cambios deseados, nuevas acciones solicitadas, implementación de cambios demandados o necesarios, y la forma como se realizarán estas acciones (estrategias, acción) y los resultados que se lograrán” (Amabilia Mayorga Moreno, sf.)

Bellenger menciona distintos modelos de gestión: el directivo, participativo y por proyecto.

En el *modelo directivo*, su fundamentación es la jerarquía, el de mayor jerarquía es el único responsable de definir las funciones, las ordenes y los objetivos, el resto de los actores

solo son ejecutores.

En el *modelo participativo*, los métodos de trabajo y las estrategias de la acción se establecen a través del consenso. Pero la política y los objetivos son responsabilidad de los que ejercen la dirección. En este modelo, los actores son considerados colaboradores.

El *modelo por proyecto* toma parte de los dos modelos anteriores. Del primero, la eficacia y del segundo la motivación.

El modelo de *Gestión por proyecto*, seleccionado como modelo de Gestión para la cátedra, es un modelo movilizador, basado en la participación y el compromiso mancomunado para la consecución de los objetivos. Según Bellenger, el trabajo realizado según este modelo plantea ocho grandes consecuencias:

1. Actualiza y fortalece las relaciones transversales.
2. Refuerza el poder de decisión de los ejecutivos, concentrándolo en aspectos importantes y bajo un mayor control de los especialistas.
3. Integra a los miembros de un equipo alrededor de objetivos propios surgidos de su misión, proporcionando sentido al trabajo y constituyéndose en factor de cohesión.
4. Hace partícipes a todos de los progresos, incentivando la optimización.
5. Moviliza a la acción como parte de los procesos.
6. Destaca y potencia las capacidades individuales, incitando a participar por afinidades.
7. Refuerza la autonomía, generando práctica en el seguimiento de los procesos y la distribución de responsabilidades.
8. Estructura a cada grupo a través de la organización de las acciones y la delegación de responsabilidades.

Esta propuesta de Modelo de Gestión intenta romper con el modelo de Gestión verticalista, eliminando rigideces y estratos. Si bien la estructura de las cátedras actualmente presenta una estructura vertical y rígida, será importante ir advirtiendo la necesidad de rever dicha estructura y por qué no pensar en un cambio.

En este modelo se privilegia la autonomía y se

libera la *sinergia* que se genera en los equipos de trabajo bien conformados. Townsend y Gerbhart, al hablar de sinergia, la definen como la *combinación de talento con espíritu de equipo*. Sinergia es esa energía imparabable que se genera en los grupos de personas altamente motivadas, productora de acciones fecundas y complementarias, donde cada integrante tiene su lugar, reconocido y acreditado por el resto de las personas con quienes trabaja. (Liliana Farjat, 1999).

Hablar de una transformación en el trabajo de los equipos de cátedra, puede generar resistencias, es por ello se ve conveniente incluir una segunda línea de acción en el Modelo de gestión para orientar y acompañar a las cátedras en su proceso de autogestión. Liliana Farjat presenta para la Gestión por proyectos, el Modelo de Task Force. Este modelo adaptado al modelo de gestión de cátedra, consiste en nombrar una persona externa como responsable de orientar y acompañar a un equipo de cátedra a alcanzar la concreción de un proyecto determinado. La incorporación es durante un tiempo perentorio, como un orientador/facilitador del proceso de gestión, trabajando mano a mano con el responsable del equipo de cátedra en las diferentes dimensiones para aprender a autogestionarse en el marco de la calidad educativa.

La calidad educativa y por tanto la estrategia para lograrla hacen énfasis en cómo se hacen las cosas, desde dónde se hacen, en qué condiciones se hacen, cuáles resultados se obtienen y cuál es el impacto de esos resultados. (Leiva, 2008)

La estrategia que se propone para gestionar la calidad, tiene en cuenta la filosofía emergente de la Norma Internacional para Gestión de Calidad ISO 9000:2000 y la estructura metodológica de la Norma Internacional Global-Gap (Eurep-Gap). La simbiosis de estos dos antecedentes, aporta una herramienta sencilla, potente, clara y adaptable, para gestionar la calidad de una cátedra de cualquier nivel.

La Norma Internacional ISO 9000:2000, contempla un modelo de un *sistema de gestión de calidad basado en procesos*. Esta referencia

a la Gestión de Procesos es una base para asegurar la calidad en dos enfoques: Centrada en el proceso de formación y centrada en los productos de la formación. (Leiva, 2008). Las ventajas del enfoque basado en procesos:

- Si es posible identificar los procesos involucrados, será posible mejorarlos.
- Si se ha trabajado para mejorarlos, se pueden construir documentos y registros de respaldo.
- Pueden capitalizarse los esfuerzos realizados transmitiendo la forma en que trabaja la organización, para generaciones venideras o simplemente “nuevos integrantes involucrados en los procesos por corto tiempo”.
- Se puede ver y reconocer qué procesos son “críticos” y “no críticos”.
- Se pueden asignar responsabilidades por procesos.
- Se reconocen las interacciones entre procesos.
- Hay una visión unificada del aporte de cada proceso a los objetivos de la organización.

La Norma Internacional Global GAP, es un sistema de gestión de la calidad que tiene el propósito de acordar los procedimientos para el desarrollo de las buenas prácticas agrícolas (Good Agricultural Practices - GAP). Este sistema nació como una iniciativa de los comerciantes minoristas y supermercados europeos en 1997. (Leiva, 2008)

La Norma ISO 9000 aporta básicamente los principios rectores de la gestión de calidad, y la Norma Internacional Global GAP, “la estrategia” de trabajo para gestionar la calidad en forma muy sencilla, organizando los procesos en las cuatro dimensiones de la gestión educativa : *Pedagógica, Administrativa, Socio-comunitario e Institucional/Organizacional.*

La estrategia propone la *Elaboración de la Lista de Verificación* que contiene:

- a) Puntos de Revisión.
- b) Criterios de Cumplimiento
- c) Pautas a tener en cuenta
- d) Declaración de prioridades
- e) Resultado de la Revisión

Contar con una herramienta tecnológica que facilite utilizar *la estrategia* de gestión de calidad, representará un aporte importante para los docentes.

2. LINEAS DE INVESTIGACION y DESARROLLO

Los ejes centrales de investigación y desarrollo del presente proyecto giran entorno a:

- La Gestión educativa particularmente tomando como unidad de organización, la Cátedra universitaria.
- Las estrategias que permitan gestionar la Calidad de los procesos de las cuatro dimensiones que involucra la gestión educativa.

Estos ejes se enmarcan en las líneas de investigación que propone el Programa de educación a Distancia: *Gestión Educativa y Calidad en la oferta Académica en modalidad a distancia.*

3. RESULTADOS OBTENIDOS/ESPERADOS

- **Resultados Obtenidos**
 - Diseño y Aplicación de entrevistas a distintas cátedras universitarias, para formar un modelo de comportamiento actual de las mismas. Entrevistas a referentes importantes del ámbito educativo: Soc. Sara Valenzuela y Dra. Liliana Martínez
 - Diseño de un Modelo de Gestión de cátedra. (En revisión por especialista en el tema)-.
 - Diseño de una Estrategia para la Gestión de la calidad de Cátedras Universitarias (En revisión por especialista en el tema)-.
 - Diseño e implementación de un espacio virtual www.gestioncatedra.blogspot.com , que registra el trabajo del equipo de investigadores del proyecto: producciones, acuerdos de reuniones presenciales, cronograma de actividades, talleres, Bibliografía, etc.
 - Diseño y desarrollo de una Plataforma Educativa en Moodle para las Cátedras Estructura y Funcionamiento de Computadoras I –Dto. de Informática– y Química orgánica Dto Biología. FCEFyN. www.exactasad.org/
 - Diseño e implementación del Blog de la cátedra: Estructura y Funcionamiento de Computadoras I –Dto. de Informática– FCEFyN.
 - Desarrollo y aplicación de una WEBQUEST de la Modalidad a Distancia para

la Cátedra Teoría de la Computación. Tercer año de la LCC. FCEFyN

Participación en Congresos y Eventos científicos y publicaciones científicas

- Exposición de la Estrategia de gestión de cátedra, en el Ateneo del Programa de Educación a Distancia de la Facultad de C,E,F y N – UNSJ – Oct. 2008

- Ponencia *La Gestión de calidad en las Cátedras*, en el I Congreso Virtual Iberoamericano de Calidad en Educación a Distancia - EduQ@2008.. CD ISBN: 978-987-24871-0-2

- Ponencias presentadas para el II Congreso Internacional de Educación Lenguaje y Sociedad. Abril 2009 (Aceptadas):

○ *Importancia de los registros en la Gestión de Cátedra.*

○ *Una nueva Visión de los equipos de cátedras universitarias. Estrategias para afrontar los desafíos de la realidad educativa actual.*

○ *Una Propuesta para la incorporación de las herramientas tecnológicas en el proceso de enseñanza aprendizaje y en la gestión de cátedra.*

- Presentación de artículo: *Aportes para la gestión de cátedra en el proceso de cambio de la presencialidad a la distancia* Revista N°7 RUEDA ISS1666-1354. (En revisión)

• Resultados en curso

○ Puesta a punto del Modelo de Gestión de cátedra y de la Estrategia de gestión de calidad propuesta

○ Diseño y prototipación de la herramienta tecnológica para gestionar la estrategia de gestión de calidad de cátedra.

○ Preparación de los materiales didácticos para el Módulo “Gestión de calidad de cátedra” en el curso *Diseño de Prácticas Pedagógicas con Tecnologías web 2.0 para EAD*”.

4. FORMACION DE RECURSOS HUMANOS

Formación de los integrantes del proyecto:

○ Maestría en tecnología Aplicada a la educación. U.N.La Plata. Tesis en desarrollo: *El problema de enseñar y aprender Ciencias Naturales en entornos educativos virtuales* .- *Diseño de un asistente para facilitar la mediación pedagógica y comunicacional*

○ Cursado de Maestría en Informática U. La Matanza. y UNSJ.

○ Asistencia a las Jornada sobre “Estrategias de Enseñanza y Aprendizaje”: Aplicación del Enfoque de Competencias docentes en el Marco del Proyecto de mejoramiento de la Formación Docente.

○ Curso *Diseño de Prácticas Pedagógicas con Tecnologías WEB 2.0 para EaD.* Aprobado

○ Talleres internos sobre la Gestión de Calidad de las Organizaciones: “Marco Normativo Existente en torno a la Gestión de Calidad” (Abril 2008) y “Marco Normativo de la Norma IRAM 3000” (Mayo 2008)

Formación de docentes y otros

○ La directora del proyecto fue Docente e intervino en la producción de los materiales didácticos del Curso de Posgrado: *Diseño de Prácticas Pedagógicas con Tecnologías web 2.0 para EAD*” Organizado por Programa de Educación a Distancia y Departamento de Posgrado, Res. N° 48/08 CD-FCEFN (junio-agosto 2008). Res. N° 97/08 CD- 467/08 FCEFN (octubre-diciembre 2008).

○ Integrantes del proyecto dictaron Seminario *La Filosofía de la Calidad y los Sistemas Normativos. Instituto Superior de Gestión de la Calidad* - Ministerio de Educación de San Juan.

○ Integrantes del proyecto dictaron Seminario: “*Gestión por Procesos*” y “*Desarrollo de Procedimientos*”. Plan de Capacitación Interna de Transportes Sur Express S.A.

○ La directora del proyecto disertó las Conferencia “*Integración de recursos web2.0 en educación*” “*Ser docente, Hoy*” en la Esc. Normal Sup. G. M Belgrano Nivel Superior de Caucete. Febrero 2008

Formación a realizar

Se prevé para el 2009:

○ Incorporación de becarios y tesistas para el diseño y prototipación de la herramienta tecnológica que permitirá gestionar la estrategia propuesta.

○ Incorporación del módulo “Gestión de calidad de cátedra” en el curso *Diseño de Prácticas Pedagógicas con Tecnologías web 2.0 para EAD*”

5. BIBLIOGRAFIA

- Ander-Egg, Ezequiel. (1996). *La planificación educativa: conceptos: métodos: estrategias y técnicas para educadores*. Buenos Aires, AR : Magisterio del Río de La Plata, . — 298 p. ISBN: 950-550-110-2
- Enriquez Eugène. *El trabajo de la muerte en las instituciones*. Kaez R. y otros *La institución y las instituciones*. Bs.As., Paidós, p.84-115
- Farjat, L. (1998). *Gestión Educativa institucional*. Buenos Aires. Lugar Editorial.
- Frigerio, G, Poggi, M. y otros.(1992) *Las Instituciones educativas. Cara y Seca. Elementos para su Gestión*. Bs. As. Troquel. Educación. Serie FLACSO-Acción.
- Garcia Leiva, L.A. *Gestión Educativa*. <http://www.prismaseducativos.com.ar/gestion.htm>
- Norma de calidad GLOBAL GAP Disponible en: http://www.globalgap.org/cms/front_content.php?client=1&changelang=3&parent=&subid=&idcat=3
- Infante Malachias, M. E., Miranda Correia, P. R. (Septiembre 2007). *Proyectos Interdisciplinarios: Superando barreras epistemológicas*. Bs As. Revista Novedades educativas, ISSN 03283534. Pag 52-65
- Norma de calidad ISO 9000:2000 .Disponible en: http://www.uladech.edu.pe/webuladech/universidad/NORMA%20ISO_9000-2000%20CONCEPTOS%20Y%20VOCABULARIO.pdf
- Leiva, A, Pósito R, Ferrari, N, Millan, F, Pelayes, S, Agüero, N. 2008. *La Gestión de Calidad en las Cátedras*. I Congreso Virtual Iberoamericano de Calidad en Educación a Distancia EduQ@2008. CD ISBN: 978-987-24871-0-2
- Litwin, Edith. (2000). *La educación a distancia*, Bs As .Amorortu ediciones.
- Pósito, R. (2007). *Camino de un equipo de trabajo*. Conferencia presentada a integrantes del Programa de educación a Distancia ExactasAD. Facultad de Ciencias Exactas F y N. UNSJ. San Juan. Disponible en http://docs.google.com/Presentation?id=dg5cpj6p_298c6d5ctcd
- Pósito, R. (Setiembre 2004). *Un entorno virtual, ¿propicia la inteligencia exitosa?* presentada en el VIII Congreso de educación a Distancia CREAD MERCOSUR/sul . Cordoba Publicado en Actas.
- Pósito, R. (2004) *Soluciones Integradas para Sistemas de Educación a Distancia*, 9ª Jornadas de Informática Educativa, 5º Congreso de Informática. San Juan Argentina, Publicado Soporte Digital.
- Pósito, R. (2006) *Acerca de la problemática del diseño de unidades didácticas para EAD*. Presentado en IV Seminario Internacional II Encuentro Nacional de Educación a Distancia. ¿Edudiseños o Tecnodesignios? CD ISBN-10-987-22880-0-3
- Pósito, R. (2007) *Una mirada reflexiva de la práctica docente desde el aporte de las Teorías de la Inteligencia humana*. Presentado en 10º Jornadas de Informática Educativa, Congreso de Informática del Nuevo Cuyo. San Juan. Disponible en www.PortalZonda.com.ar/Congreso
- Pósito, R. (2007) *Re-Significar la práctica docente de las Ciencias Naturales en los Nuevos Ambientes Educativos*. Ciclo de Conferencias sobre Problemáticas: Un paso más hacia la profesionalización docente. Organizado por Esc. Normal Sup. Fray Justo Sta María de Oro Jachal San Juan.
- Pósito, R. (2008). *Integración de recursos web2.0. Ser docente, Hoy*. Conferencia. en el marco del Proyecto de capacitación docente Esc. Normal Sup. M. Belgrano. Caucete. San Juan.
- Tilleía Perez,D (Septiembre 2007).*Proyectos Interdisciplinarios: Abrir las puertas al pensamiento complejo*, Bs As . Revista Novedades educativas, ISSN 03283534. Pag 58 -60

INTERVENCIÓN DIDÁCTICA PARA EDUCACIÓN A DISTANCIA: Reconfiguración de Medios educativos según trama disciplinar.

Gonzalez Liliana Mirna---Cattapan Adela
Programa Permanente de Educación a Distancia
Dpto de Informática---Dpto de Geofísica
Facultad de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de San Juan

CONTEXTO

El Proyecto **Intervención Didáctica para Educación a Distancia: Reconfiguración de Medios educativos según trama disciplinar** se propone identificar formas de representación y recuperación del conocimiento en diferentes disciplinas. Y a partir de ello, generar conocimiento que facilite el análisis de diferentes medios educativos según su funcionalidad y pertinencia para distintas áreas disciplinares en la modalidad a distancia, con el objeto de garantizar la calidad de las intervenciones didácticas.

En virtud de esto, se vincula con el Proyecto "Calidad en el Diseño y Producción de Materiales de Educación a Distancia"; ambos Proyectos enmarcados en el Programa Permanente de Educación a Distancia creado por Resolución 08/05 de la Facultad de Ciencias Exactas, Físicas y Naturales. El Programa se constituye en un espacio académico y de investigación, que se sustenta en la conformación de equipos de trabajo interdisciplinarios abocados a la tarea de generar experiencia, conocimientos, soporte tecnológico-comunicacional y organizativo a las propuestas académicas bajo la modalidad a distancia en nuestra Facultad.

El propósito del mencionado Programa es:

- Construir un marco de referencia que permita sistematizar los conocimientos multidisciplinares que involucra la modalidad Educación a Distancia
- Potenciar la capacidad tecnológica de la FCEfyN manifiesta en recursos humanos y equipamiento orientados a la Educación a Distancia

RESUMEN

La incidencia de las nuevas tecnologías en Educación está indicando el surgimiento de un nuevo paradigma pedagógico-tecnológico. La consolidación del uso de estas tecnologías ha generado nuevas relaciones entre los actores que intervienen en una situación de enseñanza, específicamente en la modalidad a distancia: se han replanteado y resignificado los roles docente y alumno, la función de los medios de enseñanza y la pertinencia de los recursos tecnológicos en esta modalidad. Este Proyecto busca develar las formas de representación y recuperación del conocimiento que se ponen en juego en diferentes tramas disciplinares en la modalidad a distancia. Cada disciplina requiere experiencias, simulaciones y otro tipo de ejercicios que sino fuera por los medios de enseñanza no se podrían ofrecer al alumno. El objeto de estudio son los medios de enseñanza, sus atributos y dimensiones según distintas disciplinas para EaD.

En EaD, el diálogo debe reflejarse no sólo en las interacciones comunicacionales sino también en la relación que se genera a través de los medios de enseñanza. Área Moreira (doc. Inédito) considera que los medios son "componentes sustantivos" de todo proceso de enseñanza. Tienen el rol de potenciadores de habilidades intelectuales, requieren que los estudiantes realicen una serie de operaciones cognitivas y pongan en juego estrategias que les faciliten la decodificación de los sistemas simbólicos representados.

Palabras clave: *Intervención Didáctica- Educación a Distancia- Medios Educativos- Trama Disciplinar.*

1. INTRODUCCION

Las situaciones de enseñanza se concretan en la clase, la cual constituye un "espacio educativo" en el que se dan una red de interacciones e intercambios (conocimientos, alumnos, mediatizado por la intervención docente y tecnológica); este tipo de intercambios configuran un "diálogo didáctico"(Gallino, 2007), en el que coexisten el proceso de enseñanza, el de aprendizaje y de comunicación. En este diálogo el docente busca "causar un efecto" (Sarason2002, en Sabulsky), tiene una intencionalidad. En la intervención didáctica se conjugan Procesos cognitivos simples y/o complejos, el Objeto de Conocimiento y los Procesos de Comunicación. El docente debe reconocer las potencialidades, aportes y el rol que juegan los medios en todo proceso de aprendizaje. Los medios flexibilizan y mediatizan esos procesos.

El docente asume el rol de "facilitador de la gestión de conocimientos con intervenciones"(Gallino, 2007) didácticas mediatizadas por tecnologías. Tiene el rol de "promover y acompañar el aprendizaje" del alumno, le debe brindar elementos teórico-metodológicos que contribuyan al desarrollo de competencias que lo habiliten para la vida. Al alumno le corresponde ser "protagonista" de este proceso de crecimiento interior y de desarrollo de habilidades y competencias. Escudero (en Área Moreira, 1994): considera como " medio de enseñanza es cualquier recurso tecnológico que articula en un determinado sistema de símbolos ciertos mensajes con propósitos instructivos"¹. El medio es un recurso tecnológico que soporta algún tipo de sistema de símbolos, que comunica cierto mensaje y que tiene una intencionalidad educativa. Para Escudero (1983) es posible identificar 2 elementos en todo medio:
"los modos de construcción

"los sistemas de símbolos que articulan de manera particular el código del mensaje en función de la especificidad de ese código. Área Moreira² reconoce componentes y dimensiones de los medios: El hardware, el soporte físico, material, técnico y el software, la parte blanda, el mensaje, los programas son componentes que han de identificarse en todo medio; que tiene 3 dimensiones:
"semántica que refiere a lo que el "medio dice";
"sintáctica, "cómo el medio presenta" y finalmente
"pragmática que refiere al "para qué del medio".

La elección de un medio supone una intencionalidad de quien enseña hacia quien aprende, los efectos que cada medio tiene sobre el aprendizaje depende de variables internas, de los sujetos así como del contexto. Los medios codifican el conocimiento de manera particular, pero a la vez cada sujeto tiene capacidades, condiciones, habilidades singulares para la decodificación de ese sistema de símbolos. En ese proceso de decodificación y de aprehensión del referente el sujeto se compromete de manera particular; "el problema reside, por tanto, en la transformación de las representaciones simbólicas externas a representaciones cognitivas internas, deduciéndose de aquí que el grado de semejanza, correspondencia e isomorfismo entre el modo de codificación de los mensajes y el modo de representación interna de los mismos influirá en la facilitación o no del aprendizaje"³. Traemos a colación los conceptos de Perkins y Simmos (1988) que aluden a "sistema de procesamiento de conocimientos"⁴ y de "marco de acceso". Aquí se mencionan las

² Op. cit.

³ Op. cit.

⁴ Perkins D.N; La persona - más: una visión distribuida del pensamiento y el aprendizaje; en en Cogniciones distribuidas, Consideraciones psicológicas y educativas, Buenos Aires, Amorrortu Editores, página 128

¹ Área Moreira Manuel Los Medios de Enseñanza: conceptualización y tipología. Doc. inédito. Asignatura de Tecnología Educativa. Universidad de la Laguna. Web de Tecnología Educativa

categorías que en conjunto se llaman "marco de acceso": "Conocimiento: el alumno aporta conocimientos previos (disciplinares o no), los conocimientos que le brinda libros, revistas y otra bibliografía. "Representación: el alumno utiliza todo tipo de representaciones como diagramas, mapas conceptuales, herramientas computacionales que le faciliten la representación de los conocimientos.

"Recuperación: puede utilizar índices, distintos instrumentos que le ayuden a recuperar información y/o conocimientos acerca de un tema.

"Construcción: el alumno elabora sus trabajos y estudia en un contexto que le brinda recursos para que realice con éxito las operaciones cognitivas necesarias en todo proceso de conocimientos. Estas categorías nos permiten explicar las fases de procesamiento de conocimiento y entender el proceso de aprendizaje humano. A la luz de ellas podemos analizar la forma como un alumno distribuye el pensamiento y el aprendizaje que logrará y la relevancia de los medios de que dispone. No debemos olvidar las particularidades disciplinares, al momento de incorporar los materiales. La mediación de los materiales debe contemplar las características que requiere cada disciplina. Kuhn sugiere hablar de "matriz disciplinaria"⁵: "disciplinaria" porque se refiere a la posesión común de quienes practican una disciplina particular; "matriz" porque está compuesta por elementos ordenados de varias índoles..." "...incluye entre sus componentes generalizaciones simbólicas, compromisos compartidos de creencias en modelos concretos, valores compartidos y ejemplares compartidos."

En referencia a esta temática Morin expresa "La disciplina es una categoría organizadora

dentro del conocimiento científico..." "...una disciplina tiende naturalmente, a la autonomía, por medio de la delimitación de sus fronteras, por el lenguaje que se da, por las técnicas que tiene que elaborar o utilizar y, eventualmente, por las teorías propias..."⁶. La matriz disciplinar es lo que subyace en las prácticas pedagógicas propias de cada área del conocimiento y configura el sistema de procesamiento de conocimiento.

2. LINEAS DE INVESTIGACION y DESARROLLO

Este Proyecto encuadrado en el Programa Permanente de Educación a Distancia se ubica en la línea de investigación: **Los medios, materiales o recursos de enseñanza.**

Se busca generar conocimientos que aporten a la problemática de los medios de enseñanza adecuados a la modalidad a distancia y formar equipos interdisciplinarios que se aboquen al análisis y estudio de la temática desde la visión de diferentes tramas disciplinares.

Los ejes centrales de investigación giran entorno de:

- las formas de representación y recuperación del conocimiento que se ponen en juego en diferentes disciplinas en la modalidad a distancia.
- los atributos y dimensiones de los medios según distintas disciplinas, específicamente Ciencias de la Computación, Lenguas Extranjeras, Ciencias Naturales y Ciencias Sociales.

3. RESULTADOS OBTENIDOS/ESPERADOS

- **Publicaciones Científicas con Referato.**
 - Blogs en la formación de competencias para la gestión de información. Revista Tecnología y Comunicación Educativa. ILCE. 2008. México.
 - Tipología de Estrategias de Aprendizaje para la Comprensión Lectora en Inglés en la Modalidad EaD. Revista de Educación a Distancia (RED) N° 20.

⁵ Kuhn Thomas(1988). La estructura se las revoluciones científicas. Argentina. F.C.E. Pág.279-280

⁶ Morin, Edgar (2002). La Cabeza bien puesta. Repensar la reforma. Reformar el pensamiento. Buenos Aires. Ediciones Nueva Visión.

Publicación en Línea. Murcia, España. Año VIII Número 20. Septiembre de 2008. <http://www.um.es/ead/red/20/>

▪ **Participación y Asistencia a Eventos Científicos:**

- Gestión de Información y Blogs: Tipología de alumnos. Anales 5ta Jornada de Informática y Educación- Universidad Nacional de Villa María. Córdoba. noviembre 2008.
- La Gestión de calidad en las Cátedras. I Congreso Virtual Iberoamericano de Calidad en Educación a Distancia - EduQ@2008.. Octubre 2008.
- Análisis de una Intervención Didáctica en modalidad Semi-Presencial. II Congreso Internacional Educación, Lenguaje y Sociedad "La educación en los nuevos escenarios socioculturales". Gral Pico. La Pampa. abril 2009.
- Una nueva Visión de los equipos de cátedras universitarias. Estrategias para afrontar los desafíos de la realidad educativa actual. ídem anterior
- Asistencia a la LVIII Reunión de Comunicaciones Científicas, XXXI Reunión de Educación Matemática y XX Reunión de Estudiantes de Matemática. U.N.C. 2008
- Participación en Ateneos Inter-Proyectos del Programa Permanente de Educación a Distancias de la FCEFYN.

▪ **Acciones de Transferencia/Producción de Materiales**

1. Diseño e Implementación en Moodle de Unidades Didácticas para:

- Curso de Formación de Tutores en Moodle. <http://moodle.exactasad.org/course/view.php?id=52>

- Ciencia, Tecnología y Sociedad I. <http://campus.exactasad.org/course/view.php?id=5>
- Ciencia, Tecnología y Sociedad II, ambas de la Lic. en Sistemas de Información. <http://campus.exactasad.org/course/view.php?id=26>
- Computación (Lic. en Geofísica).
- Computación I (Lic. en Astronomía). <http://campus.adrenalinix.org/inicio/courses/view.php?id=6>

2. Diseño e Implementación de Materiales de Aprendizaje para Cursos a Distancia.

- Wesquest: Formación de Competencias. http://phpwebquest.org/wq25/webquest/soporte_tabbed_w.php?id_actividad=67178&id_pagina=1
- Diseño de Asistente Pedagógico para la elaboración de prácticas de Aprendizaje. En el Marco de la tesis de la maestría en tecnología Informática Aplicada a la educación. *El problema de enseñar y aprender Ciencias Naturales en entornos educativos virtuales*. En desarrollo.
- Wesquest: Palabras con -ed en textos científicos. Diseñada e implementada para la Cátedra Inglés Nivel 1 .Dpto. de Biología. http://phpwebquest.org/newphp/webquest/soporte_tablon_w.php?id_actividad=1459&id_pagina01
- Diseño e Implementación en Moodle de Unidades Didácticas n° 19 Hormonas de la Cátedra Química Orgánica. Dpto. Biología. 2008.
- Exelearning Prácticas de Aprendizaje.
- Diseño e implementación Blog ¡Error! Referencia de hipervínculo no válida. Escuela Normal Superior General Manuel Belgrano, Nivel Terciario.
- Diseño e implementación Blog La tecnología Web como herramienta didáctica". <http://seminariodeensenanza2.blogspot.com>. Cátedra "Seminario de Enseñanza II".

4. FORMACION DE RECURSOS HUMANOS

- Integrantes del Proyecto Diseñaron e Implementaron Unidades Didácticas para el Curso de Postgrado “Diseño de Prácticas Pedagógicas con Tecnologías web 2.0 para EAD” Postgrado, FCEFYN, UNSJ, 2008, Res. N^o 48/08 CD-FCEFYN (junio-agosto 2008); Res. N^o 97/08 CD- 467/08 FCEFYN(octubre-diciembre 2008)
- La directora del Proyecto fue Docente Responsable Curso de Actualización: Diseño de Prácticas Pedagógicas con Tecnologías Web 2.0 para EAD” Organizado por Departamento de Posgrado, FCEFYN, UNSJ, 2008, Res. N^o 97/08 CD-467/08 FCEFYN(octubre-diciembre 2008)
- Integrantes del Proyecto fueron responsables de módulo de los Cursos:
 - * Diseño de Prácticas Pedagógicas con Tecnologías Web 2.0 para EAD”
 - * Tecnologías Web y Educación: Formación de Competencias para la Gestión de Información. Organizado por Departamento de Posgrado, FCEFYN, UNSJ, 2007, Res. N^o 58/07 CD-FCEFYN (marzo-junio 2008)
- Miembros de este equipo están cursando:
 - Maestría Procesos Educativos mediados por Tecnologías UNC.
 - Maestría tecnología Informática Aplicada a la educación de la UNLP: Desarrollo de Tesis:
 - El problema de enseñar y aprender Ciencias Naturales en entornos virtuales. diseño de un asistente para facilitar la mediación pedagógica comunicacional.
 - Una propuesta tecnológica para mejorar la enseñanza de la comprensión del lenguaje simbólico matemático aplicando Tics en el nivel polimodal.
- Cursos Aprobados por miembros del equipo:
 - Taller Tecnologías Web 2.0 y Educación: Exe-Learning FCEFyN
 - Curso de Actualización: Diseño de Prácticas Pedagógicas con

Tecnologías Web 2.0 para EAD” Postgrado, FCEFYN, UNSJ, 2008, Res. N^o 97/08 CD- 467/08 FCEFYN(octubre-diciembre 2008)

- "Tutor Virtual y Enseñanza de lenguas. Roles y desafíos". 2008; con evaluación. U.N.C. Dpto. de Lenguas Extranjeras.
- Estadística en las Ciencias Sociales con el manejo del SSPS a cargo de las Dras Adriana Mallea y Miriam Herrera. FFHA7-

5. BIBLIOGRAFIA

1. Área Moreira, M. (1990). Los medios de enseñanza: Conceptualización y tipología. Doc. inédito: Asignatura Tecnología Educativa, Universidad de La Laguna. Web de tecnología educativa.
http://www.uclm.es/PROFESORADO/RICARDO/Clasificaciones_medios/doc_ConcepMed.html
2. Área Moreira, M. (2001) “Educar en la Sociedad de la Información”. Ed. Desclée. 2001. España
3. Perkins D.N; La persona - más: una visión distribuida del pensamiento y el aprendizaje; en Cogniciones distribuidas. Buenos Aires. Amorrortu Editores, página 128.
4. Kuhn Thomas (1988). La estructura se las revoluciones científicas. Argentina. F.C.E. pág.279-280
5. Morin, Edgar (2002). La Cabeza bien puesta. Repensar la reforma. Reformar el pensamiento. Buenos Aires. Ediciones Nueva Visión.
6. Litwin, Edith. Tecnologías educativas en tiempos de Internet. Amorrortu Editores. 2005. Argentina.
7. Salomón, Gabriel (2001) Cogniciones distribuidas. Amorrortu ediciones. Bs. As.
8. Kaplún, G. (2005). Aprender y enseñar en tiempos de Internet. Capítulo 3. La pedagogía de la EaD

con NTIC: ¿Transmisión o construcción de conocimientos? Montevideo: CINTERFOR/OIT, 2005. 197 p. ISBN 92-9088-199-2.<http://www.cinterfor.org.uy/public/spanish/region/ampro/cinterfor/publ/kaplun/pdf/cap3.pdf>.

Educación Basada en la Web

Perla Señas – Mercedes Vitturini – Laura Benedetti – Carolina Fernández Coria – Marcelo Zanconi

Laboratorio de Investigación y Desarrollo en Informática y Educación (LIDInE)
Departamento de Ciencias de la Computación. Universidad Nacional del Sur
Bahía Blanca. Argentina - psenas@cs.uns.edu.ar

Resumen

Se presenta en este trabajo el estado de avance de un proyecto trienal que comenzó a ejecutarse en 2008. Es la continuación de otros dos proyectos relacionados con las temáticas que relacionan las TICs con los procesos educativos, todos desarrollados en el Laboratorio de Investigación y Desarrollo en Informática y Educación (LIDInE) de la Universidad Nacional del Sur y financiados por su Secretaría General de Ciencia Tecnología. El objetivo general de este proyecto apunta a satisfacer las actuales expectativas sobre los sistemas de aprendizaje basados en la Web (SABW). Las investigaciones se orientan a temas relacionados con la Web Semántica, los Objetos de Aprendizaje, los Almacenes para alojarlos y las Web Quest. Las arquitecturas basadas en ontologías y orientadas a conceptos y la tecnología móvil son ítems prometedores para el desarrollo de tales sistemas. Las búsquedas de organizaciones basadas en lo conceptual son valoradas para organizar, procesar, y visualizar los dominios de conocimiento en los SABW. Además, los procesos de educación basada en la Web (EBW), en el caso de desarrollarse con presencialidad parcial o nula, requieren de metodologías didácticas especiales y de nuevas herramientas tecnológicas que asistan a los docentes en sus quehaceres específicos.

Palabras clave: e-learning - m-learning – EBW – Web Semántica - SABW

1. Introducción

Esta investigación continúa con los estudios realizados en el marco de los proyectos “Agentes Pedagógicos para Sistemas de Aprendizaje Interactivos” y “Aprendizajes basados en la Web”, que finalizaron en diciembre de 2004 y diciembre de 2007 respectivamente. La orientación actual es hacia las aplicaciones educativas basadas en la Web, tanto en lo referente a e-learning como a m-learning, fundamentalmente con grados de presencialidad parcial o nula.

El aprendizaje y la enseñanza basados en la Web introducen nuevas variaciones en los modelos o supuestos de la educación formal. El aprender a aprender, el aprendizaje autónomo, las comunidades de aprendizaje, la formación continua, la promoción de un auténtico interés en el alumno y el aprendizaje colaborativo han adquirido relevancia notoria. A ello se agrega la cognición e información situada, así como la inteligencia distribuida, procesos que permiten que solidariamente se aborde la identificación de problemas y el planeamiento y ejecución colectiva de las opciones más productivas de solución a los mismos. Todo ello presiona para la definición de un nuevo paradigma educativo en el que las Ciencias de la Computación y las Ciencias de la Educación tienen mucho que aportar [3],[4],[5],[9].

Permanentemente se buscan superaciones tecnológicas para poder diseñar y fundamentalmente mantener ambientes de aprendizaje personalizados, con contenidos y materiales apropiados para las aspiraciones y necesidades de cada uno. Para que ello sea posible quedan aún muchos problemas por resolver, entre otros los relacionados con la evaluación, el seguimiento permanente de los aprendices, la relación número de estudiantes vs. número de tutores, la organización conceptual de los contenidos y su contextualización en la red [15],[29]. Actualmente se persiguen dos objetivos generales: Por un lado se buscan esquemas para la representación de conocimiento en Agentes Pedagógicos que sean lo suficientemente flexibles como para ser usados por los expertos de materia, formas precisas para su especificación de tal forma de posibilitar el razonamiento

automático y representaciones gráficas adecuadas para su visualización [17],[26]. Por otro lado se busca desarrollar esquemas de SABW, siguiendo los lineamientos de la Web Semántica y que sean aptos para abordar una política de formación continua para profesionales y para el dictado de cursos con diferentes grados de presencialidad, con acceso tecnológico fijo y/o móvil [8],[18]. En tal sentido cobran un renovado valor los mapas conceptuales hipermediales (MCH) [14].

2. Significado de la Investigación: Interés e importancia del tema

Para los SABW, los Agentes Pedagógicos son una propuesta muy interesante. Son agentes autónomos que apoyan el aprendizaje humano integrando junto con estudiantes, ambientes de aprendizaje interactivos, tienen capacidad para mantener un espectro amplio de interacciones instruccionales efectivas con los alumnos que componen el entorno de aprendizaje. Forman parte de sistemas donde colaboran agentes humanos y de software, integrando acción con instrucción. Son capaces de aprender, de proveer a los estudiantes retroalimentación continua durante su trabajo en el entorno, tienen capacidad de presentarse dando la sensación de estar vivos y de inducir en los aprendices los mismos tipos de respuestas afectivas que generan otro tipo de entes vivos [13],[16].

Vale también destacar que una de las actividades más recientes en los desarrollos orientados a la Web es el de la Web Semántica cuya finalidad es dotar de significado a todas las clases de información sobre la red. Un subconjunto importante de esa información lo representan los Objetos de Aprendizaje, que son recursos digitales que se pueden reutilizar en diferentes contextos para lograr un objetivo de aprendizaje particular. Algunos de los elementos propuestos para la Web Semántica son: XML, RDF, PICS, las ontologías y los agentes [1],[8],[18].

Desde el área de las Ciencias de Computación puede hacerse un aporte significativo al área de Educación, que vaya más allá de lo meramente operacional. En este sentido cobran gran interés los SABW, en particular aquellos diseñados como Sistemas Multiagentes Mixtos. En estos sistemas la representación de conocimiento tiene un doble propósito, permitir hacer razonamiento automatizado y ser un recurso pedagógico eficaz para la construcción del conocimiento en los seres humanos.

Desde el área de Ciencias de la Educación el aporte importante a la temática del proyecto, se centra en la adaptación de metodologías existentes y la creación de otras nuevas, todas relacionadas con las Didácticas Especiales en el contexto del paradigma educativo basado en la Web [30],[31].

3. Originalidad y Finalidades Específicas de la Propuesta

Con el aumento de la EBW, existe un incremento proporcional en las expectativas y los requisitos hacia los SABW. Una meta a alcanzar en las investigaciones actuales es el desarrollo de sistemas con mayor grado de adaptación e inteligencia, con soporte individual para los estudiantes, para que puedan lograr una mejor recuperación, evaluación, comprensión, y retención de la información y con soporte eficaz para lograr resolver los problemas y realizar las tareas que se les proponen. Las arquitecturas basadas en ontologías y orientadas a conceptos se constituyen como una opción prometedora en el desarrollo de tales sistemas. Encontrar nuevas organizaciones con base en lo conceptual, con potencial para organizar, procesar, y visualizar los dominios de conocimiento en los SABW sigue representando un desafío aún en la actualidad; las tareas de visualización y navegación basadas en conceptos permiten que el sistema ayude a los estudiantes a orientarse dentro del dominio formando su propia comprensión y asociación conceptual. La importancia de lo conceptual y del uso de ontologías en ambientes de aprendizaje está recibiendo una considerable atención; la organización de sistemas basada en conceptos dentro de ambientes educativos ha sido apoyada por varios investigadores. El aspecto común de sus propuestas es utilizar una representación explícita de un sistema de conceptos del dominio, acordado y bien fundado, para avanzar en la interoperabilidad y el conocimiento compartido. En esta propuesta se centra la atención en considerar a los MCH como soporte organizacional de SABW. En tal sentido en este proyecto se trabaja sobre los siguientes tópicos:

- a) Uso de estructuras conceptuales en los SABW, con el propósito de apoyar:
 - La organización y procesamiento del conocimiento: adquisición, sistematización y razonamiento
 - El aprendizaje colaborativo
 - La resolución de problemas
 - La autoría de courseware colaborativos
 - Las interacciones usuario-grupo, ya sea alumno o tutor
 - La navegación y exploración con guía en lo conceptual y su contextualización
 - La recuperación de la información
- b) Aspectos del diseño y de la implementación de SABW con base conceptual:
 - Sus arquitecturas y metodologías
 - Los lenguajes de especificación
 - La visualización de estructuras conceptuales
 - El uso compartido y el reuso de estructuras conceptuales
 - Las estructuras conceptuales y las herramientas de autoría
- c) Evaluación de SABW basados en una estructura conceptual

Permanentemente se buscan superaciones tecnológicas para poder diseñar y fundamentalmente mantener ambientes de aprendizaje personalizados, con contenidos y materiales apropiados para las aspiraciones y necesidades de cada uno. Para que ello sea posible quedan aún muchos problemas por resolver, entre otros los relacionados con la evaluación, con el seguimiento de los aprendices, con la relación número de estudiantes vs. número de tutores, con la organización conceptual de los contenidos y con su contextualización en la red. En tal sentido con este proyecto se pretenden lograr aportes relacionados con la Web Semántica y con el diseño de Agentes Pedagógicos, en lo referente a la organización del conocimiento desde lo conceptual, con aplicación en los SABW.

En el mundo son muchas las universidades que disponen de alguna forma de EBW, ya sea como campus virtual o formación on-line, lo que hace posible no sólo el apoyo de la clase presencial con el aula virtual, sino también el dictado de cursos enteros y la expedición de títulos de grado y de postgrado a través de este sistema. Para que la Universidad Argentina pueda competir seriamente en el marco de este modelo, entendemos que todo aporte como el presentado en este proyecto es de valor. En la UNS, en particular estas investigaciones se advierten como un aporte de interés para tareas relacionadas con la formación continua o con la articulación que se realiza entre diferentes niveles.

4. Líneas de Investigación

Actualmente se trabaja en las siguientes líneas de investigación:

- a) *Enseñanza de la programación en el contexto de la EBW*: Se trabaja en la implementación de entornos de aprendizaje aptos para el planteo de algoritmos como forma de resolución de un problema planteado. Se buscan entornos que ofrezcan recursos para la visualización de los conceptos fundamentales y la forma en que ellos se relacionan; esto con el objeto de brindarle a la comunidad de aprendizaje herramientas que colaboren con los procesos de análisis y comprensión del problema.
- b) *Herramientas tecnológicas para la EBW*: En esta línea se trabaja fundamentalmente en los problemas relacionados con el desgranamiento, con la evaluación y con el alto cociente horas tutor/alumno en cursos no presenciales de la EBW. Se buscan herramientas tecnológicas aptas para paliar dichas cuestiones.
- c) *Enseñanza de lenguas extranjeras en el marco de la EBW*: Esta línea comenzó como una aplicación concreta de los estudios del LIDInE en un área específica en la Universidad Nacional del

Sur: la enseñanza de la lengua inglesa. Ha permitido, en un marco de investigación-acción aplicar parte de los desarrollos tecnológicos realizados en el laboratorio. La importancia que cobró el tema dio lugar a la creación de un proyecto independiente sobre esta temática.

d) *Nuevos paradigmas para la EBW*: Las experiencias sobre el uso didáctico de los recursos telemáticos son aún pocas y limitadas. Esto permite articular sólo un primer análisis parcial sobre el real impacto que estos sistemas puedan llegar a tener en la didáctica colaborativa. Pero los problemas son muchos y urgentes: ¿cómo se ubican las actividades y los objetivos en función de la estrategia colaborativa adoptada y en relación al canal comunicativo elegido?, ¿cómo se deben individualizar las exigencias comunicativas para cada actividad?, ¿qué comunicaciones interpersonales conviene adoptar?, ¿cómo efectuar la transferencia de lo semi-elaborado? y finalmente, ¿cómo se debe elegir la tecnología de la comunicación más adecuada para cada una de las exigencias específicas?. En el LIDInE, en un marco de investigación –acción se trabaja para responder estos interrogantes.

e) *La EBW y la tecnología móvil*: Esta línea de investigación está basada en dos ejes: la EBW y la movilidad de los actores, entendiendo por actores a los que intervienen en todo el proceso de aprendizaje: los contenidistas, los profesores, los alumnos y el medio electrónico que brinda la posibilidad de comunicación y encuentro independizándolos de las variables espaciotemporales. El objetivo es brindar pautas tanto educacionales como tecnológicas para alcanzar la construcción de conocimientos y el desarrollo de competencias en el nuevo espacio social mediado por las TICs.

5. Resultados obtenidos/esperados

Durante la evolución de este proyecto se consolida el trabajo de los investigadores que participan en él. Se han presentado los resultados parciales obtenidos durante este primer año de trabajo en congresos nacionales e internacionales. Se concluyó con la implementación de una herramienta apta para los trabajos de evaluación en espacios de EBW, se diagramaron nuevas propuestas didácticas para SABW tanto para e-learning como m-learning y se sigue profundizado el estudio de los MCH como esquemas de representación de conocimiento en agentes pedagógicos.

6. Formación de recursos humanos

Se realiza un trabajo interdisciplinario en el que participan investigadores de Ciencias de la Computación, de Ciencias de la Educación y del área de Lenguas Extranjeras. Se trabajan temas relacionados con el proyecto en tesis de licenciatura, una de las cuales ya fue presentada. La formación de recursos humanos alcanza además a alumnos del Profesorado en Computación que realizan una introducción a la investigación sobre EBW durante el cursado de la Didáctica Especial. Se interactúa con investigadores del Laboratoire VERIMAG de la Université Joseph Fourier de Grenoble, Francia en un marco de Investigaciones Basadas en la Web (IBW) en lo que respecta a m-learning. También, a través de la dirección de tesis de maestría, con investigadores del Master Interuniversitario en Formación de un Profesorado de Calidad para la Docencia Preuniversitaria (MIFORCAL), proyecto que se encuadra en el Subprograma de Cooperación para la formación Científica y Técnica de la Unión Europea y América Latina y en el que intervienen once universidades de siete países diferentes, Se ha colaborado también con el Laboratorio de Investigación y Desarrollo en Informática (LIDI) de la Universidad de la Plata como jurados de tesis de postgrado.

4. Bibliografía

- [1] Angros, R., Scholer, A., Rickel, J. and W.L. Johnson. Teaching Animated Agents in Virtual Worlds. In AAAI Spring Symposium on Smart Graphics, Stanford, March 2000.
- [2] ARIADNE Foundation for the European Knowledge Pool. <http://www.ariadneeurope.org/>. Consult en 2008.
- [3] Bransford J., Brown A. y Cocking R. (Editores) (2000). How people learn. USA:

- [4] Committee on Developments in the Science of Learning - Commission on Behavioral and Social Sciences and Education - National Research Council. <http://books.nap.edu/html/howpeople1/>. Consultado en 2008.
- [5] Creating New Learning Experiences on a Global Scale: Second European Conference on Technology Enhanced Learning, EC-TEL 2007. Springer. 2007.
- [6] EOE Foundation. Educational Objects Economy: Building Communities that Build Knowledge, consultado en (<http://www.eoe.org>). 2003.
- [7] Extensible Markup Language (XML). Consultado en: (<http://www.w3.org/XML/>) octubre-2004.
- [8] Franconi, Kifer and May The Semantic Web: Research and Applications: 4th European Semantic Web Conference. Austria. 2007.
- [9] Garcia Penalvo. Advances in E-Learning: Experiences and Methodologies. 2007.
- [10] Johnson, W. and Shaw, E. Using Agents to Overcome Deficiencies in Web-Based Courseware. Proceedings of the AI-ED 97 Workshop on Pedagogical Agents. 1997.
- [11] Jorgensen, Shen, and Shu. Recent Advances in Computational Sciences: Selected Papers from the International Workshop on Computational Sciences and Its Education. 2007.
- [12] Lester, J. Mixed Initiative Problem Solving with Animated Pedagogical Agents. AI-ED97. Eighth World Conference on Artificial Intelligence in Education - Workshop V : Pedagogical Agents. Japan. 1997.
- [13] Lewis, W. and Shaw, E. Using Agents to Overcome Deficiencies in Web-Based Courseware. AI-ED97: World Conference on Artificial Intelligence in Education - Workshop: Pedagogical Agents. 1997.
- [14] Malet, A. y Señas P. Los Mapas Conceptuales Hipermediales y la construcción de conocimiento. V Conferencia Internacional de Ciencias de la Educación. Cuba. 1999.
- [15] MERLOT Multimedia Educational Resources for Learning and Online Teaching. Consultado en: (<http://www.merlot.org/Home.po>) octubre-2004.
- [16] Morin, J., Lelouche, R. Tutoring Knowledge Modelling as Pedagogical Agents in an ITS. AI-ED97. Eighth World Conference on Artificial Intelligence in Education. Japan, 1997.
- [17] Moroni, N., Vitturini, M., Zanconi, M., Señas, P. Una plataforma para el desarrollo de mapas conceptuales hipermediales. IV Jornadas Chilenas de Computación. Valdivia. 1996.
- [18] Nilsson, M., Pálmer, M. and Naeve, A., Semantic Web Metadata for e-Learning. Some Architectural Guidelines. (<http://www2002.org/CDROM/alternate/744/>) Consultado en 2003.
- [19] Ontology Inference Language (OIL). (<http://www.ontoknowledge.org/oil/>).
- [20] Shea, C., Lillis, O Shea & Collins. The application of e-learning and m-learning technology in the context of Life Long Learning in Irish Higher Education. In Kommers & Richards (Eds.). World Conference on Educational Multimedia, Hypermedia and Telecommunications. Chesapeake, 2006.
- [21] Reusser, Kurt. Tutoring Systems and Pedagogical Theory: Representational Tools for Understanding, Planning and Reflection in Problem Solving. En Computers as Cognitive Tools. Lajoie and Derry, ed. 1993.
- [22] Sanz, C., Zangara, A., Gonzalez, A., Ibañez, E., De Giusti, A. WebLIDI: Desarrollo de un Entorno de Aprendizaje en la WEB. CACIC-03. Argentina 2003.
- [23] Savarimuthu B. T. R., Purvis, M. A., Purvis, M. K., "Creating Ontologies for a Collaborative, Multi-agent-based Workflow System", Ontologies and Soft Methods in Knowledge Management, Katarzyniak R. (ed.), Advanced Knowledge International, Adelaide, Australia, 2006.
- [24] Savin-Baden. Problem Based Online Learning. Routledge Press. 2007.
- [25] SCORM Sharable Content Object Reference Model. Consultado en (<http://www.adlnet.org/>)
- [26] Señas, P. Tesis de Magíster: MCH como herramienta para la Representación de Conocimiento en Agentes Inteligentes. Universidad Nacional del Sur. 2000.
- [27] Smith, D., Cypher, A. and Spohrer, J. KidSim: Programming Agents without a Programming Language. En Software agents. Bradshaw, ed. California AAAIPress. 1997.
- [28] Sowa, J. Knowledge Representation: Logical, Philosophical, and Computational Foundations Brooks Cole. 2000.
- [29] Tarouco, L., Grando, A. and Pedroso Konrath, M. Projeto e produção de objetos educacionais usando conceitos de alfabetização visual. CACIC-04. Argentina. 2004.
- [30] Wegerif, Rupert Technology: Expanding the Space of Learning (Computer-Supported Collaborative Learning Series). Springer. 2007.
- [31] Wiley, D. A., Connecting learning objects to instructional design theory: A definition, a metaphor and a taxonomy. In D. Wiley (Ed.), The Instructional Use of Learning Objects. 2006.

CONTENIDOS DIDÁCTICOS Y SCORM PARA E-LEARNING EN CARRERAS DE GRADO

Vazquez, Alejandro

Universidad Tecnológica Nacional, Facultad Regional Mendoza

Ingeniería en Sistemas de Información

Licenciatura en Tecnología Educativa

M.B.A. (Maestría en Administración de Negocios)

CONTEXTO

Esta línea de investigación y desarrollo está relacionada con el proyecto de investigación “Implementación del e-learning en las carreras de grado de la UTN-FRM” vigente y aprobado en la Universidad Tecnológica Nacional, Facultad Regional Mendoza, Ingeniería en Sistemas de Información y con Tesina “Contenidos didácticos para e-learning y SCORM” de la Licenciatura en Tecnología Educativa, de la Universidad Tecnológica Nacional, Facultad Regional Mendoza.

RESUMEN

En este trabajo se analiza la utilización de SCORM¹ para mejorar la portabilidad de los materiales de estudio facilitando la gestión e implementación de proyectos educativos con e-learning.

SCORM ayuda a lograr lo que se persigue con la aplicación de un estándar para los contenidos e-learning: Durabilidad, Interoperabilidad, Accesibilidad, Reusabilidad. Esta compatibilidad ofrece muchas ventajas como proteger la inversión y aumentar la oferta de posibilidades de aprovechar los recursos de capacitación.

Según García Aretio, L. (2007) es correcto considerar los ambientes en el que el centro de su interés y preocupación sea el estudiante, lo que aprende y como lo aprende. Dadas las características o principios de la educación a distancia en cualquiera de sus manifestaciones o propuestas, el estudiante es más protagonista de su formación. Las dosis de autoaprendizaje suponen una ventaja con respecto a otras formas de incorporación de los saberes. El centro es el estudiante, el que aprende con sus propios estilos y ritmos, convirtiéndose el profesor en un facilitador de los aprendizajes. En esos ambientes se valoran especialmente los contenidos didácticos. SCORM permite aprovechar los recursos para los materiales de estudio en la implementación de proyectos de educativos con e-learning en las carreras de grado.

Palabras clave: S.C.O.R.M. - e-learning² - Ead³ - Educación a distancia – contenidos e-learning

¹ S.C.O.R.M.: Shareable Content Object Reference Model

² e-learning: Educación a distancia por Internet, utilizando plataformas educativas automatizadas, materiales didácticos

1. INTRODUCCION

El concepto de e-learning responde a una de las diferentes formas de hacer educación y formación a distancia, la que se sostiene en un entorno o contexto virtual creado para el aprendizaje, a través de Internet. La Comisión de la Comunidad Europea, como conclusión a su reunión de Lisboa del año 2000, publicó el documento “e-learning, concebir la educación del futuro”. El significado que se da a e-learning es el de “el uso de las tecnologías de información y la comunicación, incluido Internet, en la enseñanza y el aprendizaje.”

El e-learning⁴ equivale a la inevitable transformación de todas las formas de educación y aprendizaje en el siglo XXI. La influencia del e-learning en la educación tradicional no es débil y no puede ser ignorada, ya que la transformación que implica va más allá del mero aumento de la eficiencia y un tratamiento más personalizado, facilitado y mediado a través de las herramientas tecnológicas y de comunicación vinculada con Internet.

El e-learning no se contrapone a la formación tradicional: los recursos tecnológicos y los métodos didácticos se deben diseñar de acuerdo con las necesidades de los estudiantes. La tendencia habitual es desarrollar acciones formativas mixtas, poniendo el énfasis en la parte presencial o en el e-learning en función de los objetivos y el contexto del aprendizaje: Compatibilidad de horarios, localización geográfica, recursos económicos, etc. La preparación de materiales con soporte electrónico requiere un mayor esfuerzo de diseño y programación que la enseñanza tradicional, que se apoya más en las dotes personales del profesor. En este sentido, se comprende que el e-learning está asociado a la idea de mejora de la calidad de la formación y puede verse como una evolución más que como una revolución. La clave de la eficacia del e-learning está en la interactividad que se consiga: Del estudiante con los materiales formativos (mediante recursos técnicos para la realización de ejercicios con corrección automática, glosario, navegación, etc.), con

especialmente diseñados, herramientas de comunicación, seguimiento, apoyo y evaluación.

³ Ead: Educación a Distancia.

⁴ Sistema e-learning inteligente. Luis Alberto Casas, Profesor de la Universidad Nacional de San Agustín de Arequipa.

otros compañeros (aprendizaje colaborativo) y con el Docente y Tutor.

En este sentido, el "Estudio de las necesidades de formación de las organizaciones en España 2005", realizado por Millward Brown, muestra que el 95% de las empresas considera que la formación on-line no es una moda pasajera y que su interés va en aumento. Destaca asimismo el alto porcentaje de empresas (casi el 80%) que considera que las herramientas de e-learning proporcionan contenidos formativos adaptados a las necesidades de las empresas.

El conocimiento no proviene de afuera hacia adentro del sujeto, éste no es una tabla en blanco en el que la naturaleza escribe. El conocimiento se construye por acción entre el sujeto y el objeto que se logra por adaptaciones (cognitivas) sucesivas del sujeto; los genes tienen la capacidad de interactuar con el medio lo que le permite aumentar los intercambios con el medio; existe continuidad entre lo orgánico y lo cognitivo. (Piaget, J.W. 1971).

Uno de los pilares del e-learning es el de aprendizaje significativo. Según Ausubel, D.P. (1973), un aprendizaje es significativo cuando es incorporado al conjunto de conocimientos del sujeto, relacionándolo con sus conocimientos previos, por lo que puede definirse como el proceso que ocurre en el interior de la persona, donde la actividad perceptiva le permite incorporar nuevas ideas, hechos y circunstancias a su estructura cognoscitiva; además de matizarlas evidenciándolas por medio de acciones observables, comprobables y enriquecidas una vez que ha cumplido con las actividades derivadas de las estrategias de instrucción, planificadas por el facilitador y/o sus estrategias propias de aprendizaje.

No hay un estilo de aprendizaje⁵ mejor que otro, sino que lo importante es tratar de conocerlos y de adaptar nuestro estilo educativo de la mejor manera posible a las distintas condiciones y circunstancias de aprendizaje.

En e-learning es fundamental el autoaprendizaje, a través del cual, el estudio individual y el grupal se implementan con los hábitos de estudio y las herramientas de trabajo. Para ello deben ofrecerse las condiciones de trabajo adecuadas para que el estudiante pueda avanzar a un ritmo ajustado a sus intereses, capacidades y ocupaciones.

Los actores del e-learning (figura 1) tienen que trabajar en forma coordinada para el logro de los objetivos respecto de los estudiantes de carreras de grado.

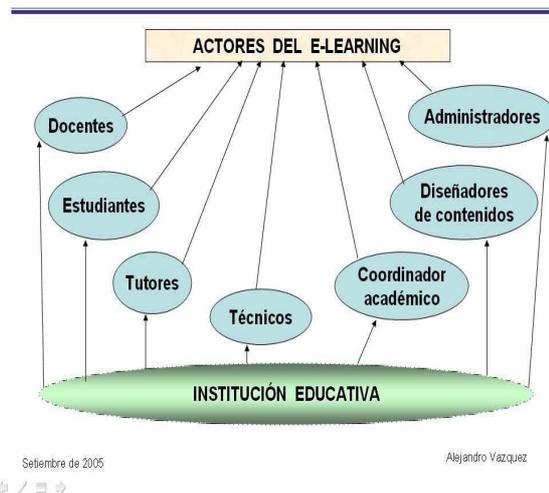


Figura 1. Actores del e-learning

El profundo cambio que ha supuesto Internet en el campo de las aplicaciones educativas no sólo afecta a aspectos puramente tecnológicos: también los paradigmas educativos se están alterando como consecuencia de las nuevas demandas sociales, en las que las redes globales de comunicación juegan un papel esencial. Los actores deben estar preparados para trabajar con estas tecnologías y ambientes. Actualmente se imponen los modelos educativos centrados en el estudiante: Ha crecido la demanda hacia una formación continua y "de por vida" en contraposición a los métodos educativos tradicionales en los que la formación se recibe en periodos determinados y centros específicos.

Los procesos educativos de e-learning tienen la clave de su éxito en la calidad de sus Docentes y Tutores, en la calidad de sus contenidos didácticos y en la habilidad para la utilización de los medios. La educación actual necesita de profesionales formados en el dominio adecuado de las técnicas y herramientas que hacen posible la implementación eficiente del e-learning.

Para evitar que el uso de las nuevas tecnologías se limite sólo a una mera trasmisión de datos más económica y rápida para consumo y adquisición por parte de los estudiantes, debería contemplarse una diversidad y flexibilidad de estrategias metodológicas en el diseño de materiales para la EaD, para adecuarse a los variados estilos de aprendizaje de los estudiantes, sus métodos de estudio y la necesidad de apoyo.

En este análisis no podemos dejar de lado que uno de los componentes principales de cualquier aplicación educativa son los contenidos educativos. Parece evidente que en última instancia, el éxito de una aplicación educativa radica en la calidad de sus contenidos. Por lo general, en la mayoría de las aplicaciones disponibles, los cursos se crean con el fin de cubrir una necesidad de aprendizaje concreta. Sin embargo, producir desde cero material educativo de alta calidad para e-learning es una labor ardua que lleva

⁵ Estilo de aprendizaje: La forma en que cada persona incorpora efectivamente los conocimientos y que depende de las características inherentes a cada uno y de su entorno específico. Cada persona aprende de manera distinta.

mucho tiempo y requiere conocimiento de diversos expertos en distintas disciplinas.

El material de estudio debe "cautivar" para ser leído y conducir al estudiante a formularse preguntas, organizarse, construir su propio aprendizaje y realizar actividades que hemos de diseñar apropiadamente para aprender "haciendo" y para "aprender a aprender".

De acuerdo con lo que explica Millward Brown (2005) "Estudio de las necesidades de formación de las organizaciones en España 2005": El desarrollo de contenidos es uno de los aspectos más importantes de un proyecto de e-learning. Durante esta fase deben considerarse las particularidades intrínsecas de esta modalidad formativa que no sólo implican un cambio importante en la metodología de enseñanza, sino que también amplían las posibilidades de personalización y adaptación de los entornos de aprendizaje a las necesidades particulares y específicas de cada usuario. El diseño y desarrollo de los contenidos que se vayan a incluir en un curso de formación on-line requieren de una adecuada estructuración y de una minuciosa planificación, que contribuirán a facilitar no sólo el proceso de aprendizaje por parte del estudiante, sino también las posibilidades de control y seguimiento por parte del formador.

Con el gran crecimiento del mercado del e-learning en todo el mundo, el incremento de los recursos didácticos disponibles, así como de herramientas para la creación de contenidos y de sistemas para su puesta en línea, ha surgido la necesidad de utilizar contenidos existentes en cualquier ambiente virtual, independientemente de la tecnología de la plataforma. Esto es posible gracias a la interoperabilidad que caracteriza los sistemas de e-learning conformes a estándares tecnológicos.

SCORM proporciona lineamientos de trabajo y una referencia de implementación detallada que permite a los contenidos y a los sistemas usar SCORM para "hablar" con otros sistemas, logrando así interoperabilidad, reusabilidad y adaptabilidad. Permite crear objetos pedagógicos estructurados. Los sistemas de gestión de contenidos en Web originales usaban formatos propietarios para los contenidos que distribuían, pero no era posible el intercambio de tales contenidos. Con SCORM se hace posible crear contenidos que puedan importarse dentro de sistemas de gestión de aprendizaje diferentes.

Los principales requerimientos que el modelo SCORM trata de satisfacer son:

- **Accesibilidad:** capacidad de acceder a los componentes de enseñanza desde un sitio distante a través de las tecnologías web, así como distribuirlos a otros sitios.
- **Adaptabilidad:** capacidad de personalizar la formación en función de las necesidades de las personas y organizaciones.
- **Durabilidad:** capacidad de resistir a la evolución de la tecnología sin necesitar una reconcepción, una reconfiguración o una reescritura del código.
- **Interoperabilidad:** capacidad de utilizarse en otro emplazamiento y con otro conjunto de herramientas o

sobre otra plataforma de componentes de enseñanza desarrolladas dentro de un sitio, con un cierto conjunto de herramientas o sobre una cierta plataforma. Existen numerosos niveles de interoperabilidad.

- **Reusabilidad:** flexibilidad que permite integrar componentes de enseñanza dentro de múltiples contextos y aplicaciones.

El modelo SCORM está compuesto por:

- **Modelo de Agregación de Contenidos (Content Aggregation Model),** que asegura métodos coherentes en materia de almacenamiento, de identificación, de condicionamiento de intercambios y de recuperación de contenidos.

- **Entorno de Ejecución (Run-Time Environment),** describe las exigencias sobre el sistema de gestión del aprendizaje que este debe implementar para que pueda gestionar el entorno de ejecución con el contenido SCORM.

- **Secuenciamiento y de navegación (Sequencing and Navigation),** permite una presentación dinámica del contenido. Describe cómo el sistema interpreta las reglas de secuenciamiento introducidas por un desarrollador de contenidos, así como los eventos de navegación lanzados por el estudiante o por el sistema.

El actual estadio del e-learning atribuye a la gestión de contenidos la tarea vital de incentivar y abanderar la innovación formativa o educativa. Sírvase de ejemplo la frenética actividad en adoptar acuerdos para la estandarización, como SCORM, Dublin Core e-learning, Learning Objects Multimedia (LOM), etc. El campo de la gestión de contenidos obtiene el papel protagonista en los procesos formativos-educativos mediados por la tecnología debido a que el acceso y uso de los contenidos es un recurso estratégico para aumentar la competitividad, la eficacia y la eficiencia en la calidad docente.

Como señala Lara Navarra, P. (2005), con el e-learning no estamos solamente introduciéndonos en una nueva tecnología del aprendizaje; se presenta una nueva forma de pensar acerca del aprendizaje. Las personas pueden aprender de formas diversas mediante el acceso a una información bien diseñada, por el uso de herramientas que mejoran el desempeño por medio de la experiencia y de otros factores. La interacción directa e ininterrumpida con los contenidos los convierte en artífices importantes del acto de aprendizaje, volviendo a demarcar la importancia de la gestión de los contenidos dentro de la educación mediada por tecnología.

El modelo SCORM consta de los siguientes elementos:

- SCORM Content Aggregation Model.
- SCORM Run-Time Environment.
- Sequencing and Navigation (incorporado en Scorm 2004).

El SCORM Content Aggregation Model define cómo se ensamblan los assets o archivos, constituyendo SCOs (Shareable Content Objects) y a su vez como estos SCOs forman Organizaciones.

Una Organización es un grupo de SCOs que conforman una lección, un capítulo o un curso.

CAM es un componente necesario para mover un contenido e-learning de un lugar a otro, los principales tópicos son: Content Model, Content Package, The Manifest, Meta-data.

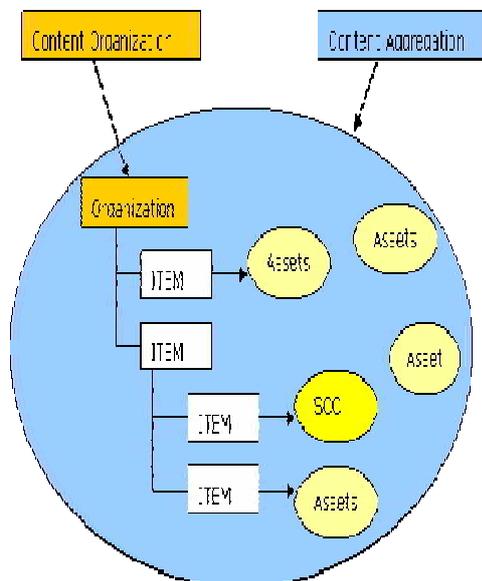


Figura 2: Elementos del modelo SCORM.

En la aplicación en un proyecto de “Utilización del e-learning en las cátedras de la carrera de grado “Ingeniería en Sistemas de Información”, como modalidad de apoyo educativo, en la UTN-FRM se realizó el siguiente análisis FODA:

<p>FORTALEZAS</p>	<ul style="list-style-type: none"> - El 11% de los Docentes están trabajando con e-learning con experiencia de años en el tema. - El 8% de los Docentes han aplicado e-learning en sus cátedras con muy buenos resultados. - Los que utilizan actualmente e-learning son apasionados del tema. - La plataforma educativa de elearning a utilizar (MOODLE) es la más utilizada en el mundo, tiene varios años de utilización en nuestra Facultad con excelentes prestaciones y aplicaciones en cursos, carreras de pregrado, grado (Licenciatura en Tecnología Educativa) y posgrado. - Está implementada la integración automática de la plataforma educativa con el servicio de correo electrónico de los Estudiantes y de los Docentes para que cuando haya participaciones, preguntas,
--------------------------	--

	<p>presentación de trabajos, etc. llegue el aviso automático a la casilla de e-mail.</p> <ul style="list-style-type: none"> - Hay un Servidor dedicado, lo que no lo hace depender de un proveedor externo. - Los Servidores de e-learning y sincronización con servidor de correo electrónico serán los mismos que actualmente se utilizan en el “Programa de Educación a Distancia” de la Facultad, lo que facilita su gestión. - Buen manejo de la tecnología por parte de los estudiantes y adaptación al cambio sin inconvenientes. - Laboratorios en el edificio de la Facultad que tiene PCs disponibles con adecuado tiempo de respuesta para aquellos que tengan tecnología limitada. - Para los Estudiantes y Docentes sólo se requiera tener acceso a una PC con cualquier tipo de conexión a Internet. - Hay 4 Docentes que están preparados para realizar materiales de estudio utilizando técnicas y recomendaciones acordadas para e-learning. - Algunas actividades didácticas se pueden adaptar con no mucho esfuerzo. - Algunos Docentes conocen métodos objetivos de evaluaciones y distintas formas de implementarlo en e-learning.
<p>OPORTUNIDADES</p>	<ul style="list-style-type: none"> - Mejorar la capacitación de Docentes en nuevas modalidades educativas y en manejo de plataforma educativa y/o tutoría y/o diseño de materiales para educación a distancia. - Aplicación de nuevas herramientas tecnológicas al servicio de la educación. - Potenciar el uso de la plataforma en la Facultad. - Mostrar a la U.T.N. y al medio los importantes resultados en el tema, desde una carrera de Ingeniería. - La plataforma educativa de elearning a utilizar está diseñada con “Orientación a Objetos” lo que es un ejemplo más que importante para los estudiantes en su integración de conocimientos

	<p>de negocios electrónicos, aplicación de Sistemas, Seguridad, tecnología, metodología de análisis y diseño, programación, etc.</p> <ul style="list-style-type: none"> - Realizar actividades prácticas o autoaprendizaje que antes se realizaba en forma presencial, con inconvenientes. - Capacitación de los Estudiantes en el uso de la plataforma educativa de e-learning. - Adecuación y actualización de materiales de estudio. - Realización de pruebas objetivas individuales y grupales con calificación automática de la plataforma educativa. - Formar recursos humanos para mejorar la integración, atención de los estudiantes y la calidad educativa, a través de la utilización del e-learning. - Mejorar la interacción entre Docentes y Estudiantes.
DEBILIDADES	<ul style="list-style-type: none"> - El 89% de los Docentes no ha trabajado con e-learning. - El 65% de los Docentes no genera cambios importantes en sus cátedras, por lo que no está acostumbrado a adaptarse en forma rápida. - El 92% de los Docentes no dispone de tiempo adicional a dedicar y no hay posibilidad de reconocimiento monetario adicional. - No hay un soporte técnico permanente por fuera de la plataforma educativa. - Bloqueos de servidor no detectados. - Insuficiente configuración de seguridad en el Sistema Operativo del Servidor. - La administración es ejercida por pocas personas, generando una cierta dependencia. - Ancho de banda de conexión a Internet insuficiente en puestos de trabajo de Docentes y Estudiantes. - Ancho de banda de conexión a Internet insuficiente en el ingreso a la Facultad. - Poco dominio de las técnicas de elaboración de materiales de estudio especializados para educación a distancia y de SCORM. - Hay pocas actividades didácticas

	<p>que sirven directamente para ser utilizadas en e-learning.</p> <ul style="list-style-type: none"> - En algunas cátedras no se respetan criterios de evaluación objetivos y uniformes en su plantel Docente. - Actualmente no es muy alto el nivel de integración entre las cátedras de los
AMENAZAS	<ul style="list-style-type: none"> - Que los usuarios presenten quejas permanentes por no entender su funcionamiento o que aparezcan problemas en los que no encuentren respuesta rápida. - El crecimiento por el uso de muchas más cátedras puede provocar saturación de los servidores. - Existe la posibilidad de que algunos estudiantes se opongan abiertamente. - Lentitud en las conexiones. - Imposibilidad de bajar ciertos archivos o ejecutar ciertos programas. - Que no haya disponibilidad para que durante los horarios de clases presenciales de las cátedras se pueda disponer de un laboratorio para que los estudiantes también tengan acceso al “aula virtual”. - Cortes de enlaces a Internet. - Que no se elaboren materiales especialmente diseñados, según modelo SCORM y mediatizados y eso afecte la motivación y uso de la plataforma educativa, o el descontento de los estudiantes. - Subir a la plataforma educativa a distancia las mismas actividades de la modalidad de educación presencial. - Realizar las mismas evaluaciones que en la modalidad de educación presencial.

2. LINEAS DE INVESTIGACION y DESARROLLO

- El notable avance del e-business e e-learning y su importancia estratégica en las organizaciones educativas y empresas.
- N.T.I.Cs.6 en e-learning.
- Diseño de contenidos didácticos para e-learning.
- La tecnología S.C.O.R.M. y su aplicación en la elaboración de materiales de estudio para e-learning.
- Integrar lo investigado en un proyecto real de e-learning en carreras de grado, como modalidad de apoyo educativo.

⁶ N.T.I.Cs. : Nuevas Tecnologías de Información y Comunicaciones.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se ha verificado que se está valorando cada día más la gran potencialidad que tienen las plataformas educativas en Internet y todas las ventajas de la educación a distancia por Internet (70% de los estudiantes consideran que el cambio de mentalidad para estudiar es el principal problema del e-learning superando ampliamente a los problemas técnicos en segundo lugar con un 25%). De la misma forma, se están logrando grandes avances en la capacitación de las personas para desempeñarse en cada una de las actividades sobre la base de los pilares del e-learning. De todas formas, aún sigue siendo la “Solidez, calidad institucional y el prestigio, reputación y diferenciación de la institución” lo que se tiene en cuenta para posicionar cursos en e-learning o para seleccionarlos (según los datos obtenidos en las encuestas realizadas).

Se concluye que el éxito de una aplicación educativa radica en gran medida en la calidad de los materiales de estudio. Por ello son de gran importancia las características, recomendaciones y pautas, respecto del diseño de contenidos educativos, como por ejemplo las 10 pautas de Lara Navarra, P. (2005). Aunque, según los resultados de las encuestas realizadas, en nuestro medio aún no está arraigada la importancia de un diseño de alta calidad de los materiales de estudio. Por ejemplo en la pregunta de ¿Qué factores tiene en cuenta para medir el riesgo de un proyecto de e-learning? aparece como el más importante con el 44% “Conseguir suficientes estudiantes”, en segundo lugar con el 32% los “Organizativos internos (personal, estructura, administración, docentes, tutores) y recién en tercer lugar con el 15% el “Diseño de los contenidos didácticos”.

Con el gran crecimiento del mercado del e-learning en todo el mundo, el incremento de los recursos didácticos disponibles, así como de herramientas para la creación de contenidos y de sistemas para su puesta en línea, ha surgido la necesidad de utilizar contenidos ya existentes independientemente de la plataforma educativa. Por ello la tecnología SCORM, según las características, entornos, especificaciones, estándares, paquetes, modelos, tecnología, comportamiento permite mejorar la portabilidad de los materiales de estudio facilitando la gestión e implementación de proyectos educativos en e-learning. De esta manera, fácilmente se puede llevar cursos completos desde una plataforma educativa a otra o de paquetes de software de diseño multimedial (por ej. Dreamweaver) a alguna plataforma educativa (siempre que éstas trabajen con paquetes SCORM). Así es que en las encuestas realizadas puede observarse por ejemplo en la pregunta ¿Qué importancia tiene la portabilidad de un curso diseñado con componentes multimediales para llevarlo de una plataforma educativa a otra? Las respuestas “Muy alta” con un 29% y “Alta” con un 36% confirman la hipótesis y su utilización en las carreras de grado.

Se ha procurado acentuar, a lo largo de este trabajo, los aspectos más importantes que deben ser considerados

antes de comenzar un proyecto de e-learning, que se resumen a continuación:

- Considerar al e-learning como parte del plan estratégico de la organización.
- Planificar la integración del e-learning con el resto de las actividades de la organización, con fuerte apoyo de la Dirección superior, en todos sus aspectos organizativos, de gestión, etc. No considerar el e-learning simplemente como una ventana al exterior de la empresa, sino como un proceso que afecta a toda la organización educativa.
- Conocer a los destinatarios, y anticiparse a sus necesidades y estilos de aprendizaje.
- Desarrollar el plan de comunicación en Internet y comunicación interna. A través de la comunicación interna se debe lograr un alto grado de conocimiento de todo el personal de la nueva modalidad.
- Estudiar el impacto en la organización y su entorno.
- Diseñar el proceso educativo, secuencia didáctica, actividades didácticas, evaluaciones, materiales de estudio mediatizados y su portabilidad.
- La facilidad de navegación, de uso y administración del sitio web y de la plataforma educativa.

4. FORMACION DE RECURSOS HUMANOS

El equipo de investigación está formado por Docentes y Estudiantes, lo que implica una especial formación en los temas técnicos (principalmente SCORM) y en los aspectos pedagógicos. Además, está prevista la formación del personal Docente de la carrera de grado involucrada en la aplicación del proyecto.

5. BIBLIOGRAFIA

- ADL Initiative. "SCORM 2004 2nd Edition Overview" en ADL (Advanced Distributed Learning).
- Ausubel, D.P. (1973). In defense of advance organizers: A reply to my critics Review of Educational Research.
- Beccacece, P. (2005). "E-learning: la scelta di un Learning Management System open source e la creazione di pacchetti SCORM". Osservatorio e-learning <http://almatwo.ei.unibo.it/wp-content/TESI_LMS_SCORM.
- Carnegie Mellon University. (2005). "SCORM Best Practices Guide for Content Developers" . Carnegie Mellon Learning System Architecture Lab.
- Córca, J. (2005). Diseño de Materiales para Educación a Distancia. Universidad Tecnológica Nacional, Facultad Regional Mendoza, Programa de Educación a Distancia.
- DMR Consulting/SEDISI (2002). Las tecnologías de la sociedad de la información en la empresa española 2001. Dirección Académica: Cátedra UCM -DMR Consulting de Administración de Negocios en Internet, Madrid.

- Duart, J.M. y Lupiáñez, F. (2005). Estrategias en la introducción y uso de las TIC en la universidad. Editorial J.M.
- Fernández Diez de la Lastra, R. (2001). La formación on line y sus mitos. Instituto Universitario Euroforum Escorial.
- García Aretio, L. (2007) Tipos de Ambiente en la EaD. Editorial del BENED.
- García Aretio, L. (2002). ¿Por qué e-learning? Editorial del BENED.
- García Aretio, L. (2001). La educación a distancia, de la teoría a la práctica. Editorial Ariel Educación, Barcelona.
- Griffiths, J.; Blat, J.; García, R. y Sayago, S. (2003). La aportación de IMS Learning Design a la creación de recursos pedagógicos reutilizables. ADL.
- Hernández Aguilar, M. (2005). Características de la Educación a Distancia. Universidad Tecnológica Nacional, Facultad Regional Mendoza, Programa de Educación a Distancia.
- IMS Learnig Resource Metadata Specification (2003). <http://www.imsglobal.org/metadata/index.cfm>
- IMS Learning Design Specification (2003). <http://www.imsglobal.org/learningdesign/index.cfm>
- Lara Navarra, P. y Duart Montoliu, J. (2005). Gestión de contenidos en el e-learning: acceso y uso de objetos de información como *recurso estratégico*. Revista de Universidad y Sociedad del Conocimiento, Vol. 2 - N.º 2. ISSN 1698-580X
- Litwin, E. (2000). *La educación a distancia, temas para el debate de la nueva agenda educativa*, Buenos aires, Amorrortu.
- López Sánchez, J. y Sandulli, F. (2006) G.I.P.T.I.C.-U.C.M. Departamento de Organización de Empresas. Universidad Complutense de Madrid.
- Merlot, M. (2003) Multimedia Educational Resources for Learning and Online Teaching.
- Millward Brown (2005) “Estudio de las necesidades de formación de las organizaciones en España 2005”. Santillana Formación.
- Moreno, F. (2002). Tipos de interacción en el proceso formativo on-line. Bailly-Bailliére.
- Ospina Pineda, D. y Zapata, D. (2004) Diseño de materiales educativos utilizando tecnología de la información y la comunicación. Universidad de Antioquía.
- Ozollo, F. (2007) Elaboración de materiales de aprendizaje). Universidad Nacional de Cuyo.
- Piaget, J.W. (1971). Psychology and Epistemology. Middlesex, England.
- Vazquez, A. (2008). Impacto del e-business en las PyMEs de Mendoza – U.CH.
- Vygotski, L. (1933). Zonas de desarrollo próximo y andamiaje.

Digitalización de Contenidos Accesibles por WEB - DICA

Director:

Lic. Occhipinti Pedro Salvador

Investigadores:

MG. Russo Claudia

Lic. Maria Luciana Balbi

AC Monica Carolina Sarobe

AC Guruceaga Amrcelo Daniel

AC Charne Ajvier Gonzalo

Auxiliar de Investigación:

Lo Monaco Gabriel Adrian

**Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA) ;
Secretaría de Investigación, Desarrollo y Transferencia; Departamento de Informática e
Investigaciones Tecnológicas UNNOBA.**

CONTEXTO

La Universidad Nacional del Noroeste de la Provincia de Buenos Aires, UNNOBBA, tiene sus Sedes en grandes Ciudades como Junin y Pergamino, estas como centro de pueblos aledaños rurales. Si bien Internet y las nuevas tecnologías se están masificando aun hay lugares y sectores donde estas no llegan o no es tan simple el acceso. La radio como medio de comunicación masivo y de alcance seguro, es un medio que por su formato permite la creación de contenido local, llegando no solo con información, sino con instrucción en primera instancia en formato audible. Un una primera etapa el proyecto apunta a la mayor difusión de distinto tipo de información en los sectores mas alejados de las nuevas tecnologías.

En una segunda etapa, se intenta unir el alcance de un medio masivo como la radio a el potencial de las nuevas tecnologías utilizando ambas en pos de una mayor y mejor difusión no solo de contenidos locales, promoción, instrucción sino de información específica para la zona y los distintos sectores que la componen, generando un sistema de información para el desarrollo

RESUMEN

Análisis de la tecnología para digitalización de contenidos, documentos y Podcast en Educación. Análisis de la tecnología para

desarrollo y mantenimiento de interfaces WEB.

Análisis de indexación del contenido digital y gestión de acceso. Análisis de procedimientos específicos. Descripción de resultados esperados. Gestión del ciclo de digitalización.

Desarrollo de la interfaz específico para el soporte y la navegación del material digitalizado. Capacitación a recursos humanos participantes. Implantación de la tecnología y el software. Análisis de resultados

Palabras clave:

Digitalización; Contenidos, Radio; Internet, Posd Cast, Educación, NTIC,s; Rural; Contenido Local

1. INTRODUCCION

La tecnología digital abre una perspectiva totalmente nueva. Internet es el lugar de mercado para la investigación, la enseñanza, la expresión, la publicación y la comunicación de la información.[1]

Las nuevas tecnologías de información y comunicación –NTICs– pueden ser medios efectivos para proporcionar grandes cantidades de información relevante. No obstante, para tener un impacto importante en los programas de desarrollo, los servicios de NTICs deben ser fácilmente

accesibles y también significativos para amplios segmentos de la población rural.[2] ...”El desarrollo de contenidos locales es la condición no negociable y más importante para el uso de las TICs en los procesos de cambio social y en el progreso material de comunidades urbanas o rurales.

El océano de ‘conocimiento’ de la red mundial no corresponde a las necesidades de la mayoría de la población. Cada país es diferente, tiene necesidades diferentes y dentro de cada país –particularmente en el Tercer Mundo – la diversidad de culturas y problemas, demanda enfoques específicos. Necesitamos inventar y multiplicar mini-redes, pequeñas redes (webs) geográficas o redes (webs) comunitarias locales para que la red sea realmente *mundial* y útil para la mayoría de la población en el planeta.

Una vez más, la radio comunitaria nos puede enseñar mucho sobre la pertinencia local. Solamente el desarrollo de contenidos locales puede establecer una diferencia radical entre los telecentros para usos sociales y los cibercafés que sirven a una clientela que ya sabe qué, dónde y cómo buscar la información que necesita. Los cibercafés no necesitan desarrollar un contenido específico porque sus clientes corresponden al usuario típico de Internet en el mundo: varones, de treinta y cinco años de edad, con educación universitaria y un ingreso alto, urbanos, que hablan inglés – es decir, parte de una élite minoritaria para la cual Internet ha sido modelado.[3]

Los cibercafés ofrecen acceso a Internet, pero los telecentros orientados al desarrollo deberían además generar información local y regional y ponerla a disposición de la comunidad. “Un telecentro podría llegar a ser un auxiliar clave para una escuela y una clínica, ofreciendo educación continua para los profesores locales, las enfermeras, las médicas y médicos si los hay.” [4] Para adecuarse a sus usuarios – de nuevo siguiendo el ejemplo de la radio comunitaria – varios proyectos de TICs que operan a partir de la comunidad, producen contenidos locales apropiados a la población específica: campesinos, pescadores y otros grupos que

pocas veces son tomados en cuenta en los cibercafés comerciales. Entre los ejemplos de este enfoque, están los Centros de Conocimiento Comunitario en Chennai, India.” [5]

...”La información que transmiten debe ser adaptada y divulgada en formatos, lenguajes e idiomas que puedan comprender los destinatarios. También debe servir a las necesidades de las personas con relación a la cultura, al contacto humano, y al entretenimiento, necesidades que a pesar de que las sentimos con fuerza todos nosotros, muchas veces son necesidades ignoradas por los profesionales del desarrollo.

La convergencia de las TICs con la radio rural puede servir a estos propósitos, dando un fuerte apoyo para el aprovechamiento y comunicación de conocimientos para el desarrollo, asegurando el acceso más amplio a la información, canalizando y fomentando la expresión cultural y el desarrollo en el ámbito local. Esto es especialmente importante en las áreas rurales, donde la radio es un mecanismo para la divulgación rápida de conocimientos e información, en una diversidad de lenguas y formatos. La radio rural, con su larga historia y su metodología participativa comprobada en el tiempo, es el medio de comunicación popular más ampliamente conocido y usado. El uso combinado de los dos medios, no sólo permite ampliar el acceso a la riqueza de información, sino que ofrece a la vez un mecanismo eficaz para la articulación de las necesidades reales de desarrollo, desde las comunidades.”[6]

La conectividad es sólo la punta del iceberg y bajo ella existen una serie de factores complejos que impiden a la mayoría de la población del mundo apropiarse de Internet. Entre estos factores encontramos los siguientes:

- El analfabetismo – UNESCO estima que hay mil millones de adultos analfabetos en el mundo, aproximadamente 25% de la población total adulta. La mayoría del contenido de la red, especialmente el contenido orientado

al desarrollo, está en lenguaje escrito.

- Idioma – A pesar de que hay más de 6 mil idiomas en el mundo, Internet está dominado por el idioma inglés y por una docena de idiomas que tienen una presencia significativa. Un mínimo de 20% de la población mundial habla idiomas que están casi completamente excluidos de la red. [7]
- El contenido – Usted puede leer en inglés pero ¿Puede encontrar contenidos locales, relevantes o contextualizados?

Más que otro medio, la radio es local. En América Latina por ejemplo, mientras la radio se produce en el ámbito local o nacional, sólo el 30% de la televisión se produce en la misma región, con el 62% producido en los Estados Unidos [3]

Hoy se están explorando modelos alternativos que incluyen telecentros y cibercafé, proyectos de apoyo y software para la traducción de textos tanto escritos como hablados.

Internet es más accesible por estas alternativas. Durante los últimos años una serie de experimentos que combinan la radio local independiente con Internet, han dado vida a nuevos modelos. [8]

Por otro lado la multimodalidad o Interacción multimodal posibilita la extensión de la Web y facilita diferentes modos de interacción conjunta (auditiva, visual).

La interacción multimodal extiende la interfaz de usuario para permitir múltiples modelos de interacción, ofreciendo a los usuarios la posibilidad de usar la voz u otros dispositivos para introducir datos. En respuesta, los usuarios podrán escuchar comandos hablados, audio y, también ver la información gráficamente. Esta capacidad por parte del usuario de especificar el modo o el dispositivo para una interacción particular, en una situación particular, está pensada para mejorar la interfaz del usuario, su accesibilidad y fiabilidad. [9]

La idea es que además de la radio y de los archivos de interés académicos guardados en formatos de mp3 haya apoyatura de videos. Lo que buscamos es que a través de la alfabetización digital podamos masificar el contenido y que puedan acceder fácilmente. Que los contenidos los bajen en sus computadoras o que los carguen en sus mp3 y los escuchen mientras van a caminar, por ejemplo.

2. LINEAS DE INVESTIGACION y DESARROLLO

Area de Redes, Software y Hardware:

- 1) La definición y delimitación de los distintos sistemas involucrados.
- 2) El software a utilizar en algunos sistemas.
- 3) El hardware.

En el punto 1) se logra comprender el funcionamiento de emisoras de radio que tienen similitud a nuestra área. de esta forma se identifican distintos sistemas: *a*).server cast masivo: el cual repite la misma difusión a muchos usuarios (oyentes) *b*). server cast encargado de codificar la fuente de audio en streams, para luego ser transportado al server cast (dedicado), quien los difundirá masivamente por internet, al usuario (oyente) que se conecta mediante el sitio de la radio web. *c*) El software de automatización que gestiona lo referente a singles, la programación de la radio, planificación automática de difusión semanal, o bien, en vivo. Como la edición de segmento de audio como entrevistas, mezclas de sonido, etc. *d*) software adicional para la edición de audio mas elaborada, con efectos, filtros, volumen, fading, etc.

Al hacer el analisis de la funcionalidad de estos sistemas surge la delimitación, la cual acarrea ciertos obstáculos del punto 2). Dicho punto esta relacionado con el operador de la radio que en una primera instancia debería supervisar la mezcla de una difusión en vivo y en otro momento la edición de las distintas fuentes de audio

como eventos, entrevistas, comunicaciones telefónicas, etc. Sin embargo lo que concierne a nuestra área es la elección del software ya que existen sistemas específicos de automatización de radio pero tienen un costo que en esta etapa, esta fuera del presupuesto; por otro lado surge la posibilidad de usar versiones de prueba o con licencia limitada, y también de licencia especiales como en ambientes Linux. Esto ultimo lleva al planteo si el operador tendría que tener cierta capacitación o si se puede adaptar al entorno Linux y a las aplicaciones para procesar el audio y automatizar la radio.

En el punto 3) tenemos un equipo (PC1) encargado de la edición, automatización y la codificación del stream para luego ser transferida al segundo equipo (PC2), optamos este esquema para aislar las diferentes sistemas con respecto a los recursos que consumen y su funcionalidad. El equipo PC1 además debería tener la adaptación y ciertas características en la tarjeta de audio, para acoplarse a la señal proveniente de la mixer que es un hardware que cumple 2 funciones importantes, la primera es adaptar el micrófono a niveles deseados y controlar la señal analógica, mezclando varias fuentes de audio. La segunda es la independencia de PC1, en caso de fallas en el equipo, de este modo, se podría utilizar un tercer equipo provisorio y poder salir al aire en vivo inclusive.

Area de Legislacion, Propiedad Intelectual, y Accesibilidad:

Legislación:

- Leyes que reconocen el derecho de habeas data y protección de datos personales.
- El derecho de autor y copyright constituyen dos concepciones sobre la propiedad literaria y artística
 - Campo de aplicación
 - Tipos de Derecho de Autor
 - Críticas al sistema de Copyright

- Regulación del derecho de autor
- Ley de Propiedad Intelectual. Ley 11.723 del Poder Ejecutivo Nacional, Buenos Aires, 28 de septiembre de 1933.
- Qué se registra?
- Beneficios del registro
- CAPIF - Cámara Argentina de Productores de Fonogramas y Videogramas
- CESSI - Cámara Argentina de Empresas de Software y Servicios Informáticos
- Organización Mundial de la Propiedad Intelectual (OMPI)
 - Propiedad industrial (comprende las invenciones, patentes, marcas, dibujos y modelos industriales e indicaciones geográficas de origen).
 - Derechos de Autor (que comprende las obras literarias y artísticas, tales como las novelas, los poemas, las obras de teatro, las películas, las obras musicales, las obras de arte, los dibujos, pinturas, fotografías, esculturas, y los diseños arquitectónicos.
 - Derechos Conexos: comprende las interpretaciones o ejecuciones de los artistas, la producción de fonogramas y las actividades de los organismos de radiodifusión
- Registro de Marcas
- W3C Accesibilidad WEB . Niveles y aplicabilidad en nuestro proyecto. Pautas de Accesibilidad al Contenido en la Web (WCAG),

3. RESULTADOS OBTENIDOS/ESPERADOS

El objetivo del proyecto es el analisis, preparacion e implantacion de contenido digital y multimedial en esta primera etapa plasmado en la creacion de la radio virtual de la UNNOBA. Accesible por WEB,

Podcast en la educación, aclarando que un podcast es lo mismo que un programa radial, con la única diferencia de que el podcast se graba, edita y produce en un computador, para ser transmitido por medio de Internet.

También analizaremos las posibilidades educativas que la radio virtual, como medio de comunicación, puede ofrecer para el desarrollo de actividades Curriculares, Culturales, de Capacitación, Comunicación y Desarrollo Local.

El proyecto incluye también el desarrollo de un sitio web que permita a los diferentes usuarios acceder a la radio de forma on-line, así también como a contenidos pregrabados, historial de programas, transcripción de contenidos y publicación en formato de video de diferentes materiales y capacitaciones.

Para desarrollar este sitio web se determina armarlo con los lenguajes XHTML y CSS cumpliendo los estándares establecidos por W3C, permitiendo que el sitio pueda ser continuado y mantenido por cualquier colaborador del proyecto que respete los mismos.

4. FORMACION DE RECURSOS HUMANOS

En cuanto a la formación de recursos humanos esta es muy amplia y multidisciplinaria, en tecnologías y metodologías, diversas: NTICs., WEB; Radio, Multimedia, Desarrollo de contenidos; PodCast., Accesibilidad, Estándares Web

5. BIBLIOGRAFIA

[1] Fuente: Estas directrices son el resultado del trabajo de un grupo de expertos de IFLA e ICA (International Council on Archives) invitados a elaborarlas por la UNESCO. http://travesia.mcu.es/documento/s/pautas_digitalizacion.pdf

[2] Ester Zulberti; Jefa, Servicio de Extensión, Educación y Comunicación; Departamento de Desarrollo Sostenible; FAO en Secreto a voces Radio, NTICs e interactividad; Editado por Bruce Girard en colaboración con Grupo de Comunicación para el Desarrollo Servicio de Extensión, Educación y Comunicación Dirección de Investigación, Extensión y Capacitación

[3] PNUD, *Informe sobre Desarrollo Humano*, 1999, P. 34

[4] Scott Robinson, "Rethinking Telecenters: Knowledge Demands, Marginal Markets, Microbanks and Remittance Flows (Repensando los Telecentros: Demandas de conocimientos, mercados marginales, microbancos y flujos de remesas)", en *On the Internet*, Vol. 6, No. 2 (Fall/Winter 2000), una publicación de la Sociedad Internet

www.isoc.org/oti/articles/0401/robinson.html

[5] Secreto a voces Radio, NTICs e interactividad; Editado por Bruce Girard en colaboración con Grupo de Comunicación para el Desarrollo Servicio de Extensión, Educación y Comunicación Dirección de Investigación, Extensión y Capacitación Departamento de Desarrollo Sostenible, Con el apoyo de: Fundación Friedrich Ebert, Dev-Comm (Banco Mundial), Comunica, Cooperación Italiana y CISP; Organización de las Naciones Unidas para la Agricultura y la Alimentación Roma, 2004

[7] Según un estudio publicado por VilaWeb.com en 2000, basado en datos de AllTheWeb, el inglés es el idioma más común, con el 68.4 % de las páginas web, seguido por el japonés, alemán y chino. El francés está en quinto lugar con el 3% y el español en el sexto con 2.5 % cyberatlas.internet.com/big_picture/demographics/article/0,1323,5901_408521,00.html

[8] Muchos de estos experimentos fueron presentados y discutidos en dos seminarios apoyados por la Fundación Friedrich Ebert: en uno se examinaron las experiencias en Asia y en el otro, las experiencias de América Latina y el Caribe. Ver *Converging Responsibility: Broadcasting and the Internet in Developing Countries*, 2000 (Responsabilidad Convergente: la Radiodifusión e Internet en los países en vías de desarrollo) <www.comunica.org/kl/> y *Mixed, Media / Medios Enteros: Radio e Internet en América Latina y el Caribe* 1999 <www.comunica.org/tampa/>.

[9] World Wide Web Consortium (W3C) Oficina Española, Guía Breve de Interacción Multimodal, <http://www.w3c.es/Divulgacion/Guiasbreves/Multimodalidad>

Software Educativo para el aprendizaje de Clasificación de Sistemas de Información

Lic. Ángela Belcastro¹, Lic. Gabriela Oriana¹, Lic. Pamela Ritter¹, Mg. Rodolfo Bertone^{1,2}

⁽¹⁾ Facultad de Ingeniería – UNSJB

⁽²⁾ III LIDI – Instituto de Investigación en Informática - Facultad de Informática – UNLP

(angelab, orianagab, pcritter)@ing.unp.edu.ar, pbertone@lidi.info.unlp.edu.ar

1. CONTEXTO

Esta línea de Investigación forma parte del Proyecto UNPSJB - UNLP. PROYECTO DE INVESTIGACIÓN: “Casos de estudio de sistemas. Software educativo de Sistemas de Información”. Informática. Análisis de sistemas.

2. RESUMEN

Este trabajo está orientado a examinar algunos elementos provenientes del campo de Tecnología Informática aplicada en Educación, que se utilizan en los software educativos con la intención de favorecer la retención y construcción de conocimiento, y pueden ser utilizados para elaborar propuestas, estrategias de aprendizaje y ejercitación, estableciendo mejoras en el proceso de enseñanza-aprendizaje de temas puntuales del análisis de sistema.

Describe elementos de un software educativo empleado e implementado, con el objeto de favorecer el aprendizaje de los estudiantes, brindando apoyo en las clases teórico-prácticas de la asignatura Introducción al Análisis de Sistemas, sede Comodoro Rivadavia de la UNPJB. Los desarrollos se encuentran dentro del marco del proyecto titulado “Casos de estudio de sistemas. Software educativo de clasificación de Sistemas

de Información”, cuya meta final principal es construir un software que tenga por objeto mejorar la calidad educativa a nivel Universitario y Polimodal en los temas relacionados con la definición, administración y clasificación de SI.

Palabras Claves

Tecnología informática aplicada en educación. Software Educativo. Análisis de sistemas. Sistemas de información. Introducción al desarrollo de sistemas de información.

3. INTRODUCCION

Los ejes centrales del trabajo se basan en la tecnología educativa y el aprendizaje los Sistemas de Información y aspectos introductorios de su desarrollo.

Tecnología educativa: CABERO (1999) señala que la Tecnología Educativa es un término **integrador** (en tanto que ha integrado diversas ciencias, tecnologías y técnicas: física, ingeniería, pedagogía, psicología...), **vivo** (por todas las transformaciones que ha sufrido originadas tanto por los cambios del contexto educativo como por los de las ciencias básicas que la sustentan), **polisémico** (a lo largo de su historia ha ido acogiendo diversos significados) y también **contradictorio** (provoca tanto defensas radicales como oposiciones frontales). Posiblemente la definición que recoge mejor estas distintas tendencias es la que propuso la UNESCO en 1984, a partir de las propuestas

hechas en 1970 por la Commission on Instructional Technology, formulando una doble acepción de Tecnología Educativa:

1.- Originalmente ha sido concebida como el uso para fines educativos de los medios nacidos de la revolución de las comunicaciones, como los medios audiovisuales, televisión, ordenadores y otros tipos de "hardware" y "software"

2.- En un nuevo y más amplio sentido, se entiende como "el modo sistemático de concebir, aplicar y evaluar el conjunto de procesos de enseñanza y aprendizaje teniendo en cuenta a la vez los recursos técnicos y humanos y las interacciones entre ellos, como forma de obtener una más efectiva educación". (UNESCO, 1984, 43-44)

INTRODUCCIÓN A LOS SISTEMAS DE INFORMACIÓN Y SU DESARROLLO:

El empleo de un sistema de información de procesamiento de transacciones puede permitirle a la organización lograr una ventaja competitiva, proporcionando un beneficio importante a largo plazo, ya que puede aumentar la lealtad del cliente, proporcionar un servicio de alta calidad, aumentar la calidad, productividad y satisfacción del cliente, lograr efectividad y eficiencia en la realización de las operaciones y reducir drásticamente los costos. Esta clase de sistema de información (SI) requiere de una gran cantidad de datos de entrada y producen un número elevado de salidas sin requerir de un procesamiento sofisticado o complejo.

Se emplea en el desarrollo rutinario de las actividades operativas, y se utilizan como la base de otros tipos de SI, el SI para la administración (MIS, Management Information System), el sistemas de información de apoyo a la toma de decisiones y otros sistemas, como los de inteligencia artificial o

sistemas expertos (AI/ES, Artificial Intelligence/Expert Systems).

Los gerentes del nivel medio, necesitan contar con información en tiempo real, para realizar un control adecuado, ellos utilizan los MIS. Un función importante de un MIS es mejorar la efectividad al proporcionar la información correcta a la persona adecuada, en la forma y el momento convenientes. Los datos que se introducen en un MIS se originan tanto en fuentes internas como en fuentes externas. Las fuentes externas de datos pueden incluir clientes, proveedores, competidores y accionistas cuyos datos no han sido capturados por los sistemas de información de procesamiento de transacciones TPS. Los MIS requieren información actual sobre el desempeño medido contra los estándares establecidos en la organización. Para lograr su objetivo, necesitan disponer de información histórica, y depende considerablemente de ella.

Es vital para un alumno que se está formando en el área de SI no solo comprender la teoría general de sistemas, y el contexto en el que se desarrollan los SI, aspecto que en gran medida se logra gracias a la utilización de EduIAS, (www.ing.unp.edu.ar/cesbor), herramienta resultante del proyecto anterior del grupo que se encuentra en una faz netamente operativa; sino también advertir cual es el esquema básico, sus componentes y características, e identificar los diferentes tipos de sistemas de información que coexisten dentro de una empresa, su evolución, resaltando la necesidad de administrar el conocimiento de manera que sea posible a futuro su aplicación.

Los diagramas de flujos de datos DFD se emplean al desarrollar un SI. El DFD de nivel cero es el diagrama de contexto DC. Ubica al sistema dentro de un contexto o entorno.

4. LINEAS DE INVESTIGACION Y DESARROLLO

Este proyecto resulta en una continuación del proyecto denominado "Casos de estudio de sistemas, basados en organizaciones reales.", tiene por objetivo ampliar las capacidades de la herramienta EduIAS, permitiendo clasificar los

sistemas de información (SI) que coexisten dentro de una organización, y destacar en cada tipo, el esquema básico, sus componentes, características, objetivos y evolución, administrando el conocimiento de manera que sea posible a futuro su aplicación. La meta final es construir un software que tenga por objeto mejorar la calidad educativa a nivel Universitario y Polimodal en los temas relacionados con la definición, administración y clasificación de SI.

5. RESULTADOS OBTENIDOS/ ESPERADOS

Consideraciones Iniciales:

- Análisis del tema: "diseño de aplicaciones educativas y características de este tipo de sistemas", con el objeto de contar con elementos suficientes para considerar contenidos y aspectos esenciales al desarrollar cada modulo componente del software educativo.
- Aplicación de contenidos de Psicología Cognitiva, a materiales didácticos propuestos como herramientas de apoyo al aprendizaje que se incluirán en módulos del software educativo. Se examinaron algunos elementos provenientes del campo de Psicología Cognitiva, que favorecen la retención y construcción de conocimiento, y pueden ser utilizados para elaborar propuestas, estrategias de aprendizaje y ejercitación, estableciendo mejoras en el proceso de enseñanza-aprendizaje de temas puntuales del análisis de sistema.

Aspectos del diseño y Producción de materiales: Los medios empleados para proporcionar los materiales, serán impreso, digital y multimedia educativo. Las estrategias multimediales utilizadas son:

- Único (el mismo material para todos los destinatarios, en lo relativo al material impreso).
- Opcional (varios materiales en diferente soporte -impreso y digital por ejemplo - con la misma información para que los destinatarios opten por uno u otro).
- Complementario (varios materiales en diferente soporte con información complementaria, en donde un medio remite al otro y contienen información diferente).

ASPECTOS DE LA METODOLOGÍA DE ENSEÑANZA:

Uno de los elementos en consideración dentro de la metodología de enseñanza, son las actividades especiales, que incluyen:

- AE 1. Clases con el empleo de casos de aplicación con modalidad en equipos.
- AE 2. Experiencias en actividades especiales con el apoyo de módulos de software educativos, y la realización de clases colaborativas.
- AE 3. Aplicación de la técnica de jerarquización para llevar al alumno al aprendizaje por descubrimiento, al iniciarse en el tema Pert, y clases especiales con empleo de la herramienta "Módulos de entrenamiento en la confección y comprensión diagramas de contexto y de otros modelos del analista".
- AE 4. Ejercitación solicitada a los alumnos de promoción, que es de caracter individual, con el empleo de mapas conceptuales y el desarrollo de actividades de investigación en Internet, de entrega en formato digital.

ALGUNOS COMPONENTES DEL SOFTWARE EDUCATIVO

- a) **Enunciado de la actividad: "los procesos y la transformación de datos en información"**. Enunciado en pdf, y video con material multimedia necesario para desarrollar el trabajo. El estudiante realiza actividades en equipo, interpreta información que no está dada

de forma textual, sino a través de un archivo multimedia.

b) Herramienta: “módulos de entrenamiento en la confección y comprensión de diagramas de contexto y otros modelos del analista”. Integrado por tres módulos, compuesto por 72 diapositivas, con ayuda contextual incorporada, y feedback en ejercitación de comprensión propuesta. Es la herramienta con la que los estudiantes se inician en la confección de diagramas de contexto, y lleva al alumno a hacer conexiones conceptuales tanto con temas ya observados de la materia, como con aspectos asociados a Sistemas de Información y al desarrollo de Sistemas de Información. Tiene control de acceso a resoluciones, que requieren de claves especiales.

c) Enunciado de la actividad: “temas de interés del campo de la información y de los sistemas de información en las organizaciones”. Estrategia medial (Opcional). En pdf, e impreso. Motiva al estudiante a desarrollar un proceso de síntesis de un tema de interés que es contenido del curso, y a investigar acerca del mismo. Permite a los docentes conocer el perfil y los intereses de los cursantes, como conocer de qué forma, el alumno, interpreta el tema.

d) Herramienta: “Diagrama de Contexto”. Permite al alumno, partiendo de la descripción escrita del proceso que se lleva a cabo en una empresa, practicar la elaboración de diagramas de contexto, utiliza elementos interactivos. Uno de sus

componentes, proporciona diferentes enunciados que describen subsistemas, y luego de manera interactiva permite al estudiante, asociar el nombre del proceso a cuatro diagramas de contexto que aparecen en pantalla.

e) Ejercicios Interactivos Del Módulo 1. Estrategia multimedial (Único). con feed back, el estudiante recibe su puntaje a medida que lo desarrolla. Proporcionar al alumno ejercitación que le ayude a corroborar de forma inmediata los resultados alcanzados al resolver problemas de comprensión e integración de procesos de la organización.

6. FORMACION DE RECURSOS HUMANOS

El equipo de proyecto ha desarrollado experiencias educativas aplicando los módulos destacados, y otras herramientas de apoyo al aprendizaje que han surgido en el marco del proyecto, con grupos de alumnos de primer año, segundo cuatrimestre de las carreras Analista Programador Universitario y Licenciatura en Informática.

Algunas herramientas que surgieron de los esfuerzos del equipo de proyecto y están disponibles para los usuarios de Internet, se encuentran en www.ing.unp.edu.ar/cesbor, y en www.ing.unp.edu.ar/asignaturas/ias accediendo por icono del proyecto de cátedra 2.

En la UNLP, dos integrantes del equipo están actualmente cursando el Magister en Tecnología Informática aplicada en Educación, han examinado, entre otros temas, los distintos tipos de inteligencia, estudios aplicados en el proyecto, reflejados en el material de participación en Wicc 2008.

7. BIBLIOGRAFIA

- [1]- Ralph M. Stair. George W. Reynolds. Principios de Sistemas de Información. Cuarta edición. Ciencias Thomson. 2000.
- [2]- J. Senn. Análisis y diseño de sistemas de Información. McGrawHill. 1992

- [3] - Davis, William. Herramientas CASE: metodología estructurada para el desarrollo de sistemas. PARANINFO. 1992
- [4] - Cohen. Sistemas de información para la toma de decisiones. McGrawHill. McGrawHill.
- [5] - Yourdon. Análisis estructurado moderno. Prentice Hall. 1993.
- [6] - A. Belcastro. Tomo de Teoría de Introducción al Análisis de Sistemas. Capítulos 3, 4, 5 y 6. [www.ing.unp.edu.ar/asignaturas/ias](http://www.ing.unp.edu.ar/asignaturas/).
- [7] - Raymond McLeod, Jr. Sistema de información gerencial. Editorial Pearson Educación. 2000.
- [8] - Effy Oz. Administración de Sistema de información. Segunda Edición. Thomson Learning. 2001.
- [9] - Belcastro, Oriana, Morgante, De la Paz, Ritter, Bertone. EduIAS, una herramienta educativa de apoyo al proceso de enseñanza aprendizaje del análisis de sistemas.
- [10] - Kendall & Kendall. Análisis y Diseño de Sistemas. Sexta Edición. Pearson. Prentice Hall. 2005.
- [11] - Mario G. Piatini. José A Calvo Manzano. Joaquín Cervera. Luis Fernández. Aplicaciones Informáticas de Gestión. Una perspectiva de Ingeniería de Software. Alfaomega. Ra-Ma. 2004.
- [12] - Tecnología Educativa. <http://www.pangea.org/peremarques/tec.htm>
- [13] - Jorge Roberto Volpentesta. Estudios de sistemas de información para la administración. Bs. As. Librería y Editorial. 1992.
- [14] - Gordon B. Davis. Margrethe H. Olson. Sistemas de información gerencial. McGrawHill. 1994.
- [15] - Tecnologías de la Información y la Comunicación. Segunda edición. Alicia Cortagerena. Claudio Freijedo. Pearson. Prentice Hall. 2006.
- [16] - Sistemas de Telecomunicación e informáticos. Gestión del desarrollo de Sistemas de Telecomunicación e Informáticos. Ramón Ramírez Luz. THOMSON. Paraninfo. 2005.
- [17] - Chi, M y Glaser, R. (1986). Capacidad de resolución de problemas. En: Sternberg, R. B Las capacidades humanas. Un enfoque desde el procesamiento de la información. Barcelona: Labor Universitaria.
- [18] - R. (1997). Las tres claves de la inteligencia exitosa. En: Sternberg, R. Inteligencia exitosa. Como una inteligencia práctica y creativa determina el éxito en la vida. Buenos Aires: Paidós.
- [19] - Sternberg, R. (1985). “La Teoría Triárquica de la Inteligencia: comprender el autogobierno mental”. (En: Pueyo, A. (1996). Los componentes cognitivos de la inteligencia. Cáp. 5. Barcelona: Paidós. pp97 a 124 – (*))
- [20] - Psicología de la Educación . John W. Santrock. Mc Graw Hill. Santrock. 2006.
- [21] - Psicología cognitiva y de la instrucción. Roger. H. Bruning. Pearson. Prentice Hall. 2005.
- [22] - Tecnología de gestión. Segunda Edición. Alicia Cortagerena. Claudio Freijedo. Pearson. Prentice Hall. 2006.