ACTAS

XXVII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACION

CACIC 2021

4 al 8 de Octubre
CONGRESO VIRTUAL

RedUNCI
CACIC 2021
Universidad Nacional de Salta

Universidad Nacional de Salta
Acta Memorias del Congreso Argentino en Ciencias de la Computación -CACIC / compilación de Marcia I. Mac Gaul. - 1a ed. - Salta : Universidad Nacional de Salta, 2021.
Libro digital, Otros

Archivo Digital: descarga y online
ISBN 978 -987-633-574-4

1. Computación. 2. Tecnología Informática. 3. Ciencias de la Información. I. Mac Gaul, Marcia I., comp. II. Título.
CDD 004.071

Universidad Nacional de Salta
Acta Memorias del Congreso Argentino en Ciencias de la Computación-CACIC / compilación de Marcia I. Mac Gaul. - 1a ed. - Salta : Universidad Nacional de Salta, 2021.
Libro digital, Otros

Archivo Digital: descarga y online
ISBN 978-987-633-574-4

1. Computación. 2. Tecnología Informática. 3. Ciencias de la Información. I. Mac Gaul, Marcia I., comp. II. Título.
CDD 004.071

ISBN 978-987-633-574-4

INDICE

ISBN 978-987-633-574-4

ISBN 978-987-633-574-4

ISBN 978 -987-633-574-4

AUTORIDADES UNSa

Rector – Cr. Víctor Hugo CLAROS

Vicerrectora – Dra. Graciela del Valle MORALES

Decano Facultad de Ciencias Exactas – Ing. Daniel Hoyos

Vicedecano Facultad de Ciencias Exactas – Mg. Gustavo Daniel GIL

Secretaria Académica y de Investigación – Dra. María Rita Martearena

Secretario de Extensión y Bienestar – Ing. Walter Alberto Garzón

ISBN 978 -987-633-574-4

AUTORIDADES Red UNCI

COORDINADOR TITULAR

Pesado Patricia (UNLP) 2020-2022

COORDINADOR ALTERNO

Estayno Marcelo (UNLZ) 2020-2022

JUNTA DIRECTIVA

Kuna Horacio (UNaM) 2020-2022

Printista Marcela (UNSL) 2020-2022

Tugnarelli Mónica (UNER) 2020-2022

Eterovic Jorge (UNLaM) 2019-2021

Aciti Claudio (UNCPBA) 2019-2021

Arroyo Marcelo (UNRC) 2019-2021

Panizzi Marisa (UK) 2019-2021

MIEMBRO HONORARIO

De Giusti Armando (UNLP)

SECRETARÍAS

Secretaría Administrativa: Lasso Marta

Secretaría Académica: Russo Claudia

Secretaría de Ciencia y Técnica: Rodríguez Nelson

Secretaría de Asuntos Reglamentarios: De Vincenzi Marcelo

Secretaría de Vinculación Tecnológica y Profesional: Gil Gustavo

Secretaría de Congresos, Publicaciones y Difusión: Thomas Pablo

ISBN 978 -987-633-574-4

COMITÉ ORGANIZADOR

David Aguilera

Rodolfo Baspineiro

Gustavo Daniel Gil

Loraine Gimson

Claudia Ibarra

Marcia Mac Gaul

Rosa Macaione

Andrea Murillo

María Laura Palermo

Ernesto Sánchez

Jorge Silvera

ESCUELA INTERNACIONAL DE INFORMATICA

Directora: Mag. Loraine Gimson

ISBN 978 -987-633-574-4

COMITÉ ACADÉMICO

ISBN 978 -987-633-574-4

Wálter Panessi (UN Lujan)
Maria Valeria Poliche (UN Catamarca)
Eduardo Campazzo (UN La Rioja)
Alejandro Oliveros (UN Tres de Febrero)
Griselda María Luccioni (UN Tucumán)
Martín Morales (UNAJ)
Patricia Zachman (UN Chaco Austral)
Antonio Foti (UN del Oeste)
Carlos García Garino (UN de Cuyo)
Julio Cesar Doumecq (UN de Mar del Plata)
Marcelo De Vincenzi (UAI)
Alberto Guerci (UB)
Marisa Panizzi (U kennedy)
Juan Bournissen (U Adventista del Plata)
Jorge Finocchieto (UCAECE)
Adriana Alvarez (UP)
Sebastián Grieco (UCA Rosario)
Marcelo Zanitti (U Salvador)
Rosa Giménez (U Aconcagua)
Carlos Beyersdorf (U Gastón Dachary)
Ariadna Guglianone (UCEMA)
Juan Pablo Cosentino (U Austral)
Liliana Rathmann (U Atlántida Argentina)
Rodolfo Bertone (UCA La Plata)
Alicia Mon (ITBA)
Fernando Pinciroli (U Champagnat)

ISBN 978 -987-633-574-4

COMITÉ CIENTÍFICO

Claudio Aciti (Argentina)
María José Abásolo (Argentina)
Hugo Alfonso (Argentina)
Jorge Ardenghi (Argentina)
Marcelo Arroyo (Argentina)
Hernán Astudillo (Chile)
Sandra Baldasarri (España)
Javier Balladini (Argentina)
Luis Barbosa (Portugal)
Rodolfo Bertone (Argentina)
Oscar Bría (Argentina)
Nieves Brisaboa (España)
Carlos Buckle (Argentina)
Alberto Cañas (EE.UU)
Ana Casali (Argentina)
Silvia Castro (Argentina)
Alejandra Cechich (Argentina)
Edgar Chavez (México)
Carlos Coello Coello (México)
Uriel Cuckierman (Argentina)
Armando De Giusti (Argentina)
Laura De Giusti (Argentina)
Marcelo De Vincenzi (Argentina)
Claudia Deco (Argentina)
Beatriz Depetris (Argentina)
Javier Diaz (Argentina)
Juerguen Dix (Alemania)
Ramón Doallo (España)
Domingo Docampo (España)
Dujmovic Jozo (USA)
Marcelo Estayno (Argentina)
Elsa Estevez (Argentina)
Jorge Eterovic (Argentina)
Marcelo Falappa (Argentina)
Pablo Fillottrani (Argentina)
Jorge Finocchieto (Argentina)
Frati Emmanuel (Argentina)
Daniel Fridlender (Argentina)
Carlos García Garino (Argentina)
Javier García Villalba (España)

ISBN 978 -987-633-574-4

Marcela Género (España)
Sergio Gomez (Argentina)
Eduard Gröller (Austria)
Roberto Guerrero (Argentina)
Jorge Ierache (Argentina)
Tomasz Janowski (Naciones Unidas)
Kuna Horacio (Argentina)
Laura Lanzarini (Argentina)
Guillermo Leguizamón (Argentina)
Fernando Lopez Gil (España)
Ronald Prescott Loui (EEUU)
Emilio Luque (España)
Cristina Madoz (Argentina)
Alejandra Malberti (Argentina)
Cristina Manresa Yee (España)
Marco Javier (España)
Mauricio Marín (Chile)
Ramón Mas Sansó (España)
Orlando Micolini (Argentina)
Alicia Mon (Argentina)
Regina Motz (Uruguay)
Marcelo Naiouf (Argentina)
Antonio Navarro Martín (España)
José Angel Olivas Varela (España)
Ariel Pasini (Argentina)
Patricia Pesado (Argentina)
María Fabiana Piccoli (Argentina)
Marcela Printista (Argentina)
Álvaro Pardo (Uruguay)
Mario Piattini (España)
Enrico Puppo (Italia)
Hugo Ramón (Argentina)
Dolores Rexachs (España)
Nora Reyes (Argentina)
Rosabel Roig Vila (España)
Gustavo Rossi (Argentina)
Paolo Rosso (España)
Sonia Rueda (Argentina)
Francisco Ruiz (España)
Claudia Russo (Argentina)

ISBN 978 -987-633-574-4

Carolina Salto (Argentina)
Cecilia Sanz (Argentina)
Guillermo Simari (Argentina)
Osvaldo Spositto (Argentina)
Ralf Steinmetz (Alemania)
Remo Suppi (España)
Liane Tarouco (Brasil)
Francisco Tirado (España)
Luiz Velho (Brasil)
Marcelo Vénere (Argentina)
Eduardo Vendrell (España)
Horacio Villagarcía (Argentina)
Dante Zanarini (Argentina)

CACIC 2021

WORKSHOP

RedUNCI

CACIC 2021

Universidad Nacional de Salta

CACIC 2021

WORKSHOP AGENTES Y SISTEMAS INTELIGENTES

COORDINADORES

Guillermo Leguizamón (UNSL)
Carolina Salto (UNLPam)
Daniel Fridlender (UNC)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978 -987-633-574-4

Estimación de Temperatura en Servidores mediante Herramientas de Deep Learning

Federico G. D'Angiolo, Ignacio Mas, Juan Ignacio Giribet

Universidad Nacional de Avellaneda, Buenos Aires, Argentina.
Instituto Tecnológico de Buenos Aires (ITBA) y CONICET, Buenos Aires, Argentina.
Universidad de Buenos Aires e Instituto Argentino de Matemática "Alberto Calderón" (IAM) CONICET, Buenos Aires, Argentina.
fdangiolo@undav.edu.ar
imas@itba.edu.ar
jgiribet@fi.uba.ar

Resumen En este trabajo se propone el estudio de estimación de temperatura sobre un servidor, con el objetivo de poder predecir el funcionamiento del mismo bajo condiciones ambientales controladas. Para esto se propone la utilización de herramientas de Deep Learning como por ejemplo, MLP (Multi Layer Perceptron) y LSTM (Long Short-Term Memory). La utilización de éstas persigue el objetivo de poder compararlas y sacar conclusiones sobre su funcionamiento en el ámbito de un datacenter, donde se encuentra el servidor bajo estudio.

Palabras claves: Deep Learning, Redes Neuronales, MLP, LSTM, Datacenter, Temperatura.

1 Introducción

Actualmente, debido al gran avance en el desarrollo de sensores para el muestreo de datos, resulta importante analizar cómo se puede integrar esta tecnología con algoritmos de Inteligencia Artificial (IA). En el caso particular de los Datacenters, es una necesidad controlar la temperatura para que los equipos que se encuentran dentro, puedan trabajar correctamente. Por esta razón, resulta conveniente tener una estimación de la temperatura a la que se encuentra cada servidor, de manera que luego se pueda modificar la ventilación y obtener resultados que permitan cierta estabilidad en las condiciones climáticas del recinto.

En este trabajo se propone usar herramientas de Machine Learning (ML) para predecir la temperatura a la que se encuentra un servidor en particular dentro de un Datacenter. Para poder llevar a cabo este trabajo, se propone sensar la temperatura de un servidor y observar su evolución en función del tiempo para luego, mediante herramientas como MLP (Multi Layer Perceptron) y LSTM (Long Short Term Memory), predecir cómo evoluciona dicha variable un instante después, es decir, en t+1 (una unidad de tiempo después). Este análisis permite pensar en distintas acciones de ventilación sobre el ambiente dado que se puede conocer cómo varía la temperatura una unidad de tiempo

ISBN 978 -987-633-574-4

posterior, la cual, en este trabajo, es de 15 segundos. Las estimaciones obtenidas con cada una de estas redes neuronales se pondrán en comparación para obtener conclusiones sobre su funcionamiento y observar así, cuál de ellas resulta más adecuada para el caso.

La motivación de este trabajo radica en la proliferación de algoritmos de ML relacionados con el procesamiento de series de tiempo ya que permiten obtener resultados aproximados en tiempos relativamente cortos. En particular, en este trabajo, se describe el procedimiento de sensado de temperatura sobre un servidor, el cual va a conformar un dataset o conjunto de datos, que tendrá la forma de una serie temporal pues cada valor de temperatura se encuentra en función del tiempo. Este sensado tiene en cuenta herramientas de IoT (Internet of Things), para el envío de datos a un servidor remoto, el cual almacena la información. Luego, con estos datos, se procede a trabajar con las redes neuronales propuestas (MLP y LSTM) para lograr la estimación.

La distribución de este trabajo se describe de la siguiente forma: en la sección 2 se realiza una introducción a la estimación de temperatura mediante redes neuronales. Luego, en la sección 3 se describe la aplicación al caso de estudio donde se comenta cómo se toman los datos de temperatura del servidor para luego, en la sección 4 mostrar los resultados y concluir con la sección 5 donde se dan las conclusiones.

1.1 Trabajos relacionados

Actualmente la investigación en estimación de variables mediante herramientas de ML se encuentra en gran consideración. Por ejemplo, existen trabajos donde se describe el funcionamiento de una red neuronal para estimar la temperatura en ambientes, es decir, mediante el aprendizaje de esta red se puede conocer cómo puede evolucionar la temperatura dentro de un recinto[1]. Siguiendo en este misma línea, se puede citar trabajos similares donde además de usar redes reuronales para la estimación de temperatura, se añaden algoritmos genéticos [2]. Dentro de esta literatura se pueden encontrar trabajos donde se estudian distintos modelos de ML para la estimación de temperatura, teniendo en cuenta la precisión con que se realiza la misma [3]. Por su parte, si bien obtener predicciones sobre la temperatura mediante herramientas de ML resulta importante, en algunos casos en base a la predicción lograda, es vital tomar decisiones que permitan reducir el consumo de potencia de equipos que generan elevaciones térmicas [4]. A su vez, en lo respectivo a la utilización y comparación de MLP y LSTM, podemos observar trabajos donde se usan estas redes para estimar temperaturas dentro de edificios y conocer cuál de estas tiene mejor rendimiento [5], [6]. Siguiendo con esta comparación, existen trabajos donde se estudia cómo varía el flujo de caja utilizando MLP y LSTM, observando sobre todo los métodos de comparación [7]. Dado que muchos de estos estudios se basan en el análisis de series de tiempo, es importante estudiar cómo se pueden utilizar las LSTM y observar su beneficio [8].

ISBN 978 -987-633-574-4

2 Estimación de temperatura mediante Redes Neuronales.

Los datos de temperatura obtenidos tienen además la información del momento de su captura, es decir, se conoce el día, hora, minutos y segundos de cada valor sensado. Concentrando estos datos en un dataset se puede obtener lo que se llama una Serie Temporal o Serie de Tiempo la cual nos brinda información sobre la evolución de la temperatura a lo largo de un período de tiempo.

Para procesar esta información y poder lograr estimaciones, se utilizan dos redes neuronales con el fin de obtener una comparación y estudiar cuál de ellas resulta mejor en cuanto al objetivo de predicción. Estas redes neuronales son: MLP (Multi Layer Perceptron) y LSTM (Long Short-Term Memory), las cuales se comentan a continuación.

2.1 Multi Layer Perceptron

Un Perceptrón Multi Capa (MLP, Multi Layer Perceptron), resulta ser una Red compuesta por perceptrones conectados entre sí. Estas conexiones forman tres capas, denominadas: capa de entrada, capa oculta (ésta puede estar conformada por varias capas) y capa de salida, las cuales se pueden visualizar en la Fig. 1

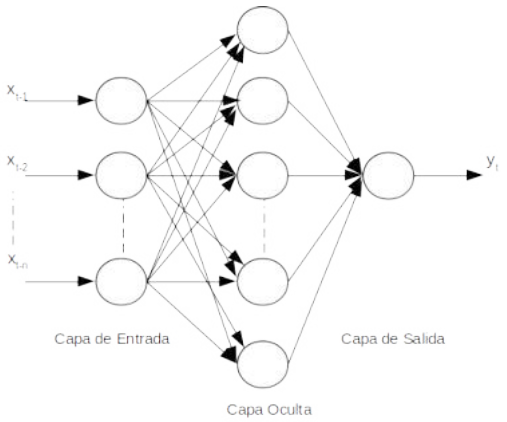


Fig. 1: Perceptrón Multi Capa

La capa de entrada toma como vector de entrada a las muestras de la serie temporal, es decir, si tenemos en cuenta que a la muestra de temperatura actual la denotamos como x_t, a las muestras anteriores las podemos llamar: $[x_{t-1}, x_{t-2}, ..., x_{t-n}]$. Con esta información, la MLP realizará el procesamiento para obtener en su salida, la estimación. Cabe aclarar que la salida de cada perceptrón se puede obtener mediante:

ISBN 978 -987-633-574-4

$$g_i = h(W_i.x + b_i) \tag{1}$$

Siendo:
h = Función de Activación
W_i = Pesos
x = Entradas
b_i = Bias

En el caso de generar una sola estimación, como se describe en este trabajo donde se busca estimar el valor de la temperatura en t+1, la capa de salida consta de un solo perceptrón, cuya salida es $\hat{y}$.

2.2 LSTM

En el caso aquí tratado sobre series de tiempo, cada uno de los valores de esta serie se encuentra relacionado con el anterior, por esta razón, sería conveniente trabajar con redes neuronales que tengan "memoria", es decir, que puedan tener en cuenta el resultado de un estado anterior para luego procesar. Esto resulta importante cuando se trata de sobre estimaciones pues el resultado a predecir está ligado a los valores anteriores. Este tipo de redes se denomina Redes Neuronales Recurrentes (RNN) y se muestran en la siguiente imagen:

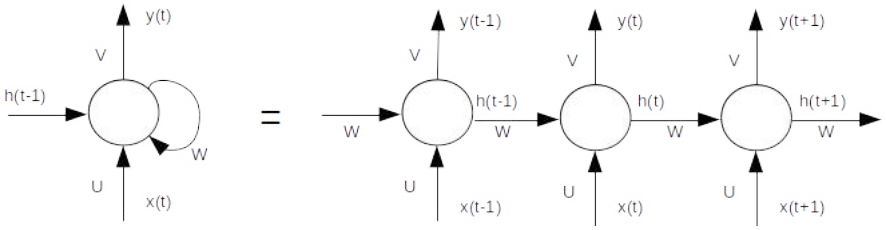


Fig. 2: Red Neuronal Recurrente

En la Fig.2 se observa la entrada a la Red como $x(t)$, la salida mediante $y(t)$ y el estado anterior $h(t-1)$. Cada uno de los nodos que conforma la red, recibe como entrada a cada una de las muestras de x(t).

En base a este sistema, se puede definir entonces el procesamiento, mediante:

$$h_t = \Phi(h_{t-1}, x_t) \tag{2}$$

Siendo:
Φ = Función de Activación.
h_{t-1} = Estado anterior.
x_t = Entradas.

ISBN 978 -987-633-574-4

El cálculo para el entrenamiento de los pesos de este tipo de red neuronal suele ser algo dificultoso, sobre todo cuando las series de tiempo tienen una longitud considerable. Por esta razón se suelen usar las LSTM (Long Short Term Memory), las cuales tienen un funcionamiento similar a las RNN pero su arquitectura se encuentra optimizada para poder solventar el problema de entrenamiento. Esto proporciona cierta robustez frente a las RNN, razón por la cual, suelen usarse este tipo de Redes frente a las Recurrentes, para ciertos problemas.

3 Aplicación al caso de estudio

Previamente al estudio en cuanto a la estimación de temperaturas sobre los servidores, resulta conveniente describir cómo se realiza el sensado de temperatura sobre dicho servidor. Para esto, en la Fig. 3, se muestra una vista superior del Datacenter y cómo se distribuyen los distintos servidores.

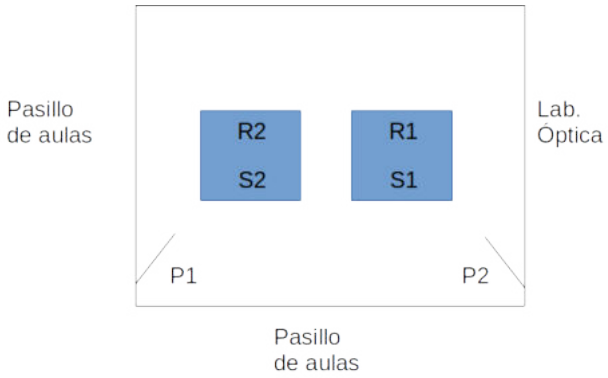


Fig. 3: Layout del Datacenter

La distribución espacial que se propone en la Fig. 3, tiene en cuenta a los Racks denotados mediante R_1 y R_2 los cuales alojan un servidor cada uno (S_1 y S_2). Para sensar la temperatura sobre cada servidor, se propone un sistema basado en un módulo WiFi ESP8266 y sensores de temperatura. Este sistema se encarga de tomar la información de temperatura a la que se encuentra el servidor bajo estudio (S_1) y la envía, mediante WiFi, a un servidor externo el cual almacena todos los datos. Como resguardo, los datos de temperatura sensados se envían a los propios servidores del Datacenter (S_1 y S_2).

En esta etapa del estudio, resulta importante tener en cuenta los datos tomados dado que el muestreo no tiene un efecto continuo en todo momento, es decir, muchas veces se producen caídas en la red de WiFi o cortes de luz momentáneos los cuales producen un reseteo del módulo. Al producirse esto último, los sensores emiten valores que no se condicen con el comportamiento que tiene el servidor hasta el momento. Este análisis es muy importante dado que con estos

ISBN 978 -987-633-574-4

datos se conforma el dataset o conjunto de datos el cual resulta ser la entrada a las redes reuronales que se estudian, si este dataset no se encuentra revisado adecuadamente, las redes podrían estimar de forma incorrecta.

3.1 Datos obtenidos

En base a los datos de temperatura tomados de los sensores, se puede obtener el gráfico de la Fig.4, que muestra cómo evoluciona la temperatura del servidor S_1, en función del tiempo.

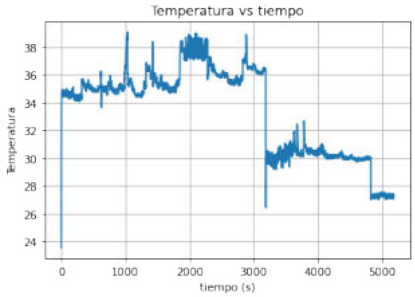


Fig. 4: Temperatura de cada servidor

En la Fig.4 se observan las variaciones de temperatura en ciertos rangos, lo cual demuestra los períodos de mayor y menor actividad del servidor. Por esta razón, resulta interesante utilizar herramientas como redes neuronales para estimar estas variaciones y tener un mejor cuidado de los servidores.

Por último, en base a la Fig.4, cabe aclarar que el muestreo de datos se realiza cada 15 segundos dadas las limitaciones del servidor donde se almacenan los datos.

4 Resultados

En esta sección se muestran los distintos resultados obtenidos a partir del entrenamiento y predicción de cada una de las redes neuronales propuestas. Los resultados obtenidos parten de experimentos realizados sobre la cantidad de muestras pasadas. Con estas y el procesamiento respectivo de cada red, se obtiene la estimación en el momento t+1. Para estos ensayos, se propone entonces configurar a las redes con cantidades similares de parámetros entrenables, iterando con distintas cantidades de muestras pasadas como por ejemplo, 5, 10, 20, 30, etc. Luego, tomando una métrica denominada Error Medio Absoluto, (MAE, Mean Absolute Error), se podrá cuantificar cómo estima cada red neuronal con cada una de las muestras mencionadas.

ISBN 978 -987-633-574-4

A continuación se describe cada detalle de la configuración del hardware, del software, del dataset y las redes neuronales, para obtener y comparar los resultados correspondientes.

4.1 Hardware, Software y Dataset.

En cuanto al Hardware, se utilizó un procesador Intel Core I7 de 2.40GHz con 4 GB de Memoria RAM. Luego, en cuanto al Software, se desarrolló en Python mediante bibliotecas de Keras y Scikit learn.

En base a los datos obtenidos mediante la toma de datos con sensores, se dividió al conjunto en dos partes: una para entrenamiento y otra para validación, siendo esta relación de 70/30, respectivamente.

En cuanto a la configuración de las redes neuronales, se utilizó una cantidad de parámetros entrenables similares en cada una para lograr una comparación. Hay que tener en cuenta que un sistema LSTM contiene mayor complejidad en cuanto a sus parámetros entrenables con respecto a una MLP, dadas las capas internas (ver fig.2). Dada esta situación, se contempla entonces una cantidad de 8749 parámetros entrenables para MLP contra 8506 de LSTM. Luego, en cuanto a la cantidad de iteraciones (epochs), para ambas redes se utilizó un valor de 100.

Por último, para evaluar el desempeño de las redes, se utilizó como métrica al Error Medio Absoluto, el cual viene dado por la siguiente expresión:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{3}$$

Siendo:

N: Cantidad de muestras para la evaluación.
y = Valor de la temperatura.
$\hat{y}$ = Estimación de temperatura.

4.2 Comparación entre Redes Neuronales.

Teniendo en cuenta el hardware, el software y las métricas utilizadas, a continuación se muestran los gráficos de estimación de cada Red Neuronal, esto es: MLP y LSTM. En la Fig. 5 se puede ver la respuesta de la MLP.

ISBN 978 -987-633-574-4

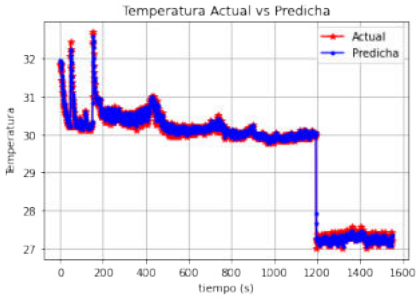
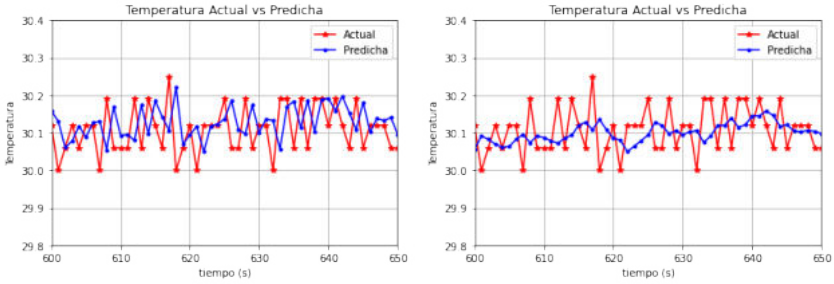


Fig. 5: Temperatura actual vs predicha con MLP

Mediante LSTM se obtiene una estimación similar a la de la Fig.5 sin embargo, dado que no se llega a visualizar cómo se realiza la estimación, a continuación se muestra un mayor detalle de la estimación de cada una de las redes:



(a) MLP (b) LSTM

Fig. 6: Temperatura actual y predicha para MLP y LSTM

Lo que se puede observar de las figuras anteriores es que con LSTM se obtiene una mejor estimación. Si bien las figuras anteriores describen cómo se produce la estimación, en el cuadro que se muestra a continuación (Tabla 1) se expone el valor de MAE para cada red neuronal teniendo en cuenta la cantidad de muestras pasadas. Con esto se podrá determinar cuál de las dos resulta conveniente a la hora de obtener estimaciones de temperatura en este Datacenter.

ISBN 978 -987-633-574-4

MAE		
Cant.Muestras pasadas	MLP	LSTM
5	0.0890	0.0939
10	0.0920	0.0909
20	0.0915	0.0915
30	0.0914	0.0915
40	0.0919	0.0920
50	0.0947	0.0907
60	0.0954	0.0890
70	0.0942	0.0855
80	0.0925	0.0851
90	0.0948	0.0884
100	0.0943	0.0883
120	0.0918	0.0876

Table 1: Tabla de Comparación entre Métricas

En la Tabla 1 se puede observar que, a medida que se incrementa la cantidad de muestras pasadas para la predicción, LSTM tiene un mejor comportamiento que MLP, es decir, el Error Medio Absoluto comienza a decrecer. Para visualizar esto, se propone la siguiente imagen, Fig. 7.

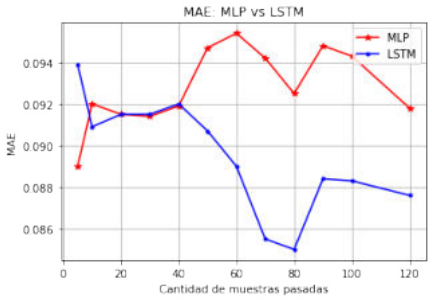


Fig. 7: MAE: MLP vs LSTM

La Fig. 7 muestra cómo evoluciona el MAE para MLP y LSTM a medida que se toma una mayor cantidad de muestras. Esto avala entonces el funcionamiento de LSTM la cual tiene en cuenta el estado previo de cada neurona en el cómputo de estimación.

5 Conclusiones

A partir de las distintas simulaciones teniendo en cuenta la cantidad de muestras pasadas, se puede observar que la mejor alternativa es LSTM. Sin embargo,

dado que ésta reviste mayor complejidad de cómputo frente a MLP, la misma tiene mayor demora en el tiempo de estimación lo cual concuerda con el modelo planteado. Si bien se plantean estas dos herramientas, existen otras como CNN y GRU que también se pueden comprobar con las mismas u otras pruebas para corroborar su efectividad. Además, existen otras herramientas de estimación como Filtro de Kalman, las cuales resultan de gran interés para su análisis.

Dado que los datos se muestrean cada 15 segundos, la acción de estimar qué sucede en t+1 mediante estas redes, especifica una idea de lo que puede suceder en los próximos 15 segundos. Por esta razón resulta importante estudiar cuál de estas redes tiene una mejor efectividad a la hora de estimar. Sobre estas acciones tomadas se puede ver que a futuro se puede realizar otro tipo de muestreo, tal vez cada 1 minuto, y observar cómo afecta a la predicción, sobre todo teniendo en cuenta que la variación de temperatura en un ambiente no suele ser demasiado rápida.

Por otro lado, se pueden seguir haciendo pruebas en cuanto a la aceleración en Hardware como por ejemplo, la utilización de una GPU, para atenuar estos tiempos de respuesta.

Referencias

1. Qiu Fang., Zhe Li.,Yaonan Wang.,Mengxuan Song., Jun Wang.: A neural-network enhanced modeling method for real-time evaluation of the temperature distribution in a data center. Springer-Verlag London Ltd., part of Springer Nature 2019.
2. Weiping Yu., Zhaoguo Wang., Yibo Xue., Lingxu Guo., Liyuan Xu.: A Combined Neural and Genetic Algorithm Model for Data Center Temperature Control. Published in CIMA@ICTAI 2018
3. Shashikant Ilager., Kotagiri Ramamohanarao.: Thermal Prediction for Efficient Energy Management of Clouds using Machine Learning. SIEEE Transactions on Parallel and Distributed Systems. Volume: 32, Issue: 5, May 1 2021. ISSN: 1045-9219.
4. Yuya Tarutani., Kazuyuki Hashimoto., Go Hasegawa., Yutaka Nakamura., Takumi Tamura., Kazuhiro Matsuda., Morito Matsuoka.: Reducing Power Consumption in Data Center by Predicting Temperature Distribution and Air Conditioner Efficiency with Machine Learning. 2016 IEEE International Conference on Cloud Engineering (IC2E). ISBN:978-1-5090-1961-8
5. Miguel Martínez Comesaña., Lara Febrero-Garrido., Francisco Troncoso-Pastoriza., Javier Martínez-Torresa.: Prediction of Building's Thermal Performance Using LSTM and MLP Neural Networks. 2020 Appl. Sci. 2020, 10(21), 7439; https://doi.org/10.3390/app10217439
6. Kim, T.-Y., Cho, S.: Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019, 182, 72–81
7. Hans Weytjens., Enrico Lohmann., Martin Kleinsteuber.: Cash Flow Prediction: MLP and LSTM compared to ARIMA and Prophet. 2019. Springer.DOI:10.1007/s10660-019-09362-7
8. Hansika Hewamalage., Christoph Bergmeir., Kasun Bandara.: Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions. 2020. Faculty of Information Technology, Monash University, Melbourne, Australia.

ISBN 978 -987-633-574-4

Performance Analysis of Simulated Annealing Using Adaptive Markov Chain Length

Carlos Bermudez[1], Hugo Alfonso[1], Gabriela Minetti[1], and Carolina Salto[1,2]

[1] Facultad de Ingeniería, Universidad Nacional de La Pampa, Argentina
[2] CONICET, Argentina
bermudezc,alfonsoh,minettig,saltoc@ing.unlpam.edu.ar

Abstract. In the Simulated Annealing (SA) algorithm, the Metropolis algorithm is applied to generate a sequence of solutions in the search space, known as the Markov chain. Usually, the algorithms employ the same Markov Chain Length (MCL) in the Metropolis cycle for each temperature. However, SA can use adaptive methods to compute the MCL. This work aims to analyze the effect of using different MCL strategies in SA behavior. This experimentation considers the Water Distribution Network Design (WDND) problem, a multimodal and NP-hard problem interesting to optimize. The results indicate that the use of adaptive MCL strategies improves the solution quality versus the static one.

Keywords: Simulated Annealing, Markov Chain Length, Water Distribution Network Design, Optimization

1 Introduction

Stochastic search optimization methods are widely used in various disciplines, such as science, engineering, management, modern statistical, machine-learning applications, to mention some. Many stochastic algorithms are inspired by a biological or physical process with some heuristic manners to find the global optimum [1]. The most common methods are simulated annealing, genetic algorithms, differential evolution, particle swarm optimization, among others [2]. In this work we focus on Simulated Annealing (SA) [3,4] due to its popularity as a search procedure because of its simple concepts, good speed, and easy implementation. SA is applied to solve NP-hard problems where it is difficult to find the optimal solution or even near-to-optimum solutions [5,6].

The Simulated Annealing algorithm is based on the principles of statistical thermodynamics. The SA simulates the energy changes in a system subjected to a cooling process until it converges to an equilibrium state (steady frozen state), where the material states correspond to problem solutions, the energy of a state to a solution cost, and the temperature to a control parameter.

The SA cooling process consists of initial and final temperatures, the cooling function, and the length of the Markov chain established by the Metropolis algorithm [7]. For each value of the temperature, the SA algorithm achieves

ISBN 978 -987-633-574-4

a certain number of Metropolis decisions. In this way, the SA consists of two cycles: one external for temperatures and the other internal, named Metropolis. Most SA literature proposals use a static Markov Chain Length (MCL) in the Metropolis cycle for each temperature [8]. But adaptive strategies to dynamically establish each MCL for the SA algorithm are also present in the literature [9,10].

The main contribution of our research is to enlarge the knowledge concerning the MCL influence on the efficiency and efficacy of a SA when solving optimization problems. In particular, we tackle the Water Distribution Network Design (WDND), which was defined as a multi-period, single-objective, and gravity-fed design optimization problem [11]. A hybrid SA (HSA), presented in [12], was used as a starting point to consider the different strategies to compute the MCL. Accordingly, research questions (*RQs*) arise out: Can the adaptive MCL strategies modify or improve the HSA performance in contrast with the static one? If they can, how do variable MC lengths affect the HSA behavior? To answer these *RQs*, we conduct experiments by applying HSA with different configurations on publicly available [13] and real-world [14] instances of the WDND problem. Furthermore, we analyze and compare these results considering the published ones in the literature.

This article is organized as follows. First, we give in Section 2 a description of the SA algorithm. In the next section, we address the strategies for computing MCL. Then, we describe the experimental design and the methodology used in Section 4. We analyze and compare the HSA behavior when solving the WDND problems in Section 5. Finally, we summarize our most important conclusions and sketch out our future work.

2 Simulated Annealing

The Simulated Annealing [3] is an efficient trajectory-based metaheuristic with the capacity of escape from local optimum. The SA generates a sequence of changes (chain) between states generated by transition probabilities, which are calculated involving the current temperature. Therefore, the SA can be modeled mathematically by Markov chains, and consists of two cycles:

- an external one, named temperature, slowly reduces the temperature to decrease defects, thus minimizing the system energy.
- an internal cycle, named Metropolis [7], generates a new potential solution (or neighbor of the current state) to the considered problem by altering the current state, according to a predefined criterion.

For each temperature, the Markov chain length usually remains without changes in the Metropolys cycle.

Figure 1 shows the general scheme of a SA algorithm, highlighting these two cycles. The SA begins with the initialization of the temperature, T, and the generation of a feasible initial solution, S_0, for the target problem. After that the two overlapping cycles begin. A new trial solution, S_1, is obtained by applying a move to the current solution, S_0, to explore other areas of the search space. At this point, S_1 is accepted with the Boltzmann probability. This process

ISBN 978 -987-633-574-4

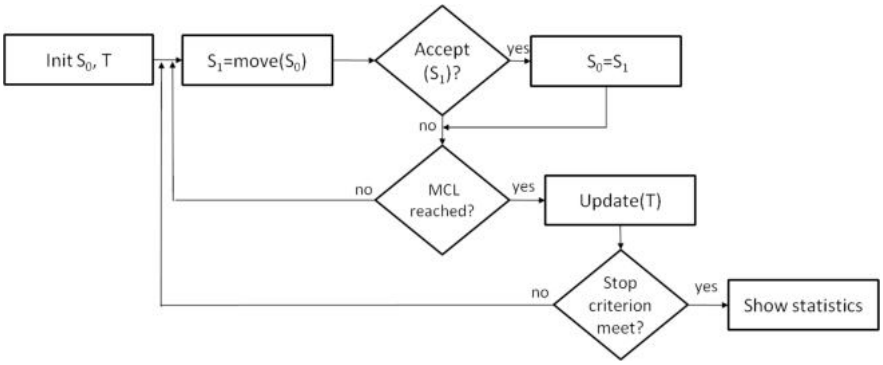


Fig. 1. Scheme of the SA algorithm.

generates a Markov chain, which is repeated until a number of steps denominated as Markov chain length. After that, the temperature in the SA algorithm is sequentially lowered until the system freezes by a cooling schedule. Finally, the SA ends the search when the total evaluation number or the temperature equilibrium ($T = 0$) is achieved.

The search space exploration is strengthened when the temperature (T) is high. But at low temperatures, the algorithm only exploits a promising region of the solution space, intensifying the search. The annealing procedure involves taking enough steps at each temperature (internal cycle). The number of steps aims to keep the system close to equilibrium until the system approaches the ground state. Traditionally, the equilibrium can be achieved by maintaining the temperature constant for a limited number of iterations; but adaptive strategies can be considered. The main objective of this work is to identify how sensitive the SA can be to the number of these iterations, by considering different strategies to compute the Markov chain length to solve NP-hard problems.

3 Markov Chain Length

The SA starts by constructing a sequence of temperatures T_1, T_2 and so on. At each step of this sequence, SA does a set of k moves to neighboring positions. Such a stochastic sequence construction is called a Markov chain, and the number of moves k is denominated Markov chain length. There are few researches in literature concerning to the effect of the MCL on the solution quality and annealing speed [9,10].

The MCL can be determined experimentally and considered static throughout the search, but also MCL can set adaptively depending on the optimization function variation. The static strategy (MCLs) assumes that each T value is held constant for a fixed number of iterations, defined before the search starts. In this work, each T value is held constant for $k = 30$ iterations, a widely used number in the scientific community. For the adaptive strategies, which depend on the characteristics of the search, we consider two different alternatives:

1. MCLa1. Cardoso et al. [9] consider that the equilibrium state is not necessarily attained at each temperature. Here, the cooling schedule is applied as soon as an improved candidate (neighbor) solution is generated. In this way, the computational effort can be drastically reduced without compromising the solution quality.
2. MCLa2. This strategy, proposed by Ali et al. [10], uses both the worst and the best solutions found in the Markov chain (inner loop) to compute the next MCL. MCLa2 increases the number of function evaluations at a given temperature if the difference between the worst and the best solutions increases. But if an improved solution is found, the MCL remains unchanged.

4 Experimental Design

In this section, we explain the experimental design tests to study the behavior of the SA introduced in [12], named HSA, using different MCL strategies to solve the WDND problem. The upcoming paragraphs briefly describe the target test problem, followed by the methodology and the parameters used.

Multi-Period Water Distribution Network Design. The mathematical formulation of the WDND is often treated as the least-cost optimization problem. The decision variables are the diameters for each pipe in the network. The problem can be characterized as simple-objective, multi-period, and gravity-fed. Two restrictions are considered: the limit of water speed in each pipe and the demand pattern that varies in time. The network can be modeled by a connected graph, which is described by a set of nodes $N = \{n_1, n_2, ...\}$, a set of pipes $P = \{p_1, p_2, ...\}$, a set of loops $L = \{l_1, l_2, ...\}$, and a set of commercially available pipe types $T = \{t_1, t_2, ...\}$. The objective of the WDND problem is to minimize the Total Investment Cost (TIC) in a water distribution network design. The TIC value is obtained by the formula shown in Equation 1.

$$\min TIC = \sum_{p \in P} \sum_{t \in T} L_p IC_t x_{p,t} \tag{1}$$

where IC_t is the cost of a pipe p of type t, L_p is the length of the pipe, and $x_{p,t}$ is the binary decision variable that determines whether the pipe p is of type t or not. The objective function is constrained by: physical laws of mass and energy conservation, minimum pressure demand in the nodes, and the maximum speed in the pipes, for each time $\tau \in \mathcal{T}$.

Methodology and Experimental Setup. To answer the *RQs* formulated in Section 1, we need the empirical verification provided by testing the HSA in a WDND test set of varying complexity. The static (MCLs) and the two adaptive (MCLa1 and MCLa2) MCL strategies are considered. Therefore, three new HSA configurations arise. The stop condition is to reach 1,500,000 evaluations of the objective function to make a fair comparison with the literature algorithms. The HSA uses the random cooling scheme [15] and 100 as seed temperature (see [16] for a justification of this parameter selection). Moreover, the testing includes 50 HydroGen instances [13] of WDND optimization problem grouped

ISBN 978 -987-633-574-4

Table 1. The best known TIC values found by our proposals and ILS.

Network	MCLs	MLCa1	MLCa2	ILS
HG-MP-1	**298000**	**298000**	**298000**	**298000**
HG-MP-2	245330	245330	245330	**245000**
HG-MP-3	310899	310706	**310493**	318000
HG-MP-4	592048	**590837**	592036	598000
HG-MP-5	**631000**	**631000**	**631000**	**631000**
HG-MP-6	617821	**609752**	614917	618000
HG-MP-7	648372	644568	**639932**	653000
HG-MP-8	795996	792436	**790037**	807000
HG-MP-9	716944	715863	**712450**	725000
HG-MP-10	730916	**712847**	727818	724000
GP-Z2-2020	355756	366684	358717	**347596**

by five different distribution networks, named as HG-MP-i with $i \in [1, 10]$, and GP-Z2-2020, a real-world case [14].

Since we deal with stochastic algorithms, we have performed 30 independent runs per WDND instance and for each HSA configuration. We have carried out a statistical analysis of the results that consists of the following steps. Before performing the statistical tests, we first check whether the data follow a normal distribution by applying the Shapiro-Wilks test. Where the data are distributed normally, we later apply an ANOVA test. Otherwise, we use the Kruskal–Wallis (KW) test. These statistical studies allow us to assess whether or not there are meaningful differences between the compared algorithms with $\alpha = 0.05$. These pairwise algorithm differences are determined by carrying out a post hoc test, as is the case of the Wilcoxon test if the KW test is used.

5 HSA Result Analysis

The result analysis is carried out considering the performance and internal behavior of the proposed HSA configurations (MCLs, MCLa1, and MCLa2).

5.1 HSA Performance

The HSA performance is analyzed considering the solutions found by each configuration, the effort required by the search, and the comparison of HSA results against the ILS ones [17], a well-known WDND solver.

To study the solution quality, we present Table 1 with the minimum TIC values for the HSA considering the three MCL strategies. Furthermore, the last column shows the TIC values corresponding to ILS. To complete the solution quality analysis, we use the relative distance between the best-known TIC value and the best TIC value of each HSA configuration as the error measure. The figures 2 and 3.a) show the distribution of the HSA errors grouped by the Hydrogen and GP-Z2-2020 networks and MCL strategies. Boxplots with different colors mean statistically different behaviors. From these results, we observe that adaptive HSA configurations improve the static one in 7 of 11 network groups. For HG-MP-i with $i \in [7, 10]$ and GP-Z2-2020 networks, the MCLs behavior is

ISBN 978 -987-633-574-4

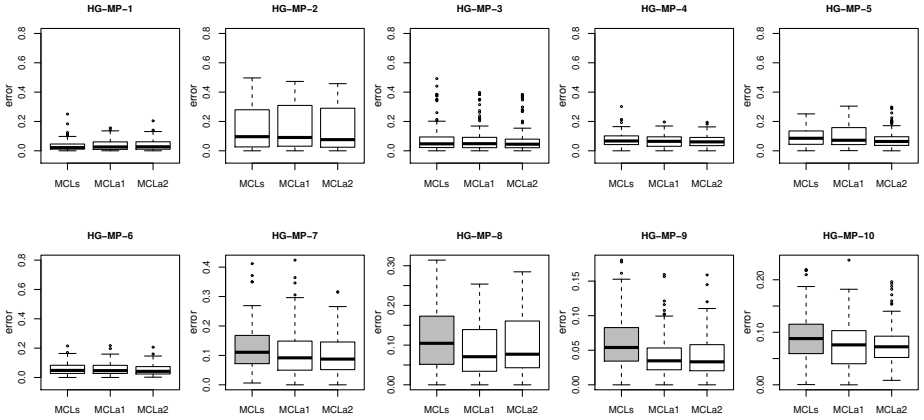


Fig. 2. BoxPlots of TIC error values found by HSA and each MCL strategies for WDND networks.

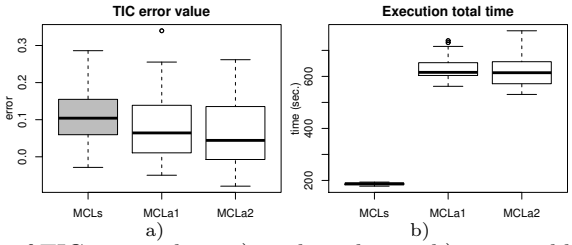


Fig. 3. BoxPlots of TIC error values, a), and total time, b), required by HSA and each MCL strategies for WDND real network (GP-Z2-2020).

significantly different (gray boxplot) to MCLa1 and MCLa2, which keep similar behavior for all cases (white boxplots). In this way, the first RQ is positively answered because the adaptive MCL strategies improve the HSA performance versus the static one, regarding efficiency and efficacy.

The following analysis is devoted to study each HSA configuration with more detail, considering the computational effort measured with the required time to execute the whole search process. Figures 4 and 3.b) show the distribution of these measures grouped by Hydrogen and GP-Z2-2020 networks and MCL strategies. First, we observe that the HSA run times grow as the instance complexity increases for all configurations. Second, MCLs is the quickest strategy for all networks, whereas adaptive HSA configurations increment significantly the total runtime. However, the MCLa1 runtimes are significantly less than the required ones by MCLa2 for HG-MP-i with $i \in [1, 4]$ networks.

Finally, we compare our results with ILS (see Table 1) from the quality point of view. In this sense, we also detect that the three HSA configurations find better average TIC values than ILS for 7 of 11 networks. Besides, all algorithms reach the same result in two cases.

ISBN 978 -987-633-574-4

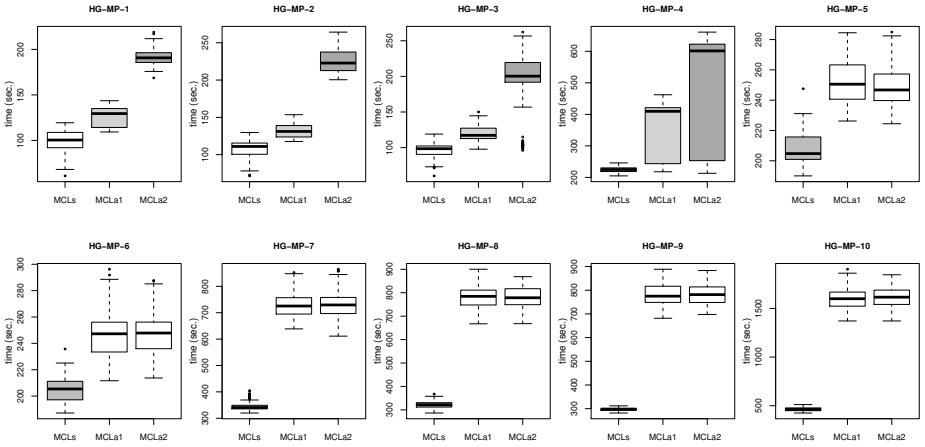


Fig. 4. BoxPlots of the total time (in seconds) required by HSA and each MCL strategies to solve the WDND networks.

5.2 HSA internal behavior

The idea behind the HSA's internal behavior analysis is to discover if the MCL strategies affect the solution quality or the temperature schedules.

Figures 5, 6, and 7 show the upper triangular matrix of scatter plots, where the correlation between the variables TIC values, MC lengths, and temperatures are graphically presented, for each HSA configuration. The Spearman's correlation coefficient, R, is calculated in every comparison and measures the linear correlation between two data sets. R belongs to the range $[-1, 1]$ and expresses the strength of association between two variables. If $R > 0$ indicates a positive relationship between the two variables (as values of one variable increase, values of the other variable also increase). When $R < 0$ indicates a negative relationship (as values of one variable increase, values of the other variable decrease). A $R = 0$ means that no linear correlation exists between the variables.

The MCLs strategy maintains constant (equal to 30) the MC length during the whole search process. Consequently, no linear correlation exists between this length and the solution quality ($R = 0$), as Fig. 5 shown. Instead, the temperature reduction is related to the solution quality because the network costs decrease during the annealing process ($R = 0.46$).

As we explain in Section 3, the adaptive strategies calculate on runtime the MC length according to different criteria. The lengths computed by MCLa1 vary in the range [730, 2850] and the calculated ones by MCLa2 belongs to [730, 3320], becoming a factor that impacts positively in the solution quality (R=0.82 and R=0.2, respectively). As figures 6 and 7 show, this impact is different when the first adaptive strategy is used, because only MCLa1 enable to decrease the MC length. Consequently, when HSA uses MCLa1 can reduce the temperature more times during the search process. This situation allows reaching a better

ISBN 978 -987-633-574-4

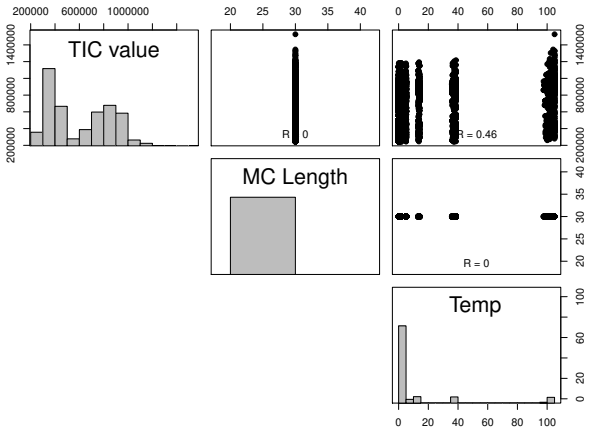


Fig. 5. Scatter plots of correlation for MCLs strategy.

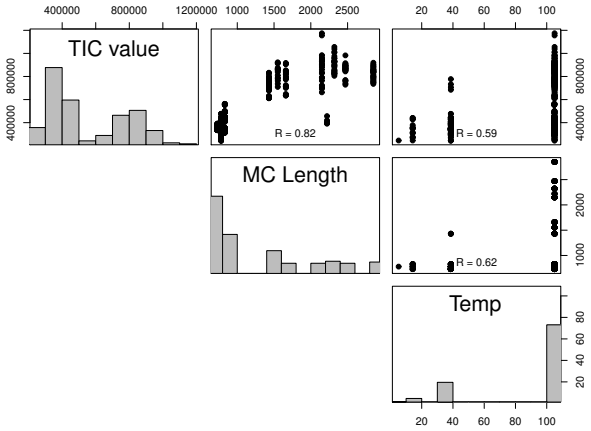


Fig. 6. Scatter plots of correlation for MCLa1 strategy.

equilibrium between exploration and exploitation in the search space, leading to a solution quality improvement.

Finally, we analyze the temperature behavior in more detail. As we can observe in the above paragraph, the variability of the MC length also affects the temperature schedule, but this relationship differs according to the adaptive strategy used. MCLa1 maintains a positive correlation (R=0.62), indicating that a diminution in the lengths is associated with a temperature reduction. Instead, these variables are inversely (R=-0.45) correlated when HSA uses MCLa2 because this strategy never decreases the MC length, but HSA always reduces the temperature after each MC ends. According to this analysis, the second *RQ* is also satisfactorily answered because HSA modifies its behavior with the MC length variability.

ISBN 978 -987-633-574-4

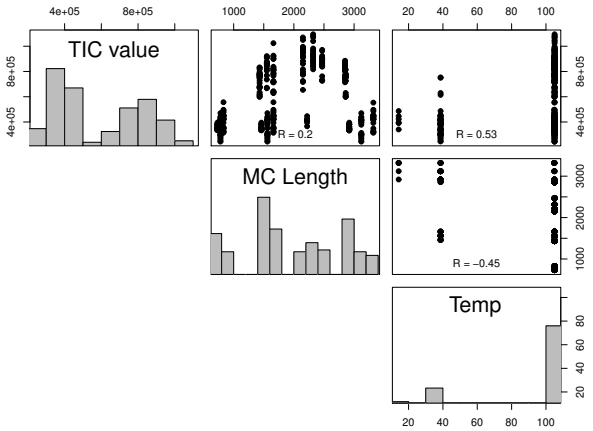


Fig. 7. Scatter plots of correlation for MCLa2 strategy.

6 Conclusions

The Simulated Annealing algorithms usually employ static Markov chain lengths in the Metropolis cycle for each temperature. However, adaptive strategies, which depend on the optimization function variability, to compute this length are also available. In this work, we contrast the static versus adaptive ones by studying the influence on the efficiency and efficacy of a SA when solving NP-hard optimization problems.

We enhance the concerning knowledge by solving several instances and real-world cases of the Water Distribution Network Design problem with a hybrid SA, in which MCL is computed by a static (MCLs) and two adaptive (MCLa1 and MCLa2) strategies. The experimentation results allowed us affirmatively to answer our research questions. The adaptive MCL strategies improve the SA performance versus the static one, modifying its behavior with the MC length variability. MCLa1 is a good trade-off between efficiency and efficacy.

A future research line consists of finding a new MCL strategy based on MCLa1 to reduce the execution time for almost all test cases. The analysis of the parallel SA behavior considering the adaptive MCL strategies for solving high-dimensional NP-hard problems is another interesting research line.

Acknowledgments

The authors acknowledge the support of Universidad Nacional de La Pampa (Project FI-CD-107/20) and the Incentive Program from MINCyT. The last author is also funded by CONICET.

References

1. J. C. Spall, *Stochastic Optimization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 173–201.
2. E.-G. Talbi, *Metaheuristics: From Design to Implementation*. Wiley, 2009.
3. S. Kirkpatrick, C. G. Jr, and M. Vecchi, "Optimization by simulated annealing," *Science*, no. 220, pp. 671–680, 1983.
4. V. Černý, "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, p. 41–51, 1985.
5. O. Borysenko and M. Byshkin, "Coolmomentum: a method for stochastic optimization by langevin dynamics with simulated annealing," *Scientific Reports*, vol. 11, p. 10705, 2021.
6. S.-W. Lin, C.-Y. Cheng, P. Pourhejazy, K.-C. Ying, and C.-H. Lee, "New benchmark algorithm for hybrid flowshop scheduling with identical machines," *Expert Systems with Applications*, vol. 183, p. 115422, 2021.
7. N. Metropolis, A. W. R. M. N. Rosenbluth, and A. H. Teller, "Nonequilibrium simulated annealing: A faster approach to combinatorial minimization," *The Journal of Chemical Physics*, vol. 21, p. 1087–1092, 1953.
8. E. Alba, C. Blum, P. Isasi, C. León, and J. A. Gómez, Eds., *Frontmatter*. John Wiley & Sons, Ltd, 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470411353
9. M. Cardoso, R. Salcedo, and S. de Azevedo, "Nonequilibrium simulated annealing: A faster approach to combinatorial minimization," *Industrial & Engineering Chemistry Research*, vol. 33, pp. 1908–1918, 1994.
10. M. Ali, A. Törn, and S. Viitanen, "A direct search variant of the simulated annealing algorithm for optimization involving continuous variables," *Computers & Operations Research*, vol. 29, no. 1, pp. 87 – 102, 2002.
11. M. Cunha and J. Sousa, "Water distribution network design optimization: Simulated annealing approach," *Journal of Water Resources Planning and Management*, vol. 125, pp. 215–221, 1999.
12. H. Alfonso, C. Bermudez, G. Minetti, and C. Salto, "A real case of multi-period water distribution network design solved by a hybrid SA," in *XXVI Congreso Argentino de Ciencias de la Computación (CACIC)*, 2020, pp. 21–3. [Online]. Available: http://sedici.unlp.edu.ar/handle/10915/113258
13. A. De Corte and K. Sörensen, "Hydrogen," available online: http://antor.uantwerpen.be/hydrogen (accessed on 27 June 2018).
14. C. Bermudez, H. Alfonso, G. Minetti, and C. Salto, "Hybrid simulated annealing to optimize the water distribution network design: A real case," in *Computer Science – CACIC 2020*, P. Pesado and J. Eterovic, Eds. Cham: Springer International Publishing, 2021, pp. 19–34.
15. C. Bermudez, G. Minetti, and C. Salto, "SA to optimize the multi-period water distribution network design," in *XXIX Congreso Argentino de Ciencias de la Computación (CACIC 2018)*, 2018, pp. 12–21.
16. C. Bermudez, C. Salto, and G. Minetti, "Designing a multi-period water distribution network with a hybrid simulated annealing," in *XLVIII JAIIO: XX Simposio Argentino de Inteligencia Artificial (ASAI 2019)*, 2019, pp. 39–52.
17. A. De Corte and K. Sörensen, "An iterated local search algorithm for water distribution network design optimization," *Network*, vol. 67, no. 3, pp. 187–198, 2016.

ISBN 978 -987-633-574-4

Aprendizaje automático aplicado al procesamiento de imágenes para la clasificación de objetos reciclables

Salina Mauro[1], Osio Jorge[1,2], Cappelletti Marcelo[1,2] , Morales Martín[1]

1 Programa de Tecnologías de la Información y la Comunicación (TIC) en aplicaciones de interés social, IIyA, UNAJ

2 Línea CeTAD, Grupo de Control Aplicado (GCA), Instituto LEICI, UNLP-CONICET

{maurosalina85} @gmail.com {josio, mcappelletti, martin.morales } @unaj.edu.ar

Abstract. El presente proyecto se basa en la utilización de técnicas de Deep Learning, específicamente se realizó el modelado de redes neuronales convolucionales (CNN) capaces de clasificar distintas imágenes de objetos reciclables, estos modelos fueron probados con una clasificación binaria (reciclable-no_reciclable) y una clasificación multiclase (plástico-vidrio-metal-papel-carton, orgánico, no_reciclable). Además, se realizaron pruebas con modelos pre entrenados, utilizando aprendizaje por transferencia (Transfer Learning) para comparar resultados. Estos modelos fueron implementados utilizando como lenguaje de programación Python, apoyándose en el Framework de backend TensorFlow y la librería de alto nivel Keras. El modelo final se probó en una aplicación (beta) implementada también en Python sobre un mini computador Raspberry Pi y un módulo de cámara (picam) en donde se toman fotos y se aplica el modelo para realizar una clasificación en tiempo real.

Keywords: Machine Learning, Deep Learning, IoT, sistema de reciclaje, Visión por computadora, procesamiento de imágenes.

1 Introducción

Una de las problemáticas actuales se centra en la importancia del reciclaje y cuidado del medio ambiente. Reciclar es de suma importancia para la sociedad, debido a que, supone la reutilización de elementos u objetos ya utilizados, los que de otro modo serían desechados contribuyendo al aumento de la formación de basura y al daño ambiental permanente [1].

En nuestro país cada dos segundos se produce una tonelada de basura y una fracción grande de ella termina en rellenos sanitarios que están al borde del colapso [2].

Se estima que los RSU son la mayoría de los desechos. Entre ellos, la basura doméstica encarna la problemática más significativa: aproximadamente la tercera parte está formada por papel y derivados, mientras que el resto se compone por plásticos, vidrio, metales y pilas [2].

Es fundamental la separación en origen, puesto que discriminar una vez que los residuos están mezclados es poco práctico y costoso. En nuestro país, aunque se han tomado

medidas para fomentar el reciclado, solo un 24% de la población se esfuerza por separar los residuos para minimizar su generación y la contaminación. Gran parte del problema radica en el esfuerzo que requiere clasificar y separar los residuos inorgánicos, es por eso que la propuesta busca desarrollar un sistema de visión por computadora que permita detectar y clasificar objetos reciclables para minimizar la cantidad de residuos que se generan diariamente.

La propuesta de trabajo se enfocó en el desarrollo de una aplicación de software encargada de realizar una clasificación de imágenes en tiempo real, mediante la cual se busca detectar la presencia de objetos reciclables contenidos en la imagen. El desarrollo se basó específicamente en el uso de redes neuronales convolucionales, las cuales han demostrado ser las más eficientes en el área del procesamiento de imágenes [3].

Para llevar a cabo la implementación del presente trabajo se utilizó una herramienta que existe desde mediados del siglo pasado, pero que ha tenido un avance significativo en la última década, la Inteligencia Artificial (IA) [4].

Una de las áreas en donde se avanzó notablemente fue en la de detección de objetos y clasificación de imágenes. Esto se debe en su mayor parte al desarrollo de nuevas técnicas de Machine Learning (Aprendizaje Automático [5-7]) como el Deep Learning (Aprendizaje Profundo [8-10]), además de las innovaciones en el manejo de Big Data (datos a gran escala) y el aumento en la capacidad de cómputo mediante el uso de diferentes tecnologías como cloud computing (computación en la nube) o el uso de GPU (unidad de procesamiento gráfico) para el análisis de información. Este avance puede verse en distintas áreas como: medicina, seguridad, turismo, finanzas, robótica, entre otras.

Machine Learning es un subcampo de la IA en el que se utilizan diferentes algoritmos para recolectar datos, y con estos realizar un aprendizaje para luego hacer una predicción o sugerencia sobre algo [7]. De esta manera se permitirá resolver problemas de forma intuitiva y automatizada, sin que el mecanismo de elección se encuentre previamente programado. En la práctica esto se traduce en una función matemática en la que se parte de una entrada y se obtiene una salida, por lo que el desafío reside en construir un modelado automático de esta función matemática.

Deep Learning es un subcampo de Machine Learning, pero existen técnicas de Machine Learning que no utilizan Deep Learning. Este último es utilizado para realizar procesos de Machine Learning empleando redes neuronales artificiales compuestas por varios niveles jerárquicos [10]. En el nivel inicial la red aprende patrones simples, y esta información se envía al siguiente nivel de la jerarquía. Este segundo nivel toma la información obtenida en el primero y la combina con nuevos patrones aprendidos en este, generando información un poco más compleja, la cual es pasada a un tercer nivel, y así sucesivamente.

2 Implementación

En este apartado se detallan las herramientas y los procedimientos que se desarrollaron para implementar el sistema.

2.1 Herramientas:

ISBN 978 -987-633-574-4

- Python: Es un lenguaje de programación multiparadigma soportando la orientación a objetos, programación imperativa y programación funcional. Es interpretado, usa tipado dinámico, open source y es multiplataforma. El framework Tensorflow y la librería de alto nivel Keras utilizados para el entrenamiento de redes neuronales son APIs que pueden ser integradas en Python, siendo esta la razón de la elección para la integración en el proyecto ([11]y[12]).
- Anaconda es una distribución libre y abierta de los lenguajes Python y R, utilizada en ciencia de datos y aprendizaje automático (machine learning). Esto incluye procesamiento de grandes volúmenes de información, análisis predictivo y cómputos científicos. Está orientado a simplificar el despliegue y administración de los paquetes de software.
- Google Colab: es un servicio cloud, basado en los Notebooks de Jupyter, que permite el uso gratuito de las GPUs y TPUs de Google, con librerías como: Scikit-learn, PyTorch, TensorFlow, Keras y OpenCV. Todo ello bajo Python 2.7 y 3.6, aún no está disponible para R y Scala.
- TensorFlow: es una librería de código abierto para Machine Learning. Esta librería de computación matemática ejecuta de forma rápida y eficiente gráficos de flujo [13]. Estos gráficos están formados por operaciones matemáticas representadas sobre nudos y cuyas entradas y salidas son vectores multidimensionales de datos, conocidos como tensores (ver Fig. 1)[14] y[15].
- Keras: Es una Interfaz de Programación de Aplicaciones (API) de alto nivel y de código abierto escrita en Python. Está especialmente diseñada para facilitar una rápida experimentación en Deep Learning [16].

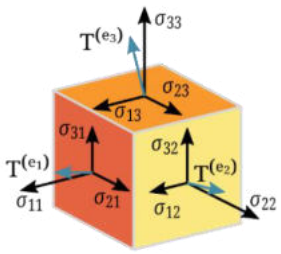


Fig. 1. Representación de un tensor

Otras librerías:

- Jupyter Notebook: El Jupyter Notebook es una aplicación web de código abierto que permite crear y compartir documentos que contienen código vivo, ecuaciones, visualizaciones y texto narrativo.

- Numpy: Librería de Python orientada al cálculo vectorial y matricial. Permite trabajar con matrices N-dimensionales, además de implementar algebra lineal, transformada de Fourier y muchas funciones relacionadas con estadísticas.

ISBN 978 -987-633-574-4

- Pandas: Es una librería de Python concebida como una extensión Numpy, que facilita el procesamiento de tablas, series temporales y otras estructuras de datos complejas.
- ScikitLearn: es una biblioteca de código abierto, que provee de herramientas científicas para el análisis de datos y el data mining [13].
- Pillow: es una librería para el manejo de todo tipo de imágenes.
- Matplotlib: es una librería específica para la generación de gráficos en Python, incluye la representación 2D y 3D de funciones e imágenes.
- OS: Permite acceder a funcionalidades dependientes del sistema operativo. Sobre todo, aquellas que refieren información sobre el entorno de este y facilitando la manipulación de la estructura de directorios.

2.2. Hardware

El uso de técnicas de Machine Learning [17], en especial de Deep Learning como es el caso de las redes neuronales convolucionales implica un costo computacional muy elevado, debido a que están compuestas por un gran número de capas y neuronas y, además, trabajan con gran cantidad de imágenes que en muchos casos tienen una alta resolución [18].

El CPU de una computadora realiza operaciones matemáticas con una velocidad muy elevada, pero las ejecuta de a una por vez, o al menos una operación por núcleo. En cambio, las tarjetas gráficas o GPU como se puede apreciar en la Fig. 2 disponen de muchos más núcleos por lo que pueden realizar muchas más operaciones en el mismo tiempo. La GPU (Unidad de Procesamiento Gráfico) es un coprocesador dedicado al procesamiento de imágenes, tiene miles de núcleos optimizados para trabajar en paralelo con operaciones sencillas, por lo que está especialmente preparada para el cálculo matricial. Básicamente, al ser otro procesador añadido, su función es la de liberar la carga del CPU, aumentando el rendimiento de nuestro ordenador. Toda esta potencia de cálculo aritmético puede emplearse para la computación de los algoritmos de Deep Learning.

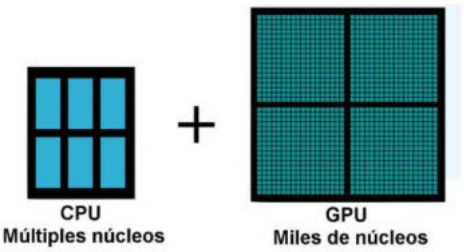


Fig. 2. Comparación entre núcleos de CPU y núcleos de GPU

2.3 Desarrollo

Durante el desarrollo del proyecto se implementaron aproximadamente 80 modelos de redes neuronales convolucionales ([19] y [20]), 60 de estos modelos fueron implementados con aprendizaje desde cero y el resto de los modelos fueron desarrollados utilizando técnicas de transferencia de aprendizaje. A su vez, las pruebas

ISBN 978 -987-633-574-4

se implementaron con 2 sets de datos, el primero con una división de dos clases y, el segundo conjunto de datos, fue dividido en 6 clases. A continuación, se detallará mejor el contenido de cada set de datos.

2.3.1 Modelo de clasificación binaria

En este caso los datos (set de datos 1) fueron divididos en 2 clases (reciclable y no reciclable), la cantidad total de imágenes de este set es de 8340 imágenes muy bien balanceadas en ambas clases, a su vez, se dividieron los datos en tres partes separando 100 imágenes para el testeo del modelo, 1708 imágenes para la validación y 6532 imágenes utilizadas para el entrenamiento. Antes de ser enviadas al modelo como entradas, se realizó un redimensionamiento por software a cada una de las imágenes para que todas tengan un tamaño de 200 x 200 píxeles. El mismo proceso de redimensionamiento se debe realizar a las nuevas imágenes que el modelo deberá clasificar luego del entrenamiento.

2.3.2 Modelo de clasificación multiclase

Para el set de datos 2 se incrementó significativamente la cantidad de imágenes y se realizó una división en 6 clases (orgánico, papel/cartón, metal, vidrio, plástico, y no reciclable). La cantidad total de imágenes de este set es de 15105 que se buscó balancear entre las 6 clases disponibles, a su vez, se dividieron los datos en tres partes separando 90 imágenes para el testeo del modelo, 3039 imágenes para la validación y 11976 imágenes utilizadas para el entrenamiento. Antes de ser enviadas al modelo como entradas, se realizó un redimensionamiento como en el modelo anterior.

3 Resultados

Durante todo el proyecto se modelaron aproximadamente 80 redes neuronales convolucionales, de las cuales los primeros 60 modelos fueron entrenados desde el inicio y los restantes 20 modelos fueron entrenados utilizando transfer learning basados en varias redes pre-entrenadas como VGG16, VGG19, ResNet50, MobileNet y XceptionV3.
Para el entrenamiento de los distintos modelos se necesitaron alrededor de dos meses de horas máquina, ya que los modelos más simples llevaron un tiempo promedio de entrenamiento de 16 horas y los modelos más complejos tuvieron un tiempo promedio de entrenamiento de 60 horas, llegando a un promedio general de 36 horas de entrenamiento por modelo propuesto.
Uno de los principales problemas que se detectó durante el desarrollo de los primeros modelos fue la presencia de overfitting, logrando buenos resultados para el set de entrenamiento y malos para el set de validación. Para los primeros modelos los resultados rondaban el 60% de acierto para la predicción de nuevas imágenes, por lo tanto, para lograr mejorar estos resultados se fueron probando pequeñas variaciones en los hiperparámetros de la red y la profundidad de los modelos. Además, se implementaron algunas de las técnicas como data augmentation, early stopping y

ISBN 978 -987-633-574-4

dropout, para lograr reducir el sobreajuste del modelo, siendo de mucha utilidad las funciones de “callbacks” que permite utilizar la librería de alto nivel Keras [16].
Las funciones “callbacks” son funciones que se ejecutan luego de cada iteración de entrenamiento, permitiendo por ejemplo realizar modificaciones al LR (learning rate) o guardar los pesos del modelo en ese momento si se cumple alguna condición predefinida. Esto se consigue debido a que estas funciones tienen acceso a todas las variables involucradas durante el entrenamiento como, por ejemplo, el porcentaje de acierto ya sea para el set de entrenamiento como para el set de validación. Las funciones elegidas fueron tres, en primer lugar, se utilizó la función “EarlyStopping” a la que se le indica la cantidad de épocas (iteraciones) que se debe esperar para terminar el entrenamiento siempre y cuando se cumpla una condición aplicada a una variable que se va monitoreando; en este caso se monitorea el valor de la función de pérdida de los datos de validación y la función corta el entrenamiento si este dato no disminuye luego de las iteraciones indicadas. En segundo lugar, se utilizó la función “ModelCheckpoint”, que permite guardar los valores de los pesos del modelo obtenido en cada iteración y una vez terminado el entrenamiento restaurar los mejores pesos, en donde el modelo se desempeña de mejor manera. Por último, la función utilizada fue “ReduceLROnPlateau”, que permite reducir el hiperparámetro LR en un factor que se define cuando se llama a la función y debe ser mayor a 0 y menor a 1, permitiendo una mayor reducción cuando el número es más cercano al nivel inferior permitido. Para el caso de los modelos planteados la función monitoreaba el valor del porcentaje de acierto del set de validación y los mejores resultados se obtuvieron cuando el factor por el que se multiplicaba el LR era “0.9” y el porcentaje de épocas a esperar para realizar el cambio era de alrededor de un 15% a un 20% de la cantidad total de épocas definidas para el entrenamiento.
Por otra parte, también se aplicó la técnica de dropout a la capa completamente conectada del modelo, variando el porcentaje de neuronas al que se le aplicaba dependiendo de la cantidad de neuronas que contenía cada capa, (este porcentaje se fue variando entre un 25% y un 75% en los distintos modelos). Esta técnica permite la desconexión de un % de neuronas para evitar que el modelo memorice un resultado.

Todas estas modificaciones permitieron una mejora considerable en el porcentaje de acierto de los modelos propuestos llegando a valores superiores al 72% para el set de datos que estaba dividido en seis clases y valores superiores al 84% para el set de datos dividido en dos clases. Una vez obtenidos estos resultados se comenzaron a aplicar los modelos pre entrenados con los cuales se logró obtener resultados mayores al 90% de acierto para el set de dos clases y valores cercanos al 80% para el set de seis clases. En el caso de los modelos que utilizaron transfer learning también se aplicaron las funciones “callbacks”.

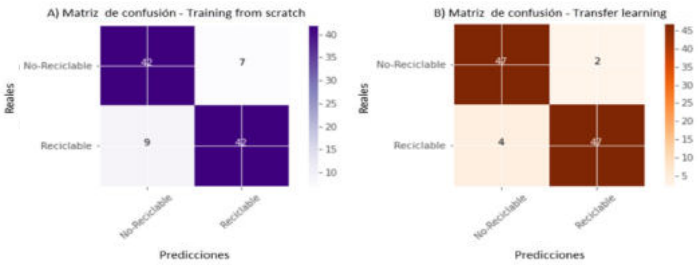


Fig. 3 – Matrices de confusion modelos binarios

ISBN 978 -987-633-574-4

Se presenta a continuación las matrices de confusión obtenidas de los 4 mejores modelos implementados, en la figura 3-A la matriz del modelo binario entrenado desde cero y la figura 3-B el modelo binario con transfer learning.

En la figura 4-A se representa la matriz de confusión correspondiente al modelo multiclase con entrenamiento desde cero, y en la figura 3-B la matriz del modelo con transfer learning.

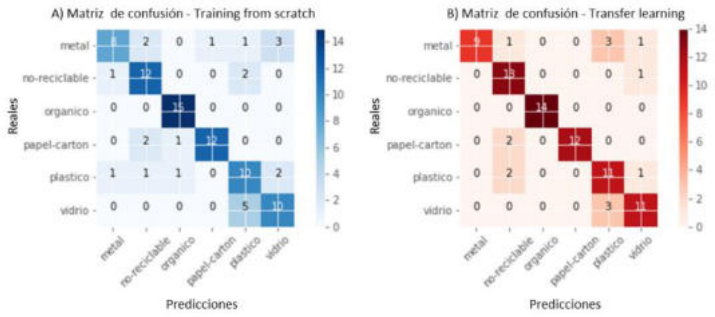


Fig. 4 – Matrices de confusión modelos multiclase

A continuación, se muestran los fragmentos de código utilizados para realizar nuevas predicciones con uno de los modelos planteados:

```
#se carga el modelo guardado ya entrenado
altura, ancho= 200,200
models='C:/temporal pps/modelo-8.4.1.vgg19/modelo.h5'
weights='C:/temporal pps/modelo-8.4.1.vgg19/pesos.h5'
cnn=tf.keras.models.load_model(models)
cnn.load_weights(weights)
#función que realiza la predicción
def predict(file):
    image=load_img(file, target_size=(altura, ancho))
    image=img_to_array(image)
    image=np.expand_dims(image, axis=0)
    prediction=cnn.predict(image)
    result=prediction[0]
    reply=np.argmax(result)
    if reply==0:
        clase='0 - METAL'
    elif reply==1:
        clase='1 - NO-RECICLABLE'
    elif reply==2:
        clase='2 - ORGANICO'
    elif reply==3:
        clase='3 - PAPEL-CARTON'
    elif reply==4:
        clase='4 - PLASTICO'
    elif reply==5:
        clase='5 - VIDRIO'
```

ISBN 978 -987-633-574-4


```
    image=mpimg.imread(file)
    plt.imshow(image, interpolation=None)
    plt.title('Categoría predicha de imagen: ' + clase)
    plt.axis('off')
    return
```

La función definida realiza la predicción sobre la imagen que recibe como parámetro e imprime el resultado de la predicción junto con la imagen analizada. A continuación, en la Fig. 5 se presentan tres ejemplos de predicciones tomadas con la cámara de la Raspberry y pasadas a la función, además se presenta un ejemplo de un falso positivo en donde se fotografió un envase de plástico y fue tomado por la red como uno de vidrio, aunque al mirar la imagen no es tan sencillo para el ojo humano hacer una buena predicción sobre la misma.

La aplicación (versión Beta), fue implementada en un miniordenador Raspberry Pi 3 Model B+, en donde se hace uso del módulo de la cámara (pi camera) de la Raspberry para tomar fotos en tiempo real y realizar la clasificación de dicha imagen determinando que tipo de objeto reciclable se encuentra en ella. Dentro de los resultados obtenidos se está evaluando no solo el porcentaje de acierto, sino también, los tiempos de predicción. Si bien se detectó que la carga inicial del modelo de red neuronal tiene una latencia de entre 30 y 40 segundos, luego, la captura de la imagen y posterior clasificación arrojó tiempos aproximados a los 10 segundos.

(a) Categoría plástico (b) Categoría metal

(c) Categoría No reciclable (d) falso positivo

Fig. 5. Resultados más significativos de las pruebas realizadas

ISBN 978 -987-633-574-4

4 Conclusiones

Durante el desarrollo e implementación de este trabajo se estudiaron algunas técnicas de Machine Learning aplicadas a la clasificación de imágenes y particularmente se buscó colaborar con el cuidado del medio ambiente utilizando la inteligencia artificial aplicada a la clasificación de objetos reciclables.

De por sí la implementación del sistema se encuentra atada a la complejidad de los altos requerimientos de una red neuronal convolucional, pero de todas formas, se logró implementar una red funcional que supera los objetivos planteados durante las primeras etapas del trabajo.

Se ha elegido Deep Learning (DL) para llevar a cabo este trabajo debido a la alta capacidad que presenta para el análisis de imágenes. Sin embargo, al programar algoritmos de DL, específicamente redes neuronales convolucionales, se presentaron algunas dificultades.

Es muy importante y complejo determinar los parámetros e hiperparámetros que mejor se adecuen al modelo propuesto impactando directamente en el resultado final. Este ajuste “fine tuning” muchas veces se considera un arte y no una ciencia para los que se inician en el campo del DL, debido a que se requiere cierta experiencia e intuición para encontrar los valores óptimos de estos hiperparámetros. En este caso, la investigación ha jugado un papel fundamental, además los parámetros e hiperparámetros se deben especificar antes de iniciar el proceso de entrenamiento. Otro aspecto que se debe tener en cuenta es la profundidad y la cantidad de capas de los modelos propuestos.

Por otro lado, la correcta elección y confección del set de datos que se utilizará para entrenar los modelos es vital para llegar a los resultados deseados. La red debe conocer de igual manera todos los elementos o clases que se quieren predecir o clasificar, por lo que, las distintas clases deben estar correctamente balanceadas, contener ejemplos representativos y, además, contar con la mayor cantidad posible de datos para lograr buenas soluciones, aunque esto se traduce a un mayor tiempo de entrenamiento. Durante el trabajo se demostró que las redes neuronales convolucionales presentan excelentes resultados en el campo de la “computer vision”, aunque uno de los inconvenientes encontrados ha sido el tiempo, ya que, a medida que los modelos se hacen más complejos crecen los tiempos de entrenamiento, llegando en el caso de este trabajo a un promedio de 36 horas para el entrenamiento de cada uno de los modelos propuestos.

Se debe tener en cuenta también que trabajar con CNN y grandes conjuntos de datos implica un alto costo computacional, por lo tanto, es muy importante contar con equipos que presenten altas prestaciones e incluso contar con GPUs (unidades de procesamiento gráfico).

En referencia a los resultados obtenidos para el conjunto de datos de prueba se puede decir que fueron satisfactorios, estos superaron las expectativas propuestas al inicio del trabajo. Además, todas las pruebas realizadas sirven como aprendizaje para futuros proyectos.

4.1 Líneas futuras

En primer lugar, se puede mejorar y aumentar el set de datos permitiendo el reentrenamiento de los modelos planteados con estas mejoras en los datos, lo que presentará una mayor robustez y mejor balanceo en las clases a clasificar.

En segundo lugar, y luego de haber mejorado el conjunto de datos, se deberían implementar nuevos modelos con distintas variaciones en los parámetros e hiperparámetros, buscando una mejora en los resultados para las distintas predicciones.

Por último, se podría implementar esta aplicación en algún sistema robótico como, por ejemplo, un cesto inteligente para separar los residuos reciclables.

Referencias

1. Maquituls España (2017), La importancia del reciclaje. Cuidemos el Medio Ambiente, https://www.maquituls.es/noticias/la-importancia-del-reciclaje-cuidemos-el-medio-ambiente/#:~:text=El%20reciclar%20o%20el%20reciclaje,de%20manera%20continua%20al%20planeta., recuperado 10 de mayo 2020.
2. Diario El Cronista (2018), Producción de basura: cuál es la realidad en Argentina y que se podría hacer, https://www.cronista.com/responsabilidad/Produccion-de-basura-cual-es-la-realidad-en-Argentina-y-que-se-podria-hacer-20180302-0075.html, recuperado 10 de mayo 2020.
3. Transfer Learning Wikipedia (2020), Transfer Learning, https://en.wikipedia.org/wiki/Transfer_learning, recuperado 01 julio de 2020.
4. Wolfgang Ertel, Introduction to Artificial Intelligence, second edition, Springer International Publishing AG 2017.
5. Kevin Murphy, Machine Learning - A probabilistic perspective, University of Cambridge, 2012
6. Zoubin Gharamani, Automatic Machine Learning, University of Cambridge, 2018
7. Miroslav Kubat, An Introduction to Machine Learning, second edition, Springer International Publishing AG 2017.
8. Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT 2017.
9. Nikhil Buduma, Fundamentals of Deep Learning, Editorial O'reilly. 2017
10. Sandro Skansi, Introduction to Deep Learning – From Logical Calculus to Artificial
11. Andreas Muller, Sarah Guido, Introduction to Machine Learning with Python, Editorial O'reilly, 2016.
12. Francois Chollet, Deep Learning with Python, MEAP edition, Manning Publications 2017.
13. Aurelien Gerón, Hands-On Machine Learning withs cikitLearn & TensorFlow, Editorial O'reilly, 2017
14. Tom Hope, Yehezkel Resheff, Itay Lieder, Learning TensorFlow, Editorial O'reilly. 2017
15. TensoFlow (2020), TensorFlow API documentation, https://www.tensorflow.org/api_docs/, recuperado 01 octubre de 2020.
16. Keras io (2020), Keras API references, https://keras.io/api/, recuperado 01 octubre de 2020.
17. Aprende Machine Learning (2020), https://www.aprendemachinelearning.com/, recuperado 25 de abril 2020.
18. Charu C. Aggarwal, Neural Network and Deep Learning, Springer International Publishing AG part of Springer Nature 2018.
19. C. Jay Kuo, Understanding Convolutional Neural Networks with A Mathematical Model, Department of Electrical Engineering University of Southern California 2016
20. Dominik Scherer, Andreas Muller, Sven Behnke, Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition, University of Bonn, Institute of Computer Science VI ,2010.

ISBN 978 -987-633-574-4

Técnicas de percepción para el uso de Inteligencia Artificial en el desarrollo de los videojuegos: Caso de Estudio Proyecto 1810

Christian Parkinson[1], Roxana Martínez[1]

[1] Centro de Altos Estudios en Tecnología Informática (CAETI)
Universidad Abierta Interamericana
Ciudad Autónoma de Buenos Aires, Argentina
{Christian.Parkinson, Roxana.Martinez}@uai.edu.ar

Abstract. La Inteligencia Artificial (IA) es aplicada en diversos ámbitos, y los videojuegos no son la excepción. Si bien existen diversos aspectos, este trabajo analiza el uso de agentes inteligentes combinados con máquinas de estados finitos en entidades autónomas. A lo largo de este documento se exponen técnicas de percepción aplicadas en NPCs (Non Playable Character) y se detalla la forma en que una entidad inteligente detecta y comprende los estímulos de su entorno, para la toma de decisiones. Luego del relevamiento realizado de juegos destacados en el mercado con características similares, se propone una técnica de percepción del entorno, que mejora la cantidad decisiones que pueden tomar los agentes de IA. Para implementar lo anteriormente dicho, se utiliza un caso de Estudio de desarrollo propio denominado "Proyecto 1810".

Keywords: Juegos Serios, Inteligencia Artificial, Técnicas de Percepción.

1 Introducción

En la actualidad, la utilización de los videojuegos se encuentra en uno de sus mejores momentos a nivel mercado internacional, como así también los juegos que se orientan al contexto de "Juegos serios". Algunos autores [1], llaman juegos serios a los "juegos diseñados específicamente para dejar un mensaje, incentivar la reflexión o el aprendizaje sobre cualquier tema que pueda considerarse útil o educativo. Pueden ser desarrollados para empresas que buscan capacitar empleados, estados que quieran entrenar a su milicia, o instituciones que intenten hacer reflexionar a la comunidad".

Para este tipo de aplicaciones, su principal objetivo no se centra en la diversión, sino en el aprendizaje. Según Rover [2] este nuevo paradigma, conlleva a la implementación de un buen modelado y creación de prototipos con el fin de proporcionar las herramientas y las tecnologías necesarias para el cumplimiento de la finalidad de este. Michael Schrage, investigador del MIT y autor del libro Juego Serio (Serious Play) [3], menciona que este tipo de juegos se definen como "cualquier herramienta, tecnología o técnica que permita a las personas mejorar la forma en que juegan en serio con la incertidumbre y que garantice el aumento de calidad de la innovación". Es decir, elevar el porcentaje de aprendizaje en las personas de una forma objetiva y funcional.

Por otro lado, Ferrer [4], define que “los videojuegos serios plantean principalmente el aprendizaje, relegando el aspecto lúdico a un segundo plano”. Además, en el entorno educativo, “los videojuegos constituyen una excelente herramienta de multiestimulación cognitivo afectiva que acelera el aprendizaje, genera placer, y potencia las habilidades digitales, el pensamiento estratégico y la creatividad, dependiendo en mayor o menor medida del tipo o género de videojuego que más se juegue” [5], es por ello que, es importante tener en cuenta el objetivo para el que se utilizará el juego, con el fin de idear temáticas particulares en el ámbito que se desee darle uso.

1.1 Características de la IA en los videojuegos

La inteligencia artificial es un concepto que hace bastante se encuentra en pleno auge, básicamente, se refiere a los procesos lógicos que se pueden desarrollar en la programación, los cuales reproducen las conductas inteligentes. Algunos autores cómo Rouhiainen, afirman que la IA es “la habilidad de los ordenadores para hacer actividades que normalmente requieren inteligencia humana. Pero, para brindar una definición más detallada, podríamos decir que la IA es la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano” [6]. Por otra parte, el autor Boden, indica que la IA posee características de “competencias psicológicas (como la percepción, la asociación, la predicción, la planificación, el control motor) que permiten a los seres humanos y demás animales alcanzar sus objetivos. La inteligencia no es una dimensión única, sino un espacio profusamente estructurado de capacidades diversas para procesar la información” [7].

Existen numerosos trabajos que tratan sobre IA en contextos de videojuegos [8], [9], [10]. Algunos autores, cómo Alcalá [11], señalan que esta incorporación “es la simulación de comportamientos de los personajes no manejados por el jugador: NPCs (Non Playable Character, Personajes No Jugadores, básicamente se refiere a que es un personaje no controlable por el jugador), enemigos, jefes finales, animales”.

1.2 Agentes Inteligentes y Máquina de Estados Finitos aplicados en videojuegos

Para el tipo de juegos del estilo de Age of Empires [12], “los avances en el campo de la inteligencia artificial en videojuegos, los enemigos eran perfectamente capaces de diseñar tácticas y estrategias que se basaban en los movimientos del jugador. Utilizaban sus recursos de forma eficiente para poder ganar la partida y anticiparse a las decisiones que pudiese estar tomando el jugador. Asimismo, podían modular la efectividad de sus tácticas para ofrecer diferente nivel de dificultad” [13]. Es decir, para el contexto de IA, pueden existir diversas funciones que están relacionadas con las distintas entidades que posee un juego.

En esta sección, se explican las técnicas utilizadas en esta propuesta, por ejemplo, la utilización del radio de visión o detección de una entidad que pueden tener los “enemigos”. Éstos son: “los radios de visión son los encargados de «avisar» a la IA del guardia cuando se acerca el jugador. Si tomamos como ejemplo juegos de plataforma,

ISBN 978 -987-633-574-4

estos tendrían un radio de detección que podría ser representado con un rectángulo o un triángulo. Para juegos en 3D se utilizan figuras sólidas como las esferas o segmentos de esferas. Todas estas figuras geométricas tienen un eje focal, es decir, el punto donde sale la vista o sensor de detección. El perímetro que detecta al jugador podría ser de cualquier figura, no necesariamente algo circular" [14]. Para el caso de juegos en 3D, se observan figuras del tipo polígonos ovalados que actúan cómo campo o radio de visión, de esta manera, el programador podrá agregar el código fuente necesario para efectuar las funciones que cumplan una lógica determina, por ejemplo: que avance o regrese a una determinada posición.

La IA en videojuegos se refiere básicamente a la implementación de NPCs, también conocidos como Bots, éstos son agentes inteligentes (autónomos), que actúan cómo "compañeros" o "enemigos". García [15] explica dos enfoques relevantes para ello:

- NPCs Competitivos (máximo rendimiento)
- NPCs Creíbles (comportamiento realista para provocar sentimientos)

García [15] también explica que los NPCs no sólo son personajes, sino que, además, pueden ser IAs de alto nivel que controlan, por ejemplo: objetos.

Un ejemplo de lo explicado anteriormente se muestra en la Figura 1, en la que se observan los sensores (para percibir el entorno, cómo ser: distancias, obstáculos, ruidos, etc.), motor de IA (calcular el comportamiento de los NPCs) y los actuadores (para modificar el entorno) [16].

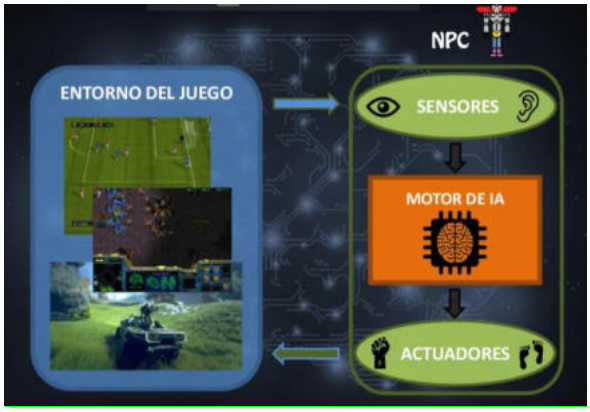


Fig. 1. Entorno de un juego y sus elementos para NPC.

Las técnicas de IA utilizadas en este trabajo son: a) ***Agentes Inteligentes:*** "Es una entidad que percibe y actúa sobre un entorno de forma razonada" [11]. Si bien existen distintos tipos de agentes, este trabajo enfoca el tipo de Agente Inteligente de Software, el cual "Es una entidad que percibe su entorno a través de sensores y actúa de forma autónoma y razonada con la mejor acción posible sobre ese entorno mediante actuadores" [11]; b) ***Máquina de estados finitos:*** "es una entidad abstracta formada por estados y transiciones entre dichos estados, las transiciones se producen por eventos sucedidos en el entorno. A su vez, la máquina genera una serie de acciones según el estado actual en el que se encuentre" [11]. Algunas de las acciones presentadas pueden ser: disparar, perseguir, entre otras.

ISBN 978 -987-633-574-4

2 Trabajos relacionados: Casos de Análisis

En esta sección se seleccionaron diferentes juegos comerciales de historia basados en guerras del tipo RPG (Role Playing Game), de los cuales se analiza la inteligencia artificial de los personajes autónomos, en función del comportamiento que adoptan en escenarios de batalla. Entre las conductas para tener en cuenta se hace foco en los criterios de persecución, ataque, huida y patrullaje.

2.1 Assassin's Creed

La saga de juegos "Assassin's Creed" de la empresa Ubisoft [17], recrea diversos escenarios históricos, donde los soldados enemigos centran su objetivo de lucha contra el personaje principal en determinadas circunstancias, lo que lleva a la conclusión del empleo de una máquina de estados. Si el jugador realiza acciones hostiles (golpear, empujar, etc.) cerca de un enemigo, éste lo persigue hasta atacarlo, en esos casos el jugador puede optar por esconderse, para ello, debe salir del campo visual del adversario, o bien responder el ataque. Cuando los enemigos encuentran un cuerpo se colocan en un modo alerta, al divisar al jugador embisten un ataque sin la necesidad de percibir una acción hostil. Si el jugador vence a varios enemigos a la vez como se muestra en la figura 2, algunos contrincantes ingresan en un "estado de pánico", dejando de atacar y huyendo aleatoriamente por el escenario, este comportamiento está sujeto a un agente inteligente que sensa la cantidad de adversarios caídos en combate para activarse.

Fig. 2. Jugador siendo atacado por varios NPCs.

2.2 Mount & Blade

Mount & Blade [18] es una saga de juegos que recrean batallas pero además incorporan recursos y administración económica, al punto que las entidades de IA se enfrentan

ISBN 978 -987-633-574-4

entre sí dependiendo del ejército al que pertenecen, como puede apreciarse en la figura 3, utilizan una máquina de estados para emplear la lógica de atacar al enemigo más cercano, e incorporan una mecánica en la que un soldado al recibir varios golpes pasa al estado de "inconsciencia" por varios segundos, permitiendo al usuario poder capturarlo y venderlo a un comerciante, o bien, al pasar el tiempo retorna el estado de pelea.

Fig. 3. Entorno de batalla de un ejército aliado contra otro.

2.3 Saga Age of Empires

La saga de juegos Age of Empires [12] posee un buen manejo de IA con una numerosa cantidad de personajes, si bien el jugador puede manejar múltiples soldados el comportamiento al detectar un enemigo es atacar. Al ejército del jugador se le puede configurar que actúe de manera defensiva, por lo cual, defenderá un área determinada sin realizar persecución alguna. Además de utilizar una máquina de estados para cambiar del modo de reposo al estado de persecución y ataque, utiliza agentes inteligentes para enviar tropas a los puntos donde los adversarios están atacando.

Fig. 4. Batalla con múltiples NPCs del juego Age of Empires.

ISBN 978 -987-633-574-4

3 Aspectos detectados en mecánicas de percepción a través de IA

En todos los juegos analizados en la sección anterior se puede observar que las entidades autónomas tienen una mecánica común a la hora de perseguir y atacar a un enemigo, éste debe ingresar dentro de un radio de visión para que el agente inteligente correspondiente cambie de estado y se focalice únicamente en ese objetivo, además, el criterio de selección del target en los juegos con múltiples actores es completamente aleatorio. Si bien el radio de visión es un elemento indispensable para cada NPC, en el presente trabajo se propone la ampliación de la funcionalidad de este basado en la toma de decisiones a la hora de seleccionar un objetivo agregando un marco de prioridades que influyan a la decisión, y, por último, la incorporación de agentes inteligentes que emulen la capacidad de audición de dichas entidades, que le permitan comprender lo que sucede en su entorno sin la necesidad de verlo.

4 Propuesta

"Proyecto 1810" [19] es un juego serio de desarrollo propio para representar las batallas de la guerra de la Independencia. De cada contienda se respeta el resultado histórico que se utiliza como punto de partida establecer la conducta que deben adoptar cada uno de los ejércitos. Si bien el juego propuesto permite realizar partidas multijugador, todos los ejércitos (aliados como enemigos) tienen personajes que participan activamente en la partida, que le brindan mayor realismo a la simulación.

Al iniciar una escena se distribuyen los ejércitos a lo largo del terreno, cada ejército está conformado por soldados que actúan y se comportan de forma autónoma. La inteligencia artificial de cada NPC contiene los siguientes aspectos: Agentes inteligentes y máquina de estado.

4.1 Agentes Inteligentes

Para simular realismo, sobre la percepción del entorno de cada soldado se establecieron dos criterios fundamentales en sus mecánicas que buscan simular los sentidos de la vista y el oído. El primer elemento es un radio de visión que consiste en utilizar un GameObject esférico alrededor del personaje, el mismo no posee renderizado, es decir, es invisible, además el collider se aplica en el modo trigger, permitiendo detectar todos los elementos que ingresan, permanecen y salen de la esfera, ignorando las leyes de la física, tal como se muestra en la figura 5. Esto le permite a la entidad tener un radio de visión de su entorno, entender cuántos aliados y enemigos hay a su alrededor, para determinar si debe perseguir, atacar o huir. El personaje se encuentra alejado del centro de la esfera, dando un mayor rango visual al frente, y menor a su espalda.

ISBN 978 -987-633-574-4

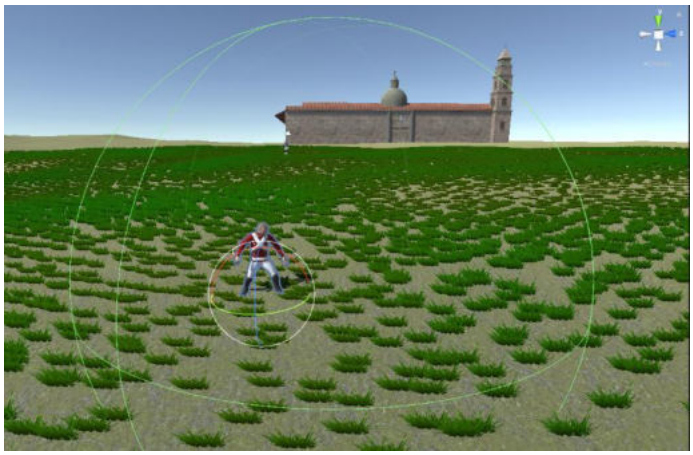


Fig. 5. Campo de visión del NPC.

El segundo elemento simula el campo auditivo, el soldado tiene la capacidad de sensar los sonidos provenientes de una fuente que puede ser de explosiones, disparos y gritos, que le permite dirigir la mirada hacia el lugar donde se emite, a fin de analizar el contexto y tomar alguna decisión. Al emitirse un sonido se utiliza un raycast (es un rayo invisible para detectar objetos) alrededor del origen que informa a cada uno de los soldados a que distancia se encuentra del mismo, como también indica de qué tipo es el origen.

Estos instrumentos son las entradas que posee el NPC que activan los procesos de los siguientes agentes inteligentes:

- Reconocimiento de aliados
- Aprendizaje de rutas de patrullaje
- Reconocimiento de nuevos objetivos de defensa
- Detección de amenazas
- Selección de objetivo de ataque
- Decisión de huida y rendición

Dependiendo del resultado histórico de la batalla representada, los soldados se inicializan con mayor o menor predisposición a rendirse, como también adoptan una postura más ofensiva o defensiva, siendo esta última la que limita a la entidad a no alejarse del objetivo que debe resguardar.

El primer agente inteligente es el “reconocimiento de aliados”, utilizando su campo visual puede reconocer soldados de su mismo bando. En caso de no estar patrullando ni peleando, y al encontrarse sin un objetivo cargado, buscará agruparse con sus pares. Al estar próximo a un NPC aliado, comienzan un intercambio de información donde actualizan las tablas de objetivos y rutas, siendo éste el segundo agente inteligente que entra en acción. El “aprendizaje de rutas de patrullaje”, permite mediante el intercambio de información entre entidades, mantener actualizados todos los sectores navegables de prioridad para la entidad. Cada ruta está determinada por diferentes puntos que un soldado puede recorrer, entre ellos están los puntos de regeneración (spawns) y

ISBN 978 -987-633-574-4

objetivos a defender, que son también detectados e interpretados gracias a la esfera de visión, pudiendo determinar también la ausencia por destrucción. O bien si el ejército aliado conquista un nuevo punto entra en actividad el tercer agente de inteligencia, “Reconocimiento de nuevos objetivos de defensa”.

El agente de “Detección de amenazas” permite a la entidad detectar en su radio de visión la cantidad de enemigos próximos, la conducta que están teniendo y los gritos que están emitiendo. El criterio que el NPC evalúa es la cantidad de aliados y enemigos para determinar si se encuentra una posición de ventaja o desventaja, también considera los gritos de “huida” emitidos por los soldados, para analizar si debe o no escapar del lugar. En el caso de estar defendiendo un objetivo, priorizará la defensa del lugar y escapará si éste ha sido destruido. Para la “Selección de objetivo de ataque”, se establece un sistema de prioridades, en primer lugar, garantiza su propia existencia, es decir, lucha con aquel soldado que lo esté atacando, si se ve atacado por más de un enemigo, elige al que tenga más chances de derrotar, en caso de paridad, toma la decisión de forma es aleatoria si las chances son las mismas. En el caso de no estar siendo atacado, el objetivo lo selecciona en base a los siguientes criterios, en primer lugar, persigue y ataca al enemigo que no se encuentre luchando, en el caso de estar todos los adversarios combatiendo, selecciona a aquel que cuente con más chances de eliminar, propiciando una ventaja competitiva al propio ejército. Cabe aclarar que las reglas de ataque quedan inhibidas en caso de que el NPC se encuentre huyendo. El agente determina la “Decisión de huida y rendición” en primera instancia basándose en el resultado de la batalla, cada soldado está condicionado al mismo, lo que hace que la decisión de rendirse quede sujeta al cumplimiento de objetivos por parte del jugador, como también a una desventaja numérica entre ejércitos con una relación de 5 enemigos a 1 aliado en adelante. Cada entidad entiende que debe huir cuando la cantidad de soldados enemigos en su rango de visión es dos veces mayor que la cantidad de aliados, por lo cual, emite un “grito” (emisión de sonido detectable para otras entidades) para indicar que debe retirarse. Cuando otro soldado recibe el sonido de retirada, dirige el rango de visión hacia el origen del “grito”, y realiza nuevamente los cálculos adicionando las nuevas detecciones provistas por la rotación de su vista, logrando advertir a otros soldados para realizar una huida colectiva. Cuando decide huir, se dirige al objetivo de defensa más próximo, y en caso de no existir, se dirige al punto de spawn. La huida se suspende en caso de encontrarse con soldados que estén patrullando sin que haya ningún enemigo en el radio de visión, esto produce que dicho soldado se acople a la patrulla. Todos los agentes inteligentes generan procesos que repercuten de forma directa en las variables que afectan a la transición de los estados del personaje.

4.2 Máquina de estados finitos

Complementario a los agentes de inteligencia está la máquina de estados finitos que proporciona las acciones básicas que cada soldado puede realizar dentro del juego, las mismas pueden apreciarse dentro de la figura 6, junto con sus respectivas transiciones.

ISBN 978 -987-633-574-4

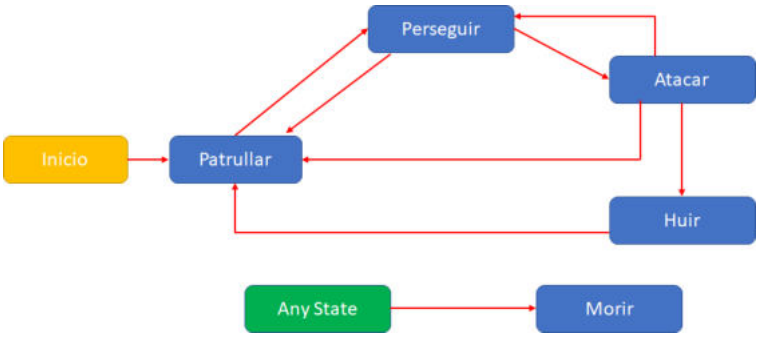


Fig. 6. Máquina de estado de los NPC de “Proyecto 1810”.

En el momento que una entidad se instancia dentro del juego la primera transición que realiza es hacia el estado de “patrullar”. Si el resultado de la batalla se corresponde con una derrota, el NPC en este estado se desplaza a los diferentes puntos de defensa, en caso de no existir ninguno, busca un soldado del mismo bando que se encuentre más próximo, y se traslada a su ubicación. En el caso de que el resultado de la contienda sea una victoria, el soldado patrulla hacia un objetivo que sus enemigos deben defender, con el fin de atacar y destruirlo.

La selección del adversario está dada por la proximidad, y en caso de igualar la distancia decide ir al objetivo con mayor cantidad de soldados aliados para mantener una ventaja dentro de la contienda, y en caso de persistir la igualdad, la decisión queda librada de forma aleatoria.

Del estado de “Patrullar” pasa al estado de “Perseguir” cuando un enemigo ingresa dentro del rango de visión, el agente inteligente de “Detección de amenazas” cambia el valor de una variable, y le indica cual el enemigo al que debe iniciar su persecución. En el momento que el enemigo sale del alcance o muere, el soldado retorna al estado de “Patrullar”, caso contrario, al alcanzar al enemigo el estado de “Persecución”, cambia al de “Atacar”, si el enemigo huye, el soldado lo perseguirá volviendo al estado de “Percusión”, en caso de ser derrotado, el soldado pasa al estado de “Patrullar”, y si el agente detecta una condición de desventaja, cambia el estado a “Huir”.

En el estado de “Huir” la entidad se desplaza a lo largo de la escena buscando un punto de defensa, soldados patrullando o bien un punto de restauración (spawn) que le permite retornar al estado de “Patrullar”.

Todos los estados tienen una transición al estado “Morir”, que ocurre cuando la vida del NPC llega a 0.

En líneas generales la máquina de estados y los agentes inteligentes se complementan para poder abordar la inteligencia de cada entidad del juego serio “Proyecto 1810”.

5 Conclusión

En el presente trabajo se realizó un análisis de la Inteligencia Artificial aplicada en el comportamiento de las entidades autónomas (NPC) de diferentes videojuegos

comerciales de gran popularidad, pudiendo apreciarse el uso de diversos agentes inteligentes para la toma de decisiones combinados con una máquina de estados finito.

En base a este análisis se realizó una implementación dentro del juego serio de propia autoría “Proyecto 1810”, incorporando un nuevo elemento para la percepción que complementa al radio visual de los soldados autónomos, lo que permitió ampliar la cantidad escenarios posibles para la toma de decisiones en cada agente de IA. Estas decisiones pueden ser afectadas por las acciones grupales de otros NPCs, como también se priorizan objetivos tanto de ataque como de defensa, en beneficio del grupo al que pertenece sin descuidar su propia existencia.

Referencias

[1] Syloper - Transformación Digital - Desarrollo de software a medida y aplicaciones, ¿Qué son los juegos serios?, https://www.syloper.com/blog/educacion/que-son-los-juegos-serios/
[2] Rover, D. T. (2005). Serious Play. Journal of Engineering Education, 94(2), 279.
[3] Schrage, M. (1999). Serious play: How the world's best companies simulate to innovate. Harvard Business Press.
[4] Ferrer, J. R. C. (2018). Juegos, videojuegos y juegos serios: Análisis de los factores que favorecen la diversión del jugador. Miguel Hernández Communication Journal, (9), 191-226.
[5] Lárez, B. E. M. (2006). Estimulación emocional de los videojuegos: efectos en el aprendizaje. Teoría de la Educación. Educación y Cultura en la Sociedad de la Información, 7(2), 128-140.
[6] Rouhiainen, L. (2018). Inteligencia artificial. Madrid: Alienta Editorial.
[7] Boden, M. A. (2017). Inteligencia artificial. Turner.
[8] Gajardo, I., Besoain, F., & Barriga, N. A. (2019, November). Introduction to behavior algorithms for fighting games. In 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) (pp. 1-6). IEEE.
[9] Angarita, L. B., Taborda, M. A. P., Márquez, J. D. E., & Díaz, J. D. V. (2021). Videojuego para el aprendizaje de lógica de programación. Revista Educación en Ingeniería, 16(31), 46-56.
[10] De Miguel Platero, C. (2018). SUMMON STORM. Programación y diseño del videojuego e inteligencia artificial.
[11] Alcalá, J. (2011). Inteligencia artificial en videojuegos. Ciclo de conferencias Game Spirit, 2.
[12] https://www.ageofempires.com/ Consultado: Agosto 2021
[13] TokioSchool, “La inteligencia artificial en videojuegos”, Disponible en: https://www.tokioschool.com/noticias/inteligencia-artificial-videojuegos/
[14] NaviGames, “El uso de las matemáticas para el desarrollo de los videojuegos”, Disponible en: https://www.navigames.es/articulos/matematicas-en-desarrollo-de-videojuegos/
[15]Inteligencia Computacional en Videojuegos (Meetup GranadAI 2019), Disponible en: https://es.slideshare.net/Slidemora/inteligencia-computacional-en-videojuegos-meetup-granadai-2019-153560968
[16] Rigau, G., & Urretavizcaya, M. Técnicas Avanzadas de Inteligencia Artificial.
[17] https://www.ubisoft.com/es-es/franchise/assassins-creed/ Consultado: Agosto 2021
[18] https://www.taleworlds.com/en/Games/Bannerlord Consultado: Agosto 2021
[19] http://1810.uai.edu.ar Consultado: Agosto 2021

ISBN 978 -987-633-574-4

A Neural Network Framework for Small Microcontrollers

César A. Estrebou[1][0000−0001−5926−8827], Martín Fleming[2], Marcos D. Saavedra[2], and Federico Adra[2]

[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata
{cesarest}@lidi.info.unlp.edu.ar
[2] Facultad de Informática, Universidad Nacional de La Plata

Abstract. This paper presents a lightweight and compact library designed to perform convolutional neural network inference for microcontrollers with severe hardware limitations. A review of similar open source libraries is included and an experiment is developed to compare their performance on different microcontrollers. The proposed library shows at least a 9 times improvement over the implementation of Google *Tensorflow Lite* with respect to memory usage and inference time.

Keywords: Machine Learning · Convolutional Neural Networks · Microcontrollers · Framework · TinyML

1 Introduction

A few years ago it was unthinkable to implement machine learning or neural network algorithms in microcontrollers, mainly for their hardware limitations. Due to cloud computing problems [1, 2] associated with computational and storage cost, network bandwidth, response latencies, power consumption, privacy and security, Edge computing started to emerge and gradually the idea of running major algorithms on microcontrollers became a reality.

On the other hand, projections made by Statista [3] estimate that the number of IoT devices connected to the Internet by 2022 will be around 16.4 billion, implying a large computing capability with low power consumption and great potential for exploitation.

Because of this, it is extremely interesting to adapt solutions from the machine learning [4] and deep learning fields so that they can run on small devices with given hardware limitations.

Today there are online platforms such as *AlwaysAI*, *Edge Impulse*, *CartesiamAI* or *Qeexo* that perform the entire process of developing a machine learning solution on a microcontroller with minimal user intervention. Companies such as *Google*, *STM*, *Mbed*, *Adafruit* and *Sparkfun* have free tools that allow implementing models created with *TensorFlow/Keras* for a limited number of microcontrollers (mainly ARM) that require 32-bit architectures with hardware that supports floating-point instructions and even SIMD or DSP instructions.

ISBN 978 -987-633-574-4

Generally, these tools are provided by microcontroller development companies or companies that provide development kits interested in promoting their products or in paying a fee to use them fully.

As a result, this limits the implementation of machine learning solutions on a large number of microcontrollers despite their popularity, low cost or additional hardware features. There are few open source machine learning libraries initiatives and very few provide support for neural networks and even fewer for convolutional networks. In general, these alternatives, besides being incipient, usually lack support and have important limitations for the wide variety of microcontrollers available in the market.

In this context we have created a small group aimed at researching and developing machine learning software for microcontrollers with significant hardware limitations, trying to cover as many of them regardless of their architecture. This paper presents an open source C/C++ library that allows to perform convolutional neural network inference on small microcontrollers without minimum hardware requirements beyond data and program memory. It also presents a tool that adapts and transforms neural network models generated with *Tensorflow/Keras* to a C, C++ or Arduino compatible version.

This article is organized as follows. Section 1 contains this introduction. Section 2 describes the process of developing machine learning models on microcontrollers.Section 3 describes open source libraries for machine learning and presents an implementation of our own.Section 3 describes open source libraries for machine learning and presents an implementation from us. In section 4, an experiment is performed to determine the performance between libraries and compare the obtained results. Finally, in section 5, conclusions and future work are presented.

2 Machine Learning in Microcontrollers

2.1 Microcontroller Development Process

Due to memory and computational capacity limitations, building machine learning models on small microcontrollers (MCUs) is a generally an impossible process. Typically, model building is done in the traditional way on a computer and then a transformation process is applied to produce a version that can be run on a microcontroller. A schematic of the steps involved in developing a machine learning model for a microcontroller is shown in Figure 1. The process starts with model selection and parameter settings. Then the model is generated using training data to finally validate its effectiveness with test data. If the result is not satisfactory, the process is restarted by reconfiguring the parameters.

Once the model is obtained, a quantization is usually performed [5, 6] in order to reduce its size and improve performance on the microcontroller. Then a tool is used to export the model, usually in C/C++ language, together with the necessary functions to carry out the inference.

Finally, the application is compiled and if the executable fits the required data and program memory size, it is deployed on the device.If the executable

ISBN 978 -987-633-574-4

does not meet the memory requirements or behaves unstable, it is returned to the model optimization point or to the development starting point to reconfigure the model parameters.

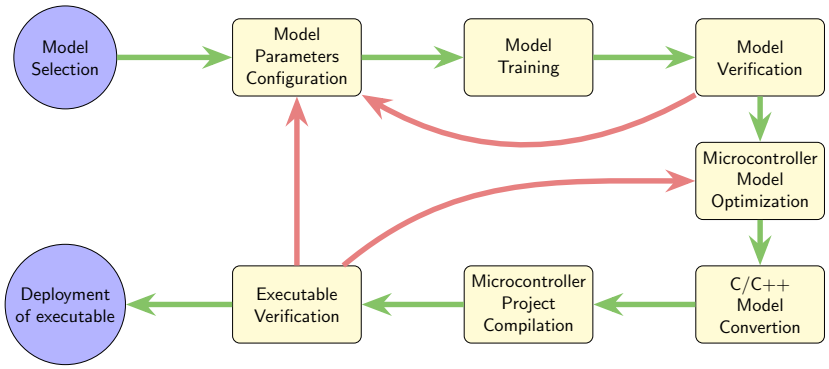


Fig. 1: Development cycle of a machine learning model for microcontrollers.

3 Neural Networks for Microcontrollers

3.1 Existing Libraries and Frameworks

In the following, this section briefly describes the open source libraries and frameworks available for the development of machine learning applications for microcontrollers.

Tensorflow Lite Micro [7, 8] for microcontrollers requires 32-bit platforms and is coded in C++ 11. It primarily supports architectures of the *ARM Cortex-M* series and has been ported to other architectures such as *Esp32*. The framework is available as an Arduino library. It can also generate projects for development environments, such as *Mbed*. It is open source and can be included in any C++ 11 project.

μTensor [9] is a lightweight machine learning inference framework built in *Tensorflow Lite* that is optimized for ARM architecture-based microcontrollers. It takes a model generated in *Tensorflow* and produces .cpp and .hpp files containing C++ 11 code to perform the inference. It does not currently support softmax functionality.

ARM CMSIS-NN [10] has a library for fully connected and convolutional neural networks named CMSIS-NN (Cortex Microcontroller Software Interface

Standard Neural Network) that maximizes the performance of Cortex-M processors with support for SIMD and DSP instructions. It includes support for 8-bit and 16-bit data types for neural networks with quantized weights.

EdgeML [11] is a library of machine learning algorithms for severely resource-constrained microcontrollers. It allows training, evaluation and deployment on various target devices and platforms. *EdgeML* is written in *Python* using *Tensorflow/Keras* and supports *PyTorch* and optimized C++ implementations for certain algorithms. Convolutional neural networks are not supported at the moment.

Eloquent TinyML [12] is an Arduino library that aims to simplify the deployment of *Tensorflow Lite* models for compatible Arduino board microcontrollers using the Arduino IDE. Starting from a model exported with *Tensorflow Lite*, this library exposes an interface to load a model and run inferences.

3.2 EmbedIA-NN, an Ultralight Library

In general, the libraries and frameworks mentioned in the 3.1 section, although they have their advantages, also have some important disadvantages, especially for microcontrollers with severe limitations. The most relevant is that they are mostly developed and optimized for specific architectures such as *ARM Cortex-M*, for 32-bit microcontrollers and/or microcontrollers with support for floating-point, DSP or SIMD instructions. This excludes devices of other architectures or devices that do not have hardware for specialized mathematical computation. Another limitation that these libraries usually have is that they are developed for C++ 11 and supported on heavy software architectures, based on objects with inheritance and polymorphism that increase the size of the programs and slow down the inference time of the algorithms. This approach may be viable for microcontrollers with good memory size and hardware resources that accelerate mathematical computation, but it is unsuitable for microcontrollers with low computational capacity and limited hardware resources.

In this article we present the development of a compact and lightweight open source library, designed for microcontrollers that are really limited both in memory and hardware. It is implemented in C, C++ and Arduino code so that it can be compiled on any platform that supports these programming languages. It provides functionalities to perform inference and debugging of the models from the microcontroller. It supports different neural network layers and activation functions including convolutional, max pooling, flatten, fully connected, ReLU and softmax. At the moment no optimizations were implemented to take advantage of advanced hardware instructions for specific microcontrollers, but there are plans to incorporate them in the future. However, optimizations are implemented for fixed-point arithmetic in 32 bits, 16 bits and 8 bits. This speeds up inferences, reduces program size and RAM usage on microcontrollers without floating point support.

ISBN 978 -987-633-574-4

In addition to the library, there is a tool that converts a model created in *Tensorflow/Keras* to C code. It also allows to generate a C, C++ or Arduino project that includes functions to perform the inference on the converted model including fixed-point optimization options.

4 Library Benchmarking Experiment

4.1 Description of the experiment

In order to determine the performance of the library, it was decided to perform an experiment by building a convolutional neural network model [13] on *Tensorflow/Keras* to recognize images of ten handwritten digits.

With the model built, a single project was developed and replicated for each library in the section 3.1 and in the four Embedia-NN implementations (8-bit, 16-bit and 32-bit floating point and fixed point). The source code for the project includes the model, the neural network functionalities to perform inference and a minimum of serial communication functionality so that each microcontroller can receive a sample and send the classification result, along with the effectiveness and time required. As part of the experiment, each project was compiled and deployed on the five selected microcontrollers. Each image of the test dataset was then submitted and each classification response was computed to determine the performance of the microcontroller-library combination. The features considered for benchmarking the different libraries were, program memory size, data memory size, inference time, and test dataset success rate.

4.2 Microcontrollers of the Experiment

The choice of the microcontrollers used in the experiment was based on aspects such as local availability, low cost, low to medium-low computational capacity and availability of open source software. Regarding connectivity it was decided to incorporate both IoT and non- IoT devices, since from the point of view of machine learning and neural networks there are many popular and interesting devices with and without this feature.

For testing purposes, 5 microcontrollers of varying characteristics were used. These MCUs are *ATmega2560*, *Arm Cortex-M3*, *Tensilica L106*, *Xtensa LX6* and *RP2040* and the technical characteristics can be seen in the table 1.

4.3 Experiment Dataset

MNIST (Modified National Institute of Standards and Technology database) is a dataset frequently used to evaluate image classification algorithms in areas of machine learning, neural networks and image processing. The chosen dataset is a reduced version of the UCI [14] repository, provided by in the *Scikit-learn* library [15]. This comprises a selection of 1797 grayscale images from the original dataset with handwritten digits centered in an 8x8 pixel area.

For model training and testing, the dataset was divided into 80An example of the dataset can be seen in Figure 2a.

ISBN 978 -987-633-574-4

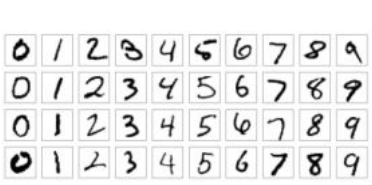
Development Board	MCU	Clock	Memory			Flot. Pt.	Connectivity
			Bits	Data	Prog.		
Arduino Mega	ATmega2560	16MHz	8	8KiB	256KiB	No	No
Stm32f103c8t6	Arm Cortex-M3	72MHz	32	20KiB	64KiB	No	No
NodeMCU ESP8266	Tensilica L106	80MHz	32	80KiB	512KiB	Si	Wi-Fi
ESP32-WROOM	Xtensa LX6	160MHz	32	320KiB	512KiB	Si	Wi-Fi+BT
Raspberry Pi Pico	RP2040	133MHz	32	264KiB	2MiB	No	No

Table 1: Relevant technical characteristics of the microcontrollers used in the experiment

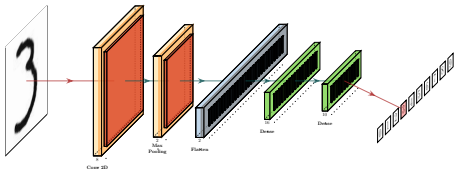
4.4 Experiment Model

A convolutional neural network (CNN or ConvNet) model [13, 16] was used to carry out the experiment tests. This type of networks are multi-layer artificial neural networks specialized in handling two-dimensional input data. Typically, their architecture is composed of combinations of convolutional, nonlinear, pooling and fully connected layers. The convolutional layer takes an image and decomposes it into different feature maps. The sequencing of various layers generates different levels of abstraction as the information progresses through the network. In the first layers low level features such as edges are obtained while in the last layers more complex and abstract structures such as parts of objects are detected. Finally the features extracted by the convolutional layers are processed by one or more layers of fully connected neurons that end up classifying the input image.

To determine the architecture of the network model we experimented with different combinations of layers in order to guarantee a good percentage of effectiveness and a low number of hyper-parameters. This last feature is of fundamental importance to maintain a small byte size to ensure that the model fits on all test microcontrollers. Figure 2b shows the architecture scheme of the convolutional neural network model generated with *Tensorflow/Keras* for testing.



(a) Digits of the MNIST dataset[15]



(b) Convolutional neural network scheme

Fig. 2: Dataset and scheme of the convolutional neural network model used in the experiment.

ISBN 978 -987-633-574-4

4.5 Comparisons and Results

The libraries *Google Tensorflow Lite*, *Eloquent TinyML*, *µTensor (microTensor)*, and *EmbedIA* in their four versions were taken to perform the tests.

It should be mentioned that two of the libraries mentioned in the 3.1 section were not considered. One of them was *CMSIS-NN* which in its official site indicates that it is possible for the library to work with processors of series prior to those supported. However, we were not able to compile the projects because it apparently requires SIMD or DSP instructions, support that the chosen microcontrollers do not have. Another one was *Microsoft EdgeML* which was also excluded from the tests because at the moment it does not support the convolutional layers included in the test model. Regarding the *µTensor* it should be mentioned that since it does not support softmax layers, the latter was replaced by a fully connected layer for the tests.

Table 1 shows the values of data memory, program memory, inference time and accuracy of the test performed for each version of the library in each microcontroller.

MCU	Library	Variant	Program Mem. (Kib)	Data Mem. (Kib)	Inference Time (μs)	Accuracy (%)
ATMega 2560 Arduino Mega	Embedia NN	Floating Pt.	14.04	5.75	75498	98.89
		Fixed Pt. 32 bits	15.49	5.75	87408	98.89
		Fixed Pt. 16 bits	**11.38**	**3.09**	**37757**	**98.89**
		Fixed Pt. 8bits	9.47	1.77	15221	89.72
STM32f103c8t6 Bluepill	µTensor	Floating Pt.	31.26	4.95	5945	98.89
	Embedia NN	Floating Pt.	23.01	0.67	9834	98.89
		Fixed Pt. 32 bits	19.02	0.67	2746	98.89
		Fixed Pt. 16 bits	**15.54**	**0.54**	**2449**	**98.89**
		Fixed Pt. 8 bits	14.32	0.48	2384	89.72
Tensilica L106 NodeMCU	Eloquent TinyML	Floating Pt.	130.17	25.16	11531	98.89
	Tensorflow Lite	Floating Pt.	115.61	23.46	11549	98.89
	Embedia NN	Floating Pt.	17.12	19.63	8213	98.89
		Fixed Pt. 32 bits	15.94	5.88	5012	98.89
		Fixed Pt. 16 bits	**13.18**	**3.39**	**1489**	**98.89**
		Fixed Pt. 8 bits	12.02	2.11	1705	89.72
Xtensa LX6 Esp 32 Devkit	Eloquent TinyML	Floating Pt.	201.49	12.96	1885	98.89
	Tensorflow Lite	Floating Pt.	191.34	8.90	794	98.89
	Embedia NN	Floating Pt.	19.03	0.94	**284**	**98.89**
		Fixed Pt. 32 bits	18.31	0.94	341	98.89
		Fixed Pt. 16 bits	**15.81**	**0.81**	367	**98.89**
		Fixed Pt. 8 bits	14.54	0.75	361	89.72
RP 2040 Raspberry Pico	Eloquent TinyML	Floating Pt.	90.98	23.28	12833	98.89
	Tensorflow Lite	Floating Pt.	129.53	21.99	10862	98.89
	µTensor	Floating Pt.	29.14	12.47	16400	98.89
	Embedia NN	Floating Pt.	9.54	6.07	9468	98.89
		Fixed Pt. 32 bits	6.32	2.38	3241	98.89
		Fixed Pt. 16 bits	**6.05**	**2.25**	**1258**	**98.89**
		Fixed Pt. 16 bits	5.98	2.19	1291	89.72

Table 2: Comparison of memory footprint and inference time required by the libraries in each microcontroller.

ISBN 978 -987-633-574-4

A significant advantage of the different EmbedIA-NN implementations over the other libraries can be seen in the table 2 and the charts in figure 3. While the 8-bit and 16-bit fixed-point implementations stand out, the latter is better because it maintains the same level of accuracy as the other libraries, while the former falls around 10%. Another remarkable aspect is that these two implementations exceed, on average, at least 9 times the memory and inference time requirements of the *Google Tensorflow Lite* and *Eloquent TinyML* libraries.

Another interesting aspect to note is the difference in performance for fixed-point arithmetic implementations on those processors such as *Stm32f103c8t6* and *RP2040* that do not have support for floating-point arithmetic.

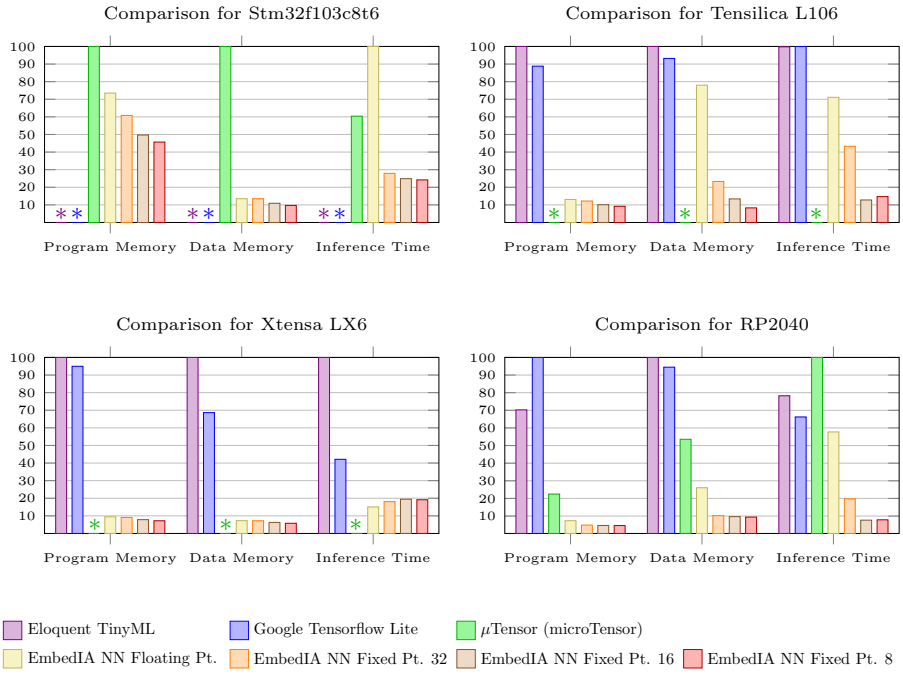


Fig. 3: Comparison of memory usage and time consumption between libraries for different microcontrollers. The unit is expressed as a percentage with respect to the library that had the highest value in the evaluated characteristic.

As the reader may have noticed, the microcontroller *ATMega2560* was not included in the graphs in the figure 3 because it only fit in memory the executables corresponding to the EmbedIA-NN implementations. As an example of the library's potential, a prototype was created for this microcontroller that recognizes handwritten digits on a 240x320 pixel graphics display with a built-in touch screen. This example integrates the experiment model, the inference func-

ISBN 978 -987-633-574-4

tions, the graphics routines code and the touch screen handling code into only 24Kib of program memory and 6Kib of RAM.

Fig. 4: ATMega2560 example integrating model, inference functions, graphics and touchscreen functions in 24Kib of program memory and 6Kib of RAM.

5 Conclusions and Future Work

This paper presented an ultralight and compact library for neural networks, designed to run on small microcontrollers with severe hardware limitations, combined with a *Tensorflow/Keras* model conversion tool and automatic code generation for C/C++ language. It was compared with other alternatives and it was shown that the 16-bit fixed-point implementation achieves at least a 9 times improvement over the memory footprint and inference time of other libraries, while maintaining the same accuracy. The advantage of EmbedIA lies in its combination with the model conversion tool that generates C language projects incorporating only the strictly necessary source code, while other C++ libraries implement class-based software architectures with inheritance and polymorphism that consume a considerable amount of data memory and program memory, and also slow down program execution. EmbedIA is an open source library that is part of a recently emerged project and will be released soon. For the future, it is planned to incorporate in a gradual way: machine learning and neural network algorithms for small microcontrollers; support for taking advantage of microcontroller hardware features with SIMD or DSP instructions; examples with practical, interesting and meaningful models for popular platforms such as Arduino.

In the short term we plan to incorporate development boards with ARM processors to compare with libraries that only support these microcontrollers.

References

1. L. Farhan, R. Kharel, O. Kaiwartya, M. Quiroz-Castellanos, A. Alissa, and M. Abdulsalam, "A concise review on internet of things (iot) -problems, challenges and opportunities," in *2018 11th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, pp. 1–6, July 2018.
2. S. Shekhar and A. Gokhale, "Dynamic resource management across cloud-edge resources for performance-sensitive applications," in *17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pp. 707–710, May 2017.
3. "Internet of things (iot) and non-iot active device connections worldwide from 2010 to 2025." https://www.statista.com/. Accessed: 2021-07-26.
4. K. Sharma and R. Nandal, "A literature study on machine learning fusion with iot," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1440–1445, April 2019.
5. R. Krishnamoorthi, "Quantizing deep convolutional networks for efficient inference: A whitepaper," *CoRR*, vol. abs/1806.08342, 2018.
6. N. Mitschke, M. Heizmann, K.-H. Noffz, and R. Wittmann, "A fixed-point quantization technique for convolutional neural networks based on weight scaling," in *IEEE International Conference on Image Processing*, pp. 3836–3840, Sep. 2019.
7. R. David, J. Duke, A. Jain, V. J. Reddi, N. Jeffries, J. Li, N. Kreeger, I. Nappier, M. Natraj, S. Regev, R. Rhodes, T. Wang, and P. Warden, "Tensorflow lite micro: Embedded machine learning on tinyml systems," 2021.
8. "Tensorflow Lite." https://www.tensorflow.org/lite. Accessed: 2021-08-01.
9. "uTensor." https://github.com/uTensor/uTensor. Accessed: 2021-08-01.
10. L. Lai, N. Suda, and V. Chandra, "Cmsis-nn: Efficient neural network kernels for arm cortex-m cpus," 2018.
11. Dennis, Don Kurian and Gopinath, Sridhar and Gupta, Chirag and Kumar, Ashish and Kusupati, Aditya and Patil, Shishir G and Simhadri, Harsha Vardhan, "EdgeML: Machine Learning for resource-constrained edge devices."
12. "Eloquent TinyML." https://github.com/eloquentarduino/EloquentTinyML/. Accessed: 2021-07-26.
13. I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. http://www.deeplearningbook.org.
14. E. Alpaydin and C. Kaynak, "Optical Recognition of Handwritten Digits." UCI Machine Learning Repository, 1998.
15. Y. LeCun, C. Cortes, and C. Burges, "Mnist handwritten digit database," *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, vol. 2, 2010.
16. L. Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems*, pp. 396–404, Morgan Kaufmann, 1990.

CACIC 2021

WORKSHOP PROCESAMIENTO DISTRIBUIDO Y PARALELO

COORDINADORES

Marcela Printista(UNSL)
Laura De Giusti (UNLP)
Carlos García Garino (UNCu)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978 -987-633-574-4

Análisis de ejecución múltiple de Funciones Serverless en AWS

Nelson Rodríguez, Hernán Atencio, Martín Gómez, Lorena Parra, Maria Murazzo

Departamento de Informática, Facultad de Ciencias Exactas Físicas y Naturales,
Uniuversidad Nacional de San Juan, San Juan, Argentina
nelson@iinfo.unsj.edu.ar, hernan.atencio.98@gmail.com, martinsj0811@gmail.com,
lorenaparra152@yahoo.com.ar,maritemurazzo@g,mail.com

Resumen. Serverless Computing es una reciente arquitectura para Cloud Computing que presenta ventajas considerables para los usuarios. Sin embargo debido a recientemente aparición, muchas de sus limitaciones o desventajas no están totalmente resueltas. Numerosos desafíos presenta esta arquitectura, que son analizados y estudiados tanto por la academia como por las empresas proveedoras de Cloud. Las plataformas Serverless permiten ejecutar funciones individuales que pueden ser administradas y ejecutadas separadamente. A diferencia de las aplicaciones tradicionales de ejecución prolongada en plataformas dedicadas, virtualizadas o basadas en contenedores, las aplicaciones serverless están diseñadas para crear instancias cuando se les llama, ejecutar una sola función y cerrarse cuando finalizan. La ejecución eficiente y de alta performance es uno de los desafíos a resolver. Aunque las funciones se ejecutan sin estado y escalan bajo demanda, la ejecución múltiple de funciones requiere que varias dificultades sean resueltas, entre ellas la latencia que presentan antes de ser ejecutadas y que impactan en la eficiencia. En el presente trabajo se realiza una serie de pruebas y análisis de los resultados, que permiten emitir conclusiones sobre el impacto que causa la ejecución de múltiples funciones.

Keywords: Serverless Computing, FaaS, Function-as-a-service, Cloud Computing

1 Introducción

La arquitectura serverless es un modelo de computación en el Cloud impulsado por eventos en el que los recursos informáticos se proporcionan como servicios escalables. En el modelo tradicional se cobraba un costo fijo y recurrente por los recursos informáticos del servidor, independientemente de la cantidad de trabajo informático realizado por el servidor. Sin embargo, la implementación de la computación Serverless ha superado esta deficiencia, ya que permite a los clientes pagar solo por el uso del servicio y no se cobran costos ocultos asociados con el tiempo de inactividad.

En este paradigma emergente, las aplicaciones de software se descomponen en múltiples funciones independientes sin estado [1] [2]. Las funciones solo se ejecutan

en respuesta a acciones desencadenantes (como interacciones de usuario, eventos de mensajería o cambios en la base de datos), y se pueden escalar de forma independiente y pueden ser efímeras (pueden durar una invocación) y están completamente administrados por el proveedor de Cloud.

Los principales proveedores de nube han propuesto diferentes plataformas informáticas sin servidor como AWS Lambda, Microsoft Azure Functions y Google Functions. Dichas plataformas facilitan y permiten que los desarrolladores se centren más en la lógica de negocios, sin la sobrecarga de escalar y aprovisionar la infraestructura, ya que el programa se ejecuta técnicamente en servidores externos con el apoyo de proveedores de servicios en el Cloud [8].

No existen muchas definiciones de Serverless. Una de estas fue publicada por Castro p., Ishakian v., Muthusamy v. y Slominski [6], la cual ofrece una descripción de la características: "La informática serverless se puede definir por su nombre, que es pensar (o preocuparse) menos por los servidores. Los desarrolladores no necesitan preocuparse por los detalles de bajo nivel de administración y escalado de servidores, y solo pagan cuando procesan solicitudes o eventos". Luego la define como: "La informática serverless es una plataforma que oculta el uso del servidor a los desarrolladores y ejecuta código que escala bajo demanda automáticamente y facturado solo por el tiempo que se ejecuta el código".

Debido a que la implementación se realiza mediante la plataforma de funciones, la computación serverless, también es denominada función como servicio (FaaS). En este enfoque, casi todas las preocupaciones operativas son abstraídas lejos de los desarrolladores. Los cuales en principio simplemente escriben código e implementan sus funciones en una plataforma sin servidor. La plataforma se encarga de la ejecución de la función, el almacenamiento y la infraestructura de contenedor, redes y tolerancia a fallas. Adicionalmente, también se encarga de escalar las funciones según la demanda real.

En la mayoría de los casos, se pueden escribir funciones en el lenguaje que el programador considere más adecuado (Node.js, Python, Go, Java y más) y utilizar herramientas de contenedor y serverlessr, como AWS SAM o la CLI de Docker, para compilar, probar e implementar las funciones.

El tamaño del mercado global de Serverless Computing se valoró en $ 3.1 millones en 2017 y se proyecta que alcance casi $ 22 millones para 2025, según un informe de Investigación y Mercados que pronostica hacia dónde se dirige la arquitectura Serverless, para 2025 [16]

Un modelo basado en funciones es particularmente adecuado para ráfagas, uso de CPU intensivo, cargas de trabajo granulares. Actualmente, los casos de uso de FaaS varían ampliamente, incluido el procesamiento de datos, el procesamiento de flujo, la computación de borde (IoT) y la computación científica [3]. Con la continua experimentación generalizada en torno a FaaS, es probable que otros casos de uso surjan en un futuro cercano.

AWS Lambda, popularizó en 2014 el concepto de informática serverless, que permiten a los desarrolladores escribir un fragmento de código que realiza una determinada tarea, ejecutarlo en el Cloud y no preocuparse por administrar la infraestructura subyacente.

AWS Lambda de Amazon [1] fue la primera plataforma serverless y definió varias dimensiones clave que incluyen costo, modelo de programación, implementación,

límites de recursos, seguridad y supervisión. Los lenguajes soportados incluyen Node.js, Java, Python y C#. y se pueden usar herramientas de contenedor y sserverless como AWS SAM o la CLI de Docker, para compilar, probar e implementar las funciones.

En sus orígenes, en Lambda cada fragmento de código suele realizar una única tarea. Por eso en 2017, AWS lanzó Step Functions, el servicio del que forma parte el nuevo Workflow Studio.

Con Step Functions, los desarrolladores pueden combinar varios fragmentos de código Lambda en flujos de trabajo que realizan varias tareas y no solo una. Estos flujos de trabajo, a su vez, pueden ser utilizados por los desarrolladores para crear aplicaciones empresariales complejas, como servicios de procesamiento de pagos y herramientas de análisis. Sin embargo también es posible integrar funciones escritas en diferentes lenguajes de programación si utilizar estos servicios especializados [15].

Serverless cubre una amplia gama de tecnologías, que se pueden agrupar en dos categorías: Backend-as-a-Service (BaaS) y Functions-as-a-Service (FaaS).

Backend-as-a-Service permite reemplazar los componentes del lado del servidor con servicios listos para usar. BaaS permite a los desarrolladores externalizar todos los aspectos detrás de una escena de una aplicación para que los desarrolladores puedan elegir escribir y mantener toda la lógica de la aplicación en el frontend. Algunos ejemplos son los sistemas de autenticación remota, la administración de bases de datos, el almacenamiento en el cloud y el hosting.

Función como servicio es un entorno en el que es posible ejecutar software. Las aplicaciones serverless son sistemas basados en el Cloud impulsados por eventos donde el desarrollo de aplicaciones se basa únicamente en una combinación de servicios de terceros, lógica del lado cliente y llamadas a procedimientos remotos hospedados en el Cloud. [14]

Existen diversos desafíos, oportunidades y problemas a resolver, entre ellos la experiencia del desarrollador [17], Interoperabilidad, testing, composición de funciones, seguridad, administración del ciclo de vida, administración de requerimientos no funcionales, performance, optimización del overhead, ingeniería para costo-performance, entre otros [9]

Un diferenciador clave de serverless es la capacidad de escalar desde cero, o no cobrar a los clientes por el tiempo de inactividad. Escalar a cero, sin embargo, conduce al problema de los arranques en frío y el pago de la penalización de obtener código serverless listo para ejecutarse.

El inicio en frío se trata del retraso entre la ejecución de una función después de que alguien la invoque. Se trata de la función en el momento de la invocación. En background, FaaS utiliza contenedores para encapsular y ejecutar las funciones. Cuando un usuario invoca una función, FaaS mantiene el contenedor en ejecución durante un período de tiempo determinado después de la ejecución de la función (caliente) y si otra solicitud entra antes del apagado, la solicitud se sirve instantáneamente. El inicio en frío es aproximadamente el tiempo que se tarda en abrir una nueva instancia de contenedor cuando no hay contenedores en caliente disponibles para la solicitud.

También es importante entender que el bajo costo del servicio se debe a que los proveedores de FaaS no necesitan ejecutar la infraestructura en previsión del uso y pueden cerrar los recursos no utilizados.

Algunos usuarios que eligen FaaS, están aprovechando las mejoras de las plataformas, por ejemplo IBM está utilizando contenedores para reducir los arranques en frío y plataformas como OpenFaaS dan a los usuarios control sobre cómo quieren utilizar los recursos.

Aunque las plataformas serverless existentes funcionan bien para aplicaciones simples, no son adecuadas para servicios más complejos, especialmente cuando la lógica de la aplicación sigue una ruta de ejecución que abarca varias funciones [10].

2 Trabajos relacionados

No existen gran cantidad de trabajos que analicen la ejecución de funciones Serverless, con el objetivo de encontrar el modo más adecuado o eficiente de ejecutarlas. La mayoría de los trabajos muestras resultados sobre el arranque en frío y aspectos relacionados y estudiados desde diferentes ópticas.

El trabajo de Johannes Manner et al [12], presenta puntos de referencia económicos y también orientados al rendimiento. Compara las plataformas AWS y Azure, y los lenguajes Java y Javascript. Analizar los factores que influyen en la duración percibida del arranque en frío mediante la realización de un banco de pruebas en AWS Lambda y Microsoft Azure Functions con 49 500 ejecuciones de funciones Cloud. Su aporte más importante es con referencia a aspectos económicos y no hace propuestas de mejoras para el arranque en frío, sino que compara el comportamiento de las plataformas.

En el trabajo de David Bermbach et al [5], presenta tres enfoques (ingenuo, el extendido y la aproximación global), que reducen el número de arranques en frío durante el tratamiento del servicio FaaS como una caja negra, implementado como parte de un middleware coreográfico liviano, utilizando composición de funciones y por lo tanto, el aprovisionamiento de nuevos contenedores antes de que el proceso de aplicación invoque la función respectiva. Las pruebas las realizan sobre AWS Lambda y OpenWhisk, con un solo lenguaje de programación Node.js.

En la publicación, de Priscilla Benedetti et al [4], se explora la conveniencia de los modos de arranque en caliente y en frío para implementar aplicaciones de IoT, teniendo en cuenta un banco de pruebas de bajos recursos comparable a un nodo en el extremo (Edge). Modelando la implementación y el análisis experimental de una plataforma serverless que incluye elementos de servicio de IoT típicos. Presenta un estudio de rendimiento en términos de consumo de recursos y latencia y para realizar las pruebas utiliza OpenFaaS, un frameworks FaaS de código abierto que permite probar una implementación de arranque en frío con una configuración precisa del tiempo de inactividad gracias a su flexibilidad.

En la publicación [7], los autores describen los principales problemas que afectan al rendimiento de las plataformas serverless y presentan algunos resultados experimentales. Esta investigación hace uso de funciones disponibles comercialmente para tres estudios de casos específicos, siendo: el entrenamiento de modelos basado en aprendizaje automático, las predicciones en vivo y la clasificación de entradas de procesamiento por lotes. Los experimentos presentados son interesantes, pero no se analiza profundidad el problema del arranque en frío.

En el trabajo [13] se describen seis malas prácticas que han sido identificadas, y propone soluciones para tratar de superarlas. Las mismas son: las llamadas asincrónicas (que pueden incrementar la complejidad y requiere un canal de respuesta alternativo), las funciones que llaman a otras funciones (que causa una depuración compleja, y puede llevar un costo extra), el código compartido entre funciones (se podrían interrumpir las funciones serverless existentes que dependen del código compartido si éste cambia), el uso de demasiadas librerías (dado que se aumenta el espacio destinado a las mismas), adopción de demasiadas tecnologías como bibliotecas, frameworks, lenguajes (incrementa la complejidad del mantenimiento y aumenta los requisitos de conocimientos de los integrantes del proyecto) y demasiadas funciones que no son reusadas (causa menor mantenibilidad y menor comprensión del sistema).

En la publicación de Jiang, Choon Lee y Zomaya [11], aparecen resultados interesantes, a pesar de que dicho trabajo se enfoca en la ejecución de worksflow científico. Demuestran que FaaS ofrece un entorno de ejecución ideal para flujo de trabajo aplicado a las ciencias, con su mecanismo dinámico de asignación de recursos y un modelo de precios conveniente y además que AWS Lambda ofrece un entorno de ejecución ideal para aplicaciones de flujo de trabajo científicas con restricciones de precedencia complejas.

Un trabajo muy interesante por la descripción de los patrones de acceso de las funciones serverless desde IoT al Cloud fue desarrollado por [9]. Proponen el uso de WebAssembly como un método alternativo para ejecutar aplicaciones serverless y demuestran cómo una plataforma basada en WebAssembly proporciona muchas de las mismas garantías de aislamiento y rendimiento de las plataformas basadas en contenedores, al tiempo que reduce los tiempos medios de inicio de las aplicaciones y los recursos necesarios para hospedarlas.

3 Ejecución múltiple

Las funciones sin servidor no suelen implementarse de forma aislada. En su lugar, se desencadenan para realizar una tarea en respuesta a una acción o un evento. En una arquitectura distribuida típica, sirven como “pegamento” entre los diferentes componentes de la aplicación: el origen de un componente y el destino de otro. Como tal, es importante establecer el ámbito de los permisos asociados con una función siguiendo el "principio de privilegios mínimos".

La mayoría de los proveedores de FaaS tienen arranques en frío de 1 a 3 segundos y esto afecta a ciertos tipos de aplicaciones donde esta latencia tendrá un impacto dramático. El arranque en frío varía según el proveedor de la nube y los lenguajes de programación. Aunque tiene casi un año de antigüedad, este estudio de referencia muestra el impacto de la latencia de arranque en frío en varias ofertas de FaaS.

El objetivo de este trabajo es analizar cuál es la estrategia más adecuada para la ejecución múltiple de funciones y además cómo impacta el arranque en frio cuando se desea ejecutar múltiples funciones en AWS Lambda y si existen estrategias que puedan minimizar su impacto.

ISBN 978 -987-633-574-4

4 Desarrollo de la pruebas

Para realizar las pruebas se utilizó el servicio AWS Lambda. Este es un servicio informático serverless que permite ejecutar código sin aprovisionar ni administrar servidores, crear una lógica de escalado de clústeres basada en la carga de trabajo, mantener integraciones de eventos o administrar tiempos de ejecución. Es un servicio flexible y rentable que le permite implementar la funcionalidad de back-end en un entorno serverless [2].

Se tuvieron en cuenta los siguientes escenarios:

a) la ejecución Se realizaron las pruebas utilizando una función que utiliza hilos de ejecución escrita en Python, en la cual por cada hilo se invoca una sola función con 20, 50 y 100 invocaciones.

b) Se procedió a realizar las pruebas de ejecución múltiple con la función Lambda Invoke perteneciente a Lambda, la cual permite que una función Lambda llame a otra función Lambda, Se realizó con 20, 50 y 100 invocaciones.

La Función Invoke, invoca una función de Lambda. Puede invocar una función de forma sincrónica (y esperar la respuesta) o de forma asincrónica

Cabe aclarar que antes de poder ejecutar este tipo de funciones, se deben de definir un conjunto de permisos en los roles de ejecución que permita este tipo de invocación.

Se trató de encontrar si existía algún impacto del arranque en frío y por otro lado si la funcionalidad provista por Lambda (en este caso invoke) mejora la performance de la ejecución de instrucciones.

5 Resultados obtenidos

Para el caso de ejecución con hilos. Al utilizar 20 hilos un resultado de 2740,77ms en tiempo de ejecución, esto se muestra en la figura 1.

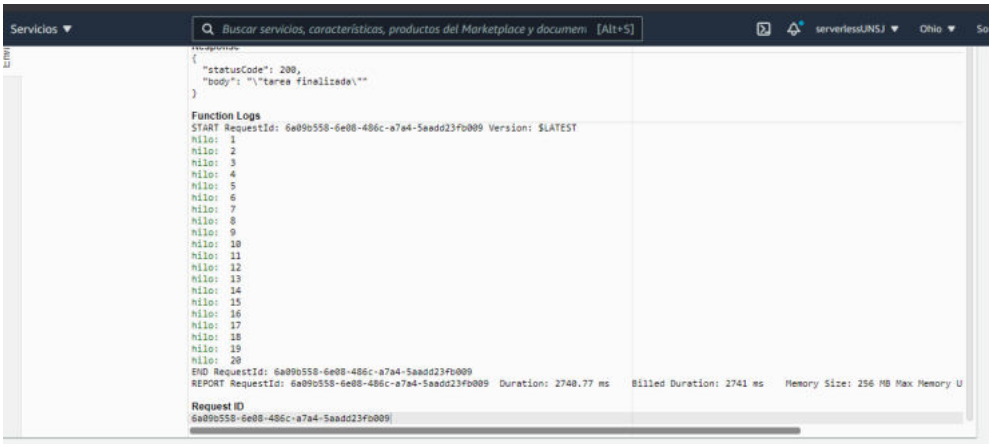


Figura 1

ISBN 978 -987-633-574-4

A continuación se probó invocando 50 funciones, donde se obtuvo un tiempo de 6112,13 ms. El tiempo obtenido es casi el doble al de la prueba anterior. Dichos resultados se muestran en la figura 2.

Figura 2

Para 100 funciones
Finalmente con 100 invocaciones de la función el resultado obtenido es 11594.42 ms, casi duplicando el tiempo obtenido anteriormente.
Como conclusión se puede afirmar que el tiempo de ejecución es proporcional a la cantidad de ejecuciones de la función, es decir, se puede equiparar a un orden de ejecución lineal, y no se observa impacto del arranque en frío
Los resultados se muestran en la figura 3

Figura 3

Para el caso b) (utilizando la función invoke):

ISBN 978 -987-633-574-4

Se partió de realizar la prueba para 20 funciones, la cual retorna un resultado de 1061,47ms en el tiempo de ejecución. Los resultados se muestran en la figura 4.

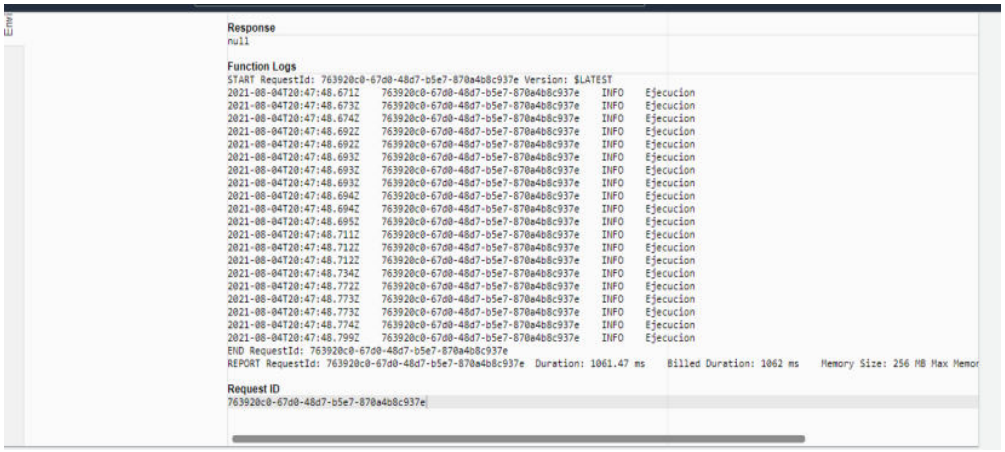


Figura 4

Luego se pasó a probar para 50 funciones, donde el tiempo aumentó proporcionalmente respecto a la prueba anterior, con un tiempo de ejecución de 1949,44 ms. Dicha ejecución se muestra en la figura 5.

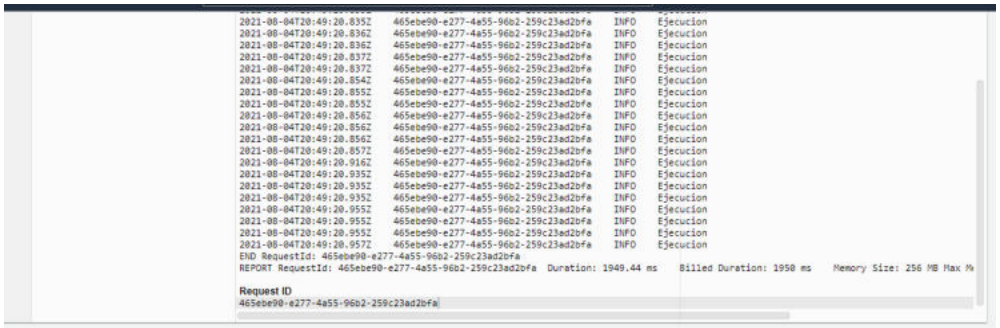


Figura 5

Por último se probó para 100 funciones, en este caso los resultados no fueron como se esperaba, ya que hubo una notable mejoría respecto del tiempo de ejecución de las anteriores pruebas, con un tiempo de ejecución de 2338,7 ms. La ejecución se muestra en la figura 6.

ISBN 978 -987-633-574-4

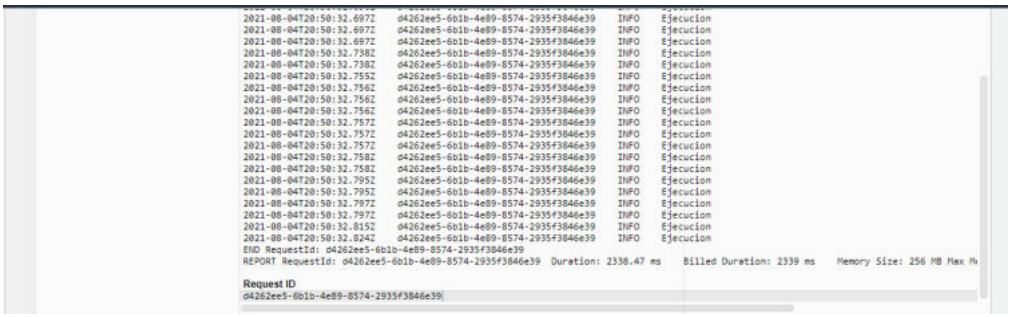


Figura 6

7 Conclusiones y Futuros trabajos

En función de los resultados obtenidos luego de realizar las seis pruebas, se puede concluir lo siguiente:

En ninguna de las pruebas se puede apreciar el impacto del arranque en frío, por lo tanto se considera que AWS provee recursos que minimizan este problema o al menos en estas pruebas no pudo apreciarse.

Por otro lado se debe considerar que Lambda Invoke es una buena alternativa a medida que se escala con la cantidad de funciones, ya que reduce la complejidad temporal lineal que se ve en la implementación con hilos.

Cabe aclarar además que los tiempos de ejecución utilizando Lambda Invoke han sido bastante menores respecto de la implementación con hilos, lo cual lo vuelve una propuesta tentadora a la hora de realizar ejecución múltiple de funciones.

Como trabajos futuros propuesto está el de aplicar concurrencia y paralelismo en la ejecución de funciones ya sea con Step Function o con alguna otra estrategia, evaluando si se mantiene la proporcionalidad en los tiempos de ejecución a medida que se escala en cantidad de funciones ejecutadas. Por otro lado, además se deberá realizar el análisis económico de la ejecución paralela (que dispondrá de más recursos por el escalado automático) que seguramente será más costoso.

También es de interés del grupo de investigación como continuación del presente trabajo, analizar los diferentes tipos de patrones de acceso de funciones serverless desde IoT, como son un solo cliente y múltiple acceso, múltiples clientes y un solo acceso y múltiples clientes, múltiples accesos.

Referencias

1. Adzic, G., Chatley, R.: Serverless computing: economic and architectural impact. In Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering (pp. 884-889). ACM.(2917)
2. AWS Lambda: aws.amazon.com/es/lambda/

3. Baldini, I., Castro, P., Chang, K., Cheng, P., Fink, S., Ishakian, V. & Suter, P.: Serverless computing: Current trends and open problems. In Research Advances in Cloud Computing (pp. 1-20). Springer, Singapore (2017).
4. Benedetti P. et al.: Experimental Analysis of the Application of Serverless Computing to IoT Platforms. Sensors (Basel, Switzerland) vol. 21,3 928. 30 Jan. 2021, doi:10.3390/s21030928
5. Bermbach D., Karakaya A., Buchholz S.: Using Application Knowledge to Reduce Cold Starts in FaaS Services. In: SAC '20, March 30-April 3, 2020, Brno, Czech Republic (2020).
6. Castro p., Ishakian v., Muthusamy v., Slominski a.: The rise of serverless computing. In: Communications of the ACM | Dec. 2019 | VOL. 62 | NO. 12 (2019).
7. Ekin Akkus I., Chen R., Rimac I., Stein M., Satzke K., Beck A., Aditya P., Hilt V.: SAND: Towards High-Performance Serverless Computing. In: 2018 USENIX Annual Technical Conference (2018).
8. Gottlieb, N. : State of the Serverless Community Survey Results.(2016) https: //serverless.com/blog/state-of-serverless-community/. (2016).
9. Hall A., Ramachandran U.: An Execution Model for Serverless Functions at the Edge. In: International Conference on Internet-of-Things Design and Implementation (IoTDI '19), April 15–18, 2019, Montreal,QC, Canada. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/ (2019).
10. Hellerstein, J.M.; Faleiro, J.M.; Gonzalez, J.E.; Schleier-Smith, J.; Sreekanti, V.; Tumanov, A.; Wu, C. Serverless Computing: One Step Forward, Two Steps Back. arXiv 2018, arXiv:1812.03651.
11. Jiang Q., Choon Lee, Zomaya A.: Serverless Execution of Scientific Workflows. In: Springer International Publishing. M. Maximilien et al. (Eds.): ICSOC 2017, LNCS 10601, pp. 706–721, 2017. https://doi.org/10.1007/978-3-319-69035-3_51 (2017).
12. Manner J., Endreb M., Heckel T., Wirtz G.: Cold Start Influencing Factors in Function as a Service. In: 2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (2018).
13. Nupponen Gofore J., Taibi D.: Serverless: What it Is, What to Do and What Not to Do. In: 2020 IEEE International Conference on Software Architecture Companion (ICSA-C) (2020).
14. Roberts M.: Serverless Architectures. https://martinfowler.com/articles/serverless.html (2016).
15. Rodríguez N., Atencio H. et al: Interoperabilidad de funciones en el Modelo de Programación de Serverless Computing. IV CICCSI. Universidad Champagnat (2020).
16. Serverless Architecture Market by Deployment Model, Application, Organization Size, and Industry Vertical. https://www.researchandmarkets.com/reports/4828585
17. Van Eyk1 E, Iosup A, Seif S., Thömmes M.: The SPEC Cloud Group's Research Vision on FaaS and Serverless Architectures. In: Proceedings of WoSC'17, Las Vegas, NV, USA, 4 pages. https://doi.org/10.475/123. (2017).

ISBN 978 -987-633-574-4

Acelerando Código Científico en Python usando Numba

Andrés Milla[1] and Enzo Rucci[2]

[1] Facultad de Informática, UNLP. La Plata (1900), Bs As, Argentina
andressmilla@gmail.com
[2] III-LIDI, Facultad de Informática, UNLP – CIC. La Plata (1900), Bs As, Argentina
erucci@lidi.info.unlp.edu.ar

Resumen En la actualidad, Python es uno de los lenguajes más utilizados en diversas áreas de aplicación. Una de ellas es el ámbito científico, donde resulta habitual la existencia de algoritmos numéricos que requieren un gran costo computacional. Sin embargo, Python presenta limitaciones a la hora de poder paralelizar esta clase de código. Para solucionar esta problemática surge Numba, un compilador JIT que traduce Python en código de máquina optimizado a través de LLVM. Esta herramienta cuenta con primitivas para paralelizar algoritmos, autovectorización mediante instrucciones SIMD, entre otras características. En este estudio, se analizan algunas capacidades y limitaciones de Numba para acelerar algoritmos numéricos, utilizando como caso de estudio *N-Body*, un problema popular en simulación y con alta demanda computacional. Partiendo desde una implementación base desarrollada en Python con NumPy, se muestra como la integración de diferentes opciones de Numba la mejoran hasta 687×, presentando rendimientos cercanos a una implementación de C+OpenMP en una arquitectura multicore Intel de 56 núcleos.

Palabras claves: Python · Numba · N-body · HPC · Multi-hilado

1. Introducción

Desde su surgimiento a comienzos de la década del 90, Python se ha convertido en uno de los lenguajes más populares en la actualidad. De acuerdo con el índice TIOBE, Python se ubica en la segunda posición de los lenguajes más populares en el 2021 [17].

Python es un lenguaje de programación de alto nivel, interpretado, interactivo, dinámico y multi-paradigma [3]. Su notable poder de programación se debe a su sintaxis limpia y clara, la cual provoca que el esfuerzo de programación sea menor comparado con otros lenguajes [7,6]. Entre las áreas de aplicación, se pueden mencionar desarrollo web, educación, aplicaciones de escritorio, desarrollo de videojuegos, inteligencia artificial, *web scraping*, procesamiento de imágenes, entre otras [8,2].

La computación científica es otra área donde Python es muy usado, en parte debido a la existencia de un diverso ecosistema formado por herramientas y

ISBN 978 -987-633-574-4

librerías tanto de propósito general como específico [16]. Aun así, Python es considerado "lento" en comparación a lenguajes compilados como C, C++ y Fortran, especialmente para aplicaciones de cómputo numérico. Entre las causas de su pobre rendimiento, se encuentran su naturaleza de lenguaje interpretado y sus limitaciones al momento de implementar soluciones multi-hiladas [13]. En particular, el principal problema es la utilización de un componente llamado *Global Interpreter Lock* (GIL), el cual permite que único un hilo se ejecute a la vez. Es por lo que la comunidad de Python ha desarrollado varias propuestas que buscan superar esta limitación y así mejorar el rendimiento.

Entre estas soluciones, se encuentra Numba, un compilador *Just-In-Time* (JIT) que traduce Python en código de máquina optimizado [4]. Numba utiliza decoradores [1] para intervenir lo menos posible en el código del programador y, de acuerdo con su documentación, es capaz de acercarse a los rendimientos de C, C++ y Fortran.

En este artículo se propone evaluar y verificar las prestaciones de Numba en el ámbito de la computación científica. Como caso de estudio, se selecciona la simulación de N cuerpos computacionales (*N-Body*), un problema *cpu-bound* que resulta popular en la comunidad científica. Mediante este estudio se espera contribuir con la comunidad Python al explorar algunas de las capacidades y limitaciones de Numba para implementar aplicaciones paralelas sobre arquitecturas CPU multicore.

El resto del artículo se organiza de la siguiente forma. La Sección 2 introduce a Numba mientras que la Sección 3 describe al problema N-Body. Luego, la Sección 4 detalla las implementaciones realizadas. A continuación, la Sección 5 analiza los resultados experimentales mientras que la Sección 6 discute trabajos relacionados. Finalmente, la Sección 7 resume las conclusiones junto al trabajo futuro.

2. Numba

Numba es un compilador JIT que permite traducir código Python a código de máquina optimizado a través de LLVM [3]. De acuerdo con su documentación, es capaz de alcanzar aceleraciones similares a las de lenguajes compilados como C, C++ y Fortran [4], sin necesidad de re-escribir su código gracias a un enfoque de anotaciones llamados decoradores [1].

Compilación JIT. La librería ofrece dos modos de compilación: (1) modo objeto, el cual permite compilar código que haga uso de objetos; (2) modo *nopython*, que le permite a Numba generar código evitando a la API de

```
1  from numba import njit
2
3  # Equivalente a indicar
4  # @jit(nopython=True)
5  @njit
6  def f(x, y):
7      return x + y
```

Figura 1: Compilación en modo *nopython*.

[3] The LLVM Compiler Infrastructure, `https://llvm.org/`

ISBN 978 -987-633-574-4


```python
from numba import njit, double

@njit(double(double[::1, :],
             double[:, ::1]))
def f(x, y):
    """
    x:  Vector 2D de tipo Double
        organizado por columnas.
    y:  Vector 2D de tipo Double
        organizado por filas.
    Retorna la suma del producto
    entre los vectores x e y.
    """
    return (x * y).sum()
```

Figura 2: Compilación en modo *nopython* con el parámetro `signature`

```python
from numba import njit, double, prange

@njit(double(double[::1]), parallel=True)
def f(x):
    """
    x: Vector 1D.
    Retorna la suma del vector x mediante
    una reducción.
    """
    N = x.shape[0]
    z = 0

    for i in prange(N):
        z += x[i]

    return z
```

Figura 3: Compilación en modo *nopython* con el parámetro `parallel`

CPython. Para indicar dichos modos, se utilizan los decoradores `@jit` y `@njit` (ver Fig. 1), respectivamente [4].

Por defecto, cada función será compilada al momento de ser invocada y se mantendrá en la caché para futuras llamadas. Sin embargo, la inclusión del parámetro `signature` provocará que la función sea compilada al momento de la declaración. Además, también posibilitará indicar los tipos de datos que usará la función y controlar la organización de los datos [4] en memoria (ver Fig. 2).

Multi-hilado. Numba permite activar un sistema de paralelización automática estableciendo el parámetro `parallel=True`, como también indicar una paralelización explícita mediante la función `prange` (ver Fig. 3), la cual distribuye las iteraciones entre los hilos de manera similar a la directiva `parallel for` de OpenMP. Además, también soporta reducciones y se encarga de declarar las variables como privadas a cada hilo si son declaradas dentro del alcance de la zona paralela. Lamentablemente, Numba aún no soporta primitivas que permitan controlar la sincronización de los hilos, como pueden ser semáforos o *locks* [4].

Vectorización. Numba delega en LLVM la autovectorización del código y la generación de instrucciones SIMD, pero le permite al programador controlar ciertos parámetros que podrían influir en esta tarea, como la precisión numérica mediante el argumento `fastmath=True`. También ofrece la posibilidad de utilizar *Intel SVML* en caso de estar disponible en el sistema [4].

Integración con NumPy. Cabe destacar que Numba soporta un gran número de funciones de NumPy, lo cual le permite al programador controlar la organización de memoria de los arreglos y realizar operaciones entre ellos [5,4].

Soporte para GPUs. Además de CPUs, Numba es capaz de aprovechar las capacidades de las GPUs, tanto de NVIDIA como de AMD.

3. N-Body

El problema consiste en simular la evolución de un sistema compuesto por N cuerpos durante una cantidad de tiempo determinada. Dados la masa y el

ISBN 978 -987-633-574-4

estado inicial (velocidad y posición) de cada cuerpo, se simula el movimiento del sistema a través de instantes discretos de tiempo. En cada uno de ellos, todo cuerpo experimenta una aceleración que surge de la atracción gravitacional del resto, lo que afecta a su estado.

La física subyacente es fundamentalmente la mecánica de Newton [18]. La simulación se realiza en 3 dimensiones espaciales y la atracción gravitacional entre dos cuerpos C_1 y C_2 se computa usando la ley de gravitación universal de Newton (ver Ecuación 1), donde F corresponde a la magnitud de la fuerza gravitacional entre los cuerpos; G corresponde a la constante de gravitación universal [4]; m_1 y m_2 corresponden a las masas de los cuerpos C_1 y C_2, respectivamente; y r corresponde a la distancia Euclídea [5] entre los cuerpos C_1 y C_2.

Cuando N es mayor a 2, la fuerza de gravitación sobre un cuerpo, se obtiene con la sumatoria de todas las fuerzas de gravitación ejercidas por los $N-1$ cuerpos restantes. La fuerza de atracción se traduce entonces en una aceleración del cuerpo mediante la aplicación de la segunda ley de Newton, la cual está dada por la Ecuación 2, donde F es el vector fuerza, calculado utilizando la magnitud obtenida con la ecuación de gravitación y la dirección y sentido del vector que va desde el cuerpo afectado hacia el cuerpo que ejerce la atracción.

La aceleración de un cuerpo se puede calcular a partir de la Ecuación 2, dividiendo la fuerza total por su masa. Durante un pequeño intervalo de tiempo dt, la aceleración a_i del cuerpo C_i es aproximadamente constante, por lo que el cambio en velocidad está dado aproximadamente por la Ecuación 3.

El cambio en la posición de un cuerpo es la integral de su velocidad y aceleración sobre el intervalo de tiempo dt, el cual se aproxima a la Ecuación 4. Esta fórmula emplea el esquema de integración Leapfrog [19], en el cual una mitad del cambio de posición emplea la velocidad *vieja* mientras que la otra considera la velocidad *nueva*.

$$F = \frac{G \times m_1 \times m_2}{r^2} \qquad (1) \qquad\qquad F = m \times a \qquad (2)$$

$$dv_i = a_i dt \qquad (3) \qquad\qquad dp_i = v_i dt + \frac{a_i}{2} dt^2 = (v_i + \frac{dv_i}{2}) dt \qquad (4)$$

4. Propuesta

En esta sección se describen las diferentes implementaciones propuestas.

4.1. Implementación Naive

Inicialmente se desarrolló una implementación Python "pura" (denominada *naive*), la cual servirá como referencia para evaluar las mejoras introducidas por el uso de Numba. Esta implementación utiliza las operaciones entre vectores que provee NumPy [5] y su código es el de las líneas 6-25 de la Fig. 4.

[4] Equivalente a $6{,}674 \times 10^{11}$

[5] Se calcula utilizando la fórmula $\sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2)}$, siendo ($x_1$, y_1, z_1) las coordenadas de $C1$ y (x_2, y_2, z_2) las coordenadas de C_2.

ISBN 978 -987-633-574-4


```
1  @njit(
2      void(int32, int32, double[:, ::1], double[::1], double[:, ::1]),
3      fastmath=True,
4      error_model="numpy"
5  )
6  def nbody(N, D, positions, masses, velocities):
7      # Para cada instante discreto de tiempo
8      for _ in range(D):
9          # Para todo cuerpo que experimenta una fuerza
10         for i in range(N):
11             # Distancias de los cuerpos hacia el cuerpo i
12             dpos = positions - positions[i]
13             # Magnitudes de las distancias
14             dsquared = (dpos ** 2.0).sum(axis=1) + SOFT
15             # Factores de masa
16             gm = masses * (masses[i] * GRAVITY)
17             # Ley de atracción gravitacional de Newton
18             d32 = dsquared ** -1.5
19             gm_d32 = (gm * d32).reshape(N, 1)
20             forces = (gm_d32 * dpos).sum(axis=0)
21             # Integración de Verlet
22             acceleration = forces / masses[i]
23             velocities[i] += acceleration * DT / 2.0
24             # Actualización de las posiciones del cuerpo i
25             positions[i] += velocities[i] * DT
```

Figura 4: Código de la implementación *Naive* con la integración de Numba.

4.2. Integración de Numba

La primera versión Numba se obtuvo al incluir un decorador en la implementación *naive* (ver líneas 1-5 de la Fig. 4). Se indicó que el código se compile con precisión relajada mediante el parámetro `fastmath` (línea 3), con el modelo de división de NumPy para evitar la verificación de división por cero (línea 4) [4] y con *Intel SVML*, el cuál es inferido por Numba por estar disponible en el sistema.

4.3. Multihilado

Para introducir paralelismo a nivel de hilos, se utilizó la sentencia `prange`. Para ello, fue necesario antes separar el bucle que itera sobre los cuerpos (línea 10 de la Fig. 4) en dos. El primer bucle se encarga de computar la ley de atracción gravitacional de Newton y la integración de Verlet, mientras que el otro actualiza la posición de los cuerpos.

4.4. Arreglos con tipos de datos simples

Se reemplazaron las operaciones vectoriales de NumPy por operaciones numéricas, y las estructuras bidimensionales fueron sustituidas por unidimensionales para ayudar a Numba a la hora de autovectorizar el código (ver Fig. 5).

4.5. Operaciones matemáticas

Se propone evaluar alternativas para el cálculo del denominador de la ley de atracción universal de Newton por: (1) calcular la potencia positiva y luego

ISBN 978 -987-633-574-4


```
# Para cada instante discreto de tiempo
for _ in range(D):
    # Para todo cuerpo que experimenta una fuerza
    for i in prange(N):
        # Inicialización del vector de fuerza
        forces_x = 0.0
        forces_y = 0.0
        forces_z = 0.0
        # Para todo cuerpo que ejerce una fuerza
        for j in range(N):
            # Distancia hacia el cuerpo i
            dpos_x = positions_x[j] - positions_x[i]
            dpos_y = positions_y[j] - positions_y[i]
            dpos_z = positions_z[j] - positions_z[i]
            # Magnitud de la distancia
            dsquared = ((dpos_x ** 2.0) + (dpos_y ** 2.0) + (dpos_z ** 2.0) + SOFT)
            # Factor de masa
            gm = GRAVITY * masses[j] * masses[i]
            # Ley de atracción gravitacional de Newton
            d32 = dsquared ** -1.5
            forces_x += gm * d32 * dpos_x
            forces_y += gm * d32 * dpos_y
            forces_z += gm * d32 * dpos_z
        # Integración de Verlet: aceleración
        aceleration_x = forces_x / masses[i]
        aceleration_y = forces_y / masses[i]
        aceleration_z = forces_z / masses[i]
        # Integración de Verlet: velocidad
        velocities_x[i] += aceleration_x * DT / 2.0
        velocities_y[i] += aceleration_y * DT / 2.0
        velocities_z[i] += aceleration_z * DT / 2.0
        # Integración de Verlet: posición
        dp_x[i] = velocities_x[i] * DT
        dp_y[i] = velocities_y[i] * DT
        dp_z[i] = velocities_z[i] * DT
    # Actualización de la posición de los cuerpos
    for i in prange(N):
        positions_x[i] += dp_x[i]
        positions_y[i] += dp_y[i]
        positions_z[i] += dp_z[i]
```

Figura 5: Código de la implementación paralela sin operaciones de NumPy.

dividir; y (2) multiplicar por el inverso multiplicativo, calculando la potencia positiva previamente. Adicionalmente, se ponen a prueba las siguientes funciones de potencias: (1) función `pow` del módulo `math` de Python; y (2) función `power` que provee NumPy.

4.6. Vectorización

Como se indicó en la Sección 2, Numba delega la autovectorización en LLVM. Aun así, se indicaron los flags `avx512f, avx512dq, avx512cd, avx512bw, avx512vl` para favorecer el uso de esta clase particular de instrucciones.

4.7. Localidad de datos

Con el fin de mejorar la localidad de los datos, se implementó una versión que itera los cuerpos de a bloques, en forma similar a [9]. Para ello, el bucle de

ISBN 978 -987-633-574-4

la línea 4 de la Fig. 5 iterará sobre bloques de cuerpos, y en otros dos bucles más internos, se calculará la fuerza de atracción gravitacional de Newton y la integración de Verlet, respectivamente.

5. Resultados Experimentales

5.1. Diseño experimental

Todas las pruebas fueron realizadas en un sistema equipado con un 2×Intel Xeon Platinum 8276 de 28 núcleos (2 hilos hw por núcleo) y 256 GB de memoria RAM. El sistema operativo fue Ubuntu 20.04.2 LTS y el intérprete utilizado fue Python v3.8.10 junto con Numba v0.52.0 y NumPy v1.20.1.

Para la evaluación de las implementaciones, se varió la carga de trabajo al usar diferentes números de cuerpos: $N = \{4096, 8192, 16384, 32768, 65536, 131072, 262144, 524288\}$ mientras que el número de pasos de simulación se mantuvo fijo (I=100). Cada optimización propuesta, fue aplicada y evaluada incrementalmente a partir de la versión *naive* [6]. Para la comparación final con la versión C+OpenMP se utilizó la implementación presentada en [15] usando el compilador ICC (versión 19.1.0.166).

5.2. Rendimiento

Para evaluar el rendimiento se emplea la métrica GFLOPS (mil millones de FLOPS), utilizando la fórmula $GFLOPS = \frac{20 \times N^2 \times I}{t \times 10^9}$, donde N es el número de cuerpos, I es el número de pasos, T es el tiempo de ejecución (en segundos) y el factor 20 representa la cantidad de operaciones en punto flotante requerida por cada interacción [7].

En la Fig. 6 se pueden observar los rendimientos al activar las opciones de compilación y aplicar multi-hilado al variar N. Aunque las opciones de compilación de Numba (`njit+fastmath+svml`) no tienen incidencia prácticamente en el rendimiento de esta versión, sí se aprecia una mejora importante al utilizar hilos para computar el problema. En particular, se puede notar una mejora en promedio de 33× y 38× para 56 y 112 hilos, respectivamente.

En la Fig. 7 se puede apreciar la mejora significativa que produce emplear arreglos con tipos de datos simples en lugar de compuestos (un promedio de 41× para el caso de 112 hilos). Si bien el segundo simplifica la codificación, también implica organizar los datos en forma de arreglo de estructuras (en inglés, *array of structures*), lo que impone limitaciones al aprovechamiento de las capacidades SIMD del procesador [11]. Adicionalmente, también se puede notar que el uso de *hyper-threading* reporta una mejora de aproximadamente 78 % en este caso.

De la Fig. 8 se puede observar que prácticamente no hay cambios en el rendimiento por el uso de los diferentes cálculos matemáticos y funciones de potencia que fueron descritos en la sección 4.5. Esto se debe a que, independientemente

[6] Cada versión previa está etiquetada como *Referencia* en todos los gráficos.
[7] Una convención ampliamente aceptada en la literatura para este problema.

ISBN 978 -987-633-574-4

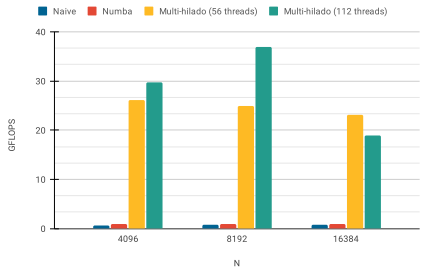


Figura 6: Rendimientos obtenidos para opciones de compilación y multi-hilado al variar N.

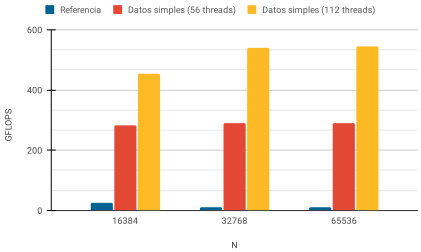


Figura 7: Rendimientos obtenidos de la optimización paralela al variar N.

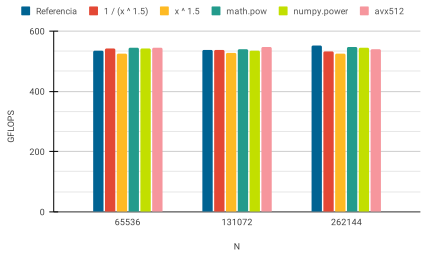


Figura 8: Rendimientos obtenidos utilizando el uso de diferentes cálculos matemáticos, funciones de potencia e instrucciones AVX512 al variar N.

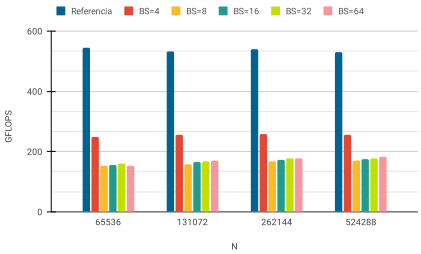


Figura 9: Rendimiento obtenido del procesamiento de a bloques al variar N.

de cuál opción se utilice, el código máquina resultante es siempre el mismo. Un caso similar se da al indicar explícitamente que utilice instrucciones AVX-512. Tal como se mencionó en la sección 2, Numba intenta autovectorizar el código a través de LLVM. Al observar el código máquina, se notó que las instrucciones generadas ya hacían uso de estas extensiones.

El procesamiento por bloques descrito en la sección 4.7 no mejoró el rendimiento de la solución, tal como se puede observar en la Fig. 9. La pérdida de rendimiento se relaciona con que esta reorganización del cómputo produce fallos en LLVM a la hora de autovectorizar. Lamentablemente, debido a que Numba no ofrece primitivas para indicar la utilización de instrucciones SIMD de forma explícita, no hay manera de enmendarlo.

En la Fig. 10 se muestran los rendimientos obtenidos para la relajación de precisión al variar el tipo de dato y la carga de trabajo (N). Se puede observar que el uso del tipo de datos `float32` (en lugar de `float64`) conlleva a una mejora de hasta 2.8× GFLOPS, a costo de una reducción en la precisión del resultado. En forma similar, se puede notar claramente la importante aceleración que

ISBN 978 -987-633-574-4

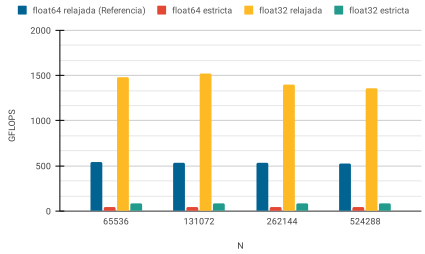


Figura 10: Rendimiento obtenido para la relajación de precisión al variar el tipo de dato y *N*.

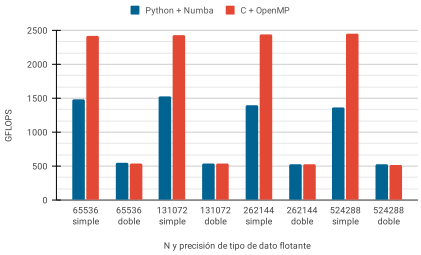


Figura 11: Comparación de rendimiento de la versión final Python+Numba y C+OpenMP variando el tipo de dato y *N*.

produce relajar la precisión en ambos tipos de datos: `float32` (17.2× en promedio) y `float64` (11.4× en promedio). En particular el pico de rendimiento es de 1524/536 GFLOPS en simple/doble precisión. Resulta importante mencionar que la versión final de Numba logra una aceleración de 687× en comparación a la implementación *naive* (caso `float64`).

Por último, en la Fig. 11 se presenta una comparación entre la implementación C+OpenMP y la versión Python+Numba más rápida. Aunque se puede notar prácticamente el mismo rendimiento al utilizar tipos de datos de doble precisión, la versión C muestra una mejora de aproximadamente 1.7× por sobre Python al cambiar a precisión simple.

6. Trabajos Relacionados

A la fecha, existen unos pocos trabajos disponibles en la literatura que exploran las capacidades de Numba. El primero es el que lo introduce propiamente como un compilador JIT capaz de optimizar código Python [12].

Posteriormente, se presentaron dos estudios [10,13] que presentan a Numba como una opción factible para acelerar cómputo numérico, estando sus resultados en línea con los del presente trabajo. En [10] se muestra como una reducción a suma en 2D logra una mejora 200× frente una versión Python pura. Por su parte, en [13] se evalúa el rendimiento de Python junto a Numba, utilizando como caso de estudio el clásico algoritmo de producto de matrices, y luego, se lo compara frente a CUDA, C y cuBLAS GEMM. Como resultado, se obtuvo que Numba aumentó el rendimiento 1415×, mientras que con C se incrementó 2162×.

Por último, un estudio reciente [14] evalúa diferentes opciones para paralelizar código Python y las compara con versiones Fortran. Como caso de estudio, se optó por el problema *Five-point stencil* y, entre los resultados, se pudo observar que Numba logra importantes aceleraciones frente a Python, aunque quedó un poco lejos del rendimiento de Fortran. Si bien en este trabajo se comparó con el lenguaje C, las tendencias encontradas son similares.

ISBN 978 -987-633-574-4

7. Conclusiones y Trabajo Futuro

En este trabajo se evaluaron las capacidades y limitaciones de Numba como optimizador de código Python. Para ello, se utilizó como caso de estudio el problema numérico *N-Body* debido a su gran costo computacional. A una implementación base desarrollada en Python+NumPy, se le integró Numba y se le aplicaron diferentes optimizaciones posibles de manera incremental. En base los resultados experimentales, se puede decir de Numba lo siguiente:

- El multi-hilado fue una técnica efectiva para mejorar el rendimiento y su introducción no requirió cambios significativos en el código.
- Numba no sólo fue capaz de auto-vectorizar el código sino que lo hizo usando las instrucciones SIMD nativas del procesador (no hubo necesidad de que el programador las especifique). Sin embargo, la ausencia de primitivas de vectorización en este traductor (como `simd` en OpenMP) puede ser una limitación cuando la auto-vectorización no es posible, como se evidenció en la versión por bloques.
- El uso de arreglos de tipos de datos simples (en lugar de compuestos) llevó a importantes mejoras de rendimientos a costo de alargar y complicar un poco el código.
- Las diferentes maneras de computar operaciones y funciones matemáticas no tuvieron impacto en las prestaciones, por lo que su elección termina siendo por preferencia.
- Tanto la reducción como la relajación de precisión produjeron aceleraciones significativas en el rendimiento; sin embargo, se debe tener en cuenta que se vio afectada la representación numérica final como contraparte.

Adicionalmente, de la comparación con la versión C+OpenMP, se pudo notar un rendimiento similar al utilizar tipos de datos de doble precisión, mientras que al utilizar simple precisión la versión de C mostró una mejora de 1.7× sobre Python+Numba. Del análisis realizado se concluye que Numba puede ser una opción muy conveniente para acelerar cómputo numérico en Python, especialmente por su bajo costo de programación.

Como trabajos futuros, resulta de interés:

- Continuar explorando otras capacidades y limitaciones de Numba no contempladas en este trabajo, como la utilización de GPUs.
- Dado que existen otras tecnologías que permitan implementar paralelismo en Python, realizar una comparación entre ellas considerando no sólo el rendimiento sino también el costo de programación.

Referencias

1. 7. Decorators — Python Tips 0.1 documentation, `https://book.pythontips.com/en/latest/decorators.html`
2. Applications for Python | Python.org, `https://www.python.org/about/apps/`

ISBN 978 -987-633-574-4

3. General Python FAQ — Python 3.9.5 documentation, https://docs.python.org/3/faq/general.html#what-is-python
4. Numba documentation — Numba 0.53.1-py3.7-linux-x86_64.egg documentation, https://numba.readthedocs.io/en/stable/index.html
5. NumPy, https://numpy.org/
6. Python vs C++ Comparison: Compare Python vs C++ Speed and More, https://www.bitdegree.org/tutorials/python-vs-c-plus-plus/
7. Python vs Java: What's The Difference? – BMC Software | Blogs, https://www.bmc.com/blogs/python-vs-java/
8. Top 12 Fascinating Python Applications in Real-World [2021] | upGrad blog, https://www.upgrad.com/blog/python-applications-in-real-world/
9. Costanzo, M., Rucci, E., Naiouf, M., Giusti, A.D.: Performance vs Programming Effort between Rust and C on Multicore Architectures: Case Study in N-Body. In: 2021 XLVII Latin American Computer Conference (CLEI). p. In press (2021)
10. Crist, J.: Dask amp; numba: Simple libraries for optimizing scientific python code. In: 2016 IEEE International Conference on Big Data (Big Data). pp. 2342–2343 (2016). https://doi.org/10.1109/BigData.2016.7840867
11. Intel Corp.: How to manipulate data structure to optimize memory use on 32-bit intel® architecture (2018), https://tinyurl.com/26h62f76
12. Lam, S.K., Pitrou, A., Seibert, S.: Numba: a LLVM-based Python JIT compiler. In: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. pp. 1–6. Austin, Texas (2015). https://doi.org/10.1145/2833157.2833162
13. Marowka, A.: Python accelerators for high-performance computing. The Journal of Supercomputing **74**(4), 1449–1460 (Apr 2018). https://doi.org/10.1007/s11227-017-2213-5, https://doi.org/10.1007/s11227-017-2213-5
14. Miranda, E.F., Stephany, S.: Common HPC Approaches in Python Evaluated for a Scientific Computing Test Case. REVISTA CEREUS **13**(2), 84–98 (Jul 2021), http://www.ojs.unirg.edu.br/index.php/1/article/view/3408, number: 2
15. Rucci, E., Moreno, E., Pousa, A., Chichizola, F.: Optimization of the n-body simulation on intel's architectures based on avx-512 instruction set. In: Computer Science – CACIC 2019. pp. 37–52. Springer International Publishing (2020)
16. SciPy.org: Scientific computing tools for Python (2021), https://www.scipy.org/about.html
17. TIOBE Software BV: TIOBE Index for August 2021 (08 2021), https://www.tiobe.com/tiobe-index/
18. Tipler, P.: Physics for Scientists and Engineers: Mechanics, Oscillations and Waves, Thermodynamics. Freeman and Co (2004)
19. Young, P.: The leapfrog method and other "symplectic" algorithms for integrating newton's laws of motion. Tech. rep., Physics Department, University of California, USA (4 2014), https://young.physics.ucsc.edu/115/leapfrog.pdf

CACIC 2021

WORKSHOP TECNOLOGIA INFORMATICA APLICADA EN EDUCACION

COORDINADORES

Gladys Gorga (UNLP)
Alejandra Malberti (UNSJ)
Claudia Russo (UNNOBA)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978 -987-633-574-4

Aplicación de Herramienta de Realidad Aumentada para la Enseñanza de Programación en el Nivel Superior

Lucas Romano[1], Ezequiel Moyano[1]

[1] IDEI, Universidad Nacional de Tierra del Fuego, H.Yrigoyen 879, 9410 Ushuaia, Argentina.
{lromano, emoyano}@untdf.edu.ar

Resumen. El presente artículo tiene por objetivo establecer las posibilidades que ofrece la incorporación de tecnologías con Realidad Aumentada a los escenarios educativos. Se realiza un marco teórico acerca de los beneficios de utilizar esta tecnología emergente y se lleva a cabo un análisis de investigaciones que afirman obtener resultados positivos por utilizar Realidad Aumentada en el nivel educativo medio y universitario.

Posteriormente se describe una experiencia propia realizada para una cátedra de la Universidad Nacional de Tierra del Fuego y se exponen resultados obtenidos respecto de variables vinculadas a una mejora académica de los estudiantes.

Palabras Claves: Realidad Aumentada, Innovación Educativa, Educación Secundaria y Universitaria, Tecnología Informática Aplicada a la Educación.

1. Introducción

Reflexionar y rediseñar los procesos de enseñanza y aprendizaje en nuevas propuestas pedagógicas es uno de los mayores desafíos de la educación; el uso de nuevas tecnologías como recursos didácticos, se afianzan cada vez más en el quehacer del aula y en la actividad docente en busca de nuevas formas creativas y colaborativas.

La incorporación de nuevas herramientas digitales en los contenidos curriculares debe permitir acercar a los estudiantes a nuevas posibilidades de construcción del conocimiento, que implique un aprendizaje significativo. En la actualidad una de las tecnologías de mayor impulso e importancia es la Realidad Aumentada.

Hablar de Realidad Aumentada (RA) es referirse a una tecnología que permite percibir e interactuar con el mundo real, creando un escenario real aumentado con información adicional generada por mediaciones pedagógicas y tecnológicas[1]. Estas tecnologías ofrecen una gran variedad de posibilidades educativas que permiten enriquecer el proceso de enseñanza-aprendizaje junto al acceso y uso de contenido multimedial significativo y variado, que sería inaccesible en otras circunstancias.

El alcance de esta tecnología es increíblemente abarcativa, su adaptación a los diferentes niveles educativos (desde los niveles iniciales hasta la educación superior)

ISBN 978 -987-633-574-4

como en personas con capacidades diferentes, hacen de ella uno de los dispositivos didácticos de mayor impacto en los procesos educativos.

2. Utilización de Tecnologías de Realidad Aumentada en los Escenarios Educativos

Son muchas las posibilidades que ofrecen las tecnologías con RA, entre ellas el poder representar e interactuar con objetos virtuales en un espacio tridimensional, potenciando la adquisición de habilidades tales como la capacidad espacial y habilidades prácticas. Como se sostiene en [2] esta característica permite avanzar en el aprendizaje de disciplinas donde los conceptos resultan abstractos o confusos para los estudiantes, bien por su complejidad o bien porque no se pueden concretar completamente en el mundo físico.

La enseñanza de la RA permite aumentar la autonomía de los estudiantes, permitiéndoles llevar su propio ritmo de estudio, y maximizar el tiempo y recursos disponibles [3]. Lo señalado, incentiva el aprendizaje significativo, ya que permite al alumno experimentar y relacionar el contenido nuevo con experiencias y aprendizajes anteriores.

La RA es un recurso tecnológico que potencia la experimentación, la interactividad, las percepciones positivas de los participantes y el trabajo colaborativo, generando con ello un aprendizaje significativo, constructivista y por descubrimiento [4,5].

La utilización de esta herramienta produce en los estudiantes altos niveles de participación y disfrute. El ambiente de aprendizaje es atractivo y estimulante. Los estudiantes que han experimentado con esta herramienta expresan su satisfacción en cuanto al material utilizado, la posibilidad de recibir información en diferentes formatos y la sensación de tener el control de la actividad, ya que pueden explorar los temas en el orden que eligen, e incluso volver a visitar los materiales cuantas veces consideren necesarias [3,6].

El uso de esta tecnología permite la creación de escenarios simulados, lo cual otorga el acceso a técnicas, o herramientas que abordan la práctica para formar profesionales en alguna especialidad concreta. Los usuarios logran adquirir competencias prácticas, mejoran el desarrollo de habilidades y construyen actitudes positivas evitando riesgos físicos [7,8].

Desde un punto de vista tecnológico la Realidad Aumentada compensa algunas de las deficiencias presentes en la educación tales como experimentos o prácticas que no pueden ser realizadas debido a los costes del equipamiento, o a la relación entre el número de equipos disponibles y los estudiantes matriculados; la disponibilidad de las instalaciones, ya sea por espacio y/o por tiempo [2].

Las bondades de la RA favorecen el aprendizaje en los entornos de e-learning, permitiendo a los estudiantes manejar su propio ritmo de aprendizaje. También

permite enriquecer un libro o cualquier material impreso con contenido virtual. Estos nuevos mecanismos de acceso a la información despiertan el interés y la curiosidad de las nuevas generaciones que emplean cada vez más la tecnología [7,9].

Sin embargo, si bien la RA constituye una valiosa herramienta para la educación, lo que favorece la calidad del aprendizaje es su encuadre dentro de un planteamiento pedagógico adecuado [10,11]. Su integración en el aula se debe enmarcar dentro de un proyecto educativo que anteponga lo pedagógico a lo tecnológico [12].

Comienza a ser imprescindible el utilizar metodologías innovadoras, como el aprendizaje basado en problemas, en el descubrimiento, y en el juego; el aprendizaje colaborativo, y otras metodologías que estimulen la creatividad, el aprendizaje autónomo y la adquisición de competencias, dentro de ambientes de aprendizaje activos y constructivistas, donde el alumno sea responsable de la construcción de su conocimiento [13].

3. Antecedentes en el Uso de RA en Escenarios Educativos.

Existen muchas experiencias relacionadas al uso de RA en diversos entornos educativos de todos los niveles, (salones de clases, laboratorios, aulas virtuales, etc.) tanto formales como informales cuyos resultados han sido muy satisfactorios y positivos en los procesos de enseñanza-aprendizaje.

A los efectos de tomar las más representativas, de acuerdo al objetivo del trabajo, en la primera etapa del proyecto se realizó un mapeo sistemático de literatura bajo el protocolo propuesto por Kitchenham [14]; a través del cual se tomaron las más apropiadas a los propósitos del presente trabajo.

Un trabajo denominado ***"Improving stroke education with augmented reality: A randomized control trial"***[5], tuvo por objetivo evaluar la eficacia del uso de RA para impartir una lección sobre fisiología, fisiopatología y anatomía del accidente cerebrovascular. Se llevó a cabo en un ensayo controlado aleatorio para evaluar el impacto de presentar información sobre el accidente cerebrovascular, a través de una aplicación con RA. Se desarrolló en Unity 3D, utilizando codificación C# para elementos interactivos.

Participaron del estudio 101 estudiantes de entre 18 y 25 años de una Universidad de Australia (sin educación formal previa sobre el cerebro o accidentes cerebrovasculares); 51 participantes fueron asignados al azar para utilizar RA y 50 en el grupo de folletos. El grupo de folletos recibió la información en forma de un recurso escrito e impreso con ilustraciones visuales y texto, mientras el grupo de RA se le presentó con imágenes interactivas en 3D del cerebro y un audio con información relacionada.

La experiencia con RA utilizó un cubo impreso en 3D de 6 cm, que contenía un patrón de color abstracto diferente en cada lado, el cual era reconocido por el dispositivo y al colocarlo frente a la cámara, la pantalla reemplazaba al cubo con un

ISBN 978 -987-633-574-4

modelo virtual aumentado del cerebro a medida que avanzaba la narración. Al rotar el cubo el modelo 3D del cerebro rotaba en tiempo real. Por otro lado, el folleto contenía una réplica exacta palabra por palabra de la transcripción de audio utilizada con capturas de pantalla de la aplicación resaltadas como ayudas visuales.

Aunque no hubo una diferencia de aumento específico en los puntajes o el aprendizaje obtenido en los grupos, los participantes que usaron RA manifestaron una satisfacción mucho mayor con el recurso y percibieron una experiencia de aprendizaje mejorada.

En el artículo ***"Motivation and Academic Improvement using Augmented Reality for 3D Architectural Visualization"***[15]**,** el objetivo básico del proyecto fue observar posibles diferencias entre estudiantes de distintas titulaciones (Arquitectura e Ingeniería). La idea principal era diseñar nuevas metodologías para incorporar tecnologías avanzadas en las clases, con el fin de mejorar el rendimiento académico. El estudio involucró cursos académicos 2012-2015 con estudiantes del primer y segundo año de las titulaciones de Ingeniería de Edificación, Arquitectura, Ingeniería Civil y Multimedia.

Los cursos se organizaron con 4 horas de conferencias (2hs cada una) y 3hs adicionales de sesiones prácticas. El objetivo fue proporcionar a los estudiantes habilidades básicas en la interpretación de modelos complejos y reproducción en 2D y 3D, como explorar métodos de visualización interactiva, a través de la publicación en blogs personales y la exhibición de modelos con RA al final del curso.

El instrumento fue diseñado con el objetivo de recoger datos tomando la evaluación heurística y atributos de usabilidad (Nielsen, 1993), la utilidad percibida y facilidad de uso (modelo de Davis, 1989), y cuestionarios de satisfacción de usabilidad (Lewis, 1995) como referencias. Participaron un total de 35 estudiantes del último año.

El trabajo mostró una relación directa entre la motivación de uso y los resultados de la experiencia del usuario, y cómo esa relación afecta el grado de progreso en el uso de una tecnología en particular. Se demostró que el uso de tecnologías, específicamente RA en la docencia, mejoran las capacidades del alumno, especialmente en función de la motivación de este, lo cual se refleja en sus resultados académicos.

Por último, ***"Eficacia del Aprendizaje mediante Flipped Learning con Realidad Aumentada en la Educación Sanitaria Escolar"***[16], tuvo como objetivo conocer la eficacia de una metodología innovadora mediante la combinación del Flipped Learning y la tecnología de RA frente a una tradicional, para el aprendizaje de contenidos asociados con los protocolos de soporte vital básico (SVB) y las pautas recomendadas para la realización de la reanimación cardiopulmonar (RCP).

El trabajo fue un diseño experimental de nivel descriptivo y correlacional, siguiendo un método cuantitativo, en el cual se analizaron y compararon dos grupos de estudiantes (control y experimental). Participaron 60 estudiantes del tercer curso de nivel Secundario, elegidos por medio de un muestreo intencional. Los estudiantes fueron matriculados en dos grupos dentro del mismo nivel académico.

ISBN 978 -987-633-574-4

La información se obtuvo por medio de un cuestionario ad hoc, confeccionado en base a las exigencias y requerimientos del estudio; compuesto por 44 preguntas clasificadas en 3 dimensiones (Social, curricular y Grado de aprendizaje). El grupo A recibió la instrucción de los contenidos de forma tradicional, sin la utilización de ningún recurso digital; el grupo B efectuó un proceso de aprendizaje mediado por las TIC, siguiendo un enfoque invertido a través de un Flipped Learning desarrollado fuera del centro educativo de manera ubicua. Las sesiones presenciales de este segundo grupo fueron enriquecidas mediante la tecnología de RA.

Las mejoras conseguidas en las variables presentadas originaron una mayor proyección, productividad y eficacia en la consecución de los objetivos didácticos formulados por los docentes. Es posible concluir que el proceso de aprendizaje mediado por las TIC y en concreto el flipped learning complementado con RA es de gran utilidad, contribuyendo a la obtención de un mayor grado de eficacia en las destrezas y conocimientos alcanzados por los estudiantes.

4. Experiencia Aplicada: “El Asistente Virtual de Cátedra”

El presente artículo refleja el diseño metodológico planteado en la investigación a través de una experiencia áulica con grupos contrastados de estudiantes en una cátedra (Expresión de problemas y algoritmos) del primer año de la carrera de Licenciatura en Sistemas de la Universidad Nacional de Tierra del Fuego.

La experiencia se realizó sobre una muestra de 63 alumnos. El espacio curricular está dividido en dos comisiones, por lo cual se procedió a implementar material con tecnologías de RA sobre una de ellas, denominada grupo-experiencia, denominando a la otra grupo-control, sobre la cual se trabajó el material tradicional.

En base a lo manifestado se diseñó un prototipo de material que utiliza RA sobre una de las unidades de la currícula. Con el fin de realizar la experiencia se utilizó una aplicación con desarrollo de proyectos aumentados ad hoc, como es el caso de la aplicación gratuita denominada “Augmented Class”, diseñada para realizar proyectos educativos con Realidad Aumentada.

A partir de la aplicación se crearon los proyectos para la experiencia con escenas aumentadas, que luego fueron compartidas a los estudiantes a los efectos de ser descargadas en sus dispositivos e importadas desde la misma aplicación.

La propuesta consistió en crear un robot como un personaje 3D animado que acompañe y asista a los estudiantes sobre la temática a desarrollar en la cursada, y a la cual los estudiantes accedían mediante tecnologías con RA.

A diferencia de como venían trabajando los estudiantes en las clases en relación al robot (visualmente se lo representaba mediante un punto), el mismo tomó una forma física animada, con características de humanoide, es decir contenía cuerpo, brazos, piernas y realizar movimientos. Se lo presentó como el “nuevo asistente virtual de la

ISBN 978 -987-633-574-4

cátedra", para acompañar a los estudiantes para realizar sus ejercicios desde la virtualidad.

Se diseñó y entregó al grupo de la experiencia una guía aumentada de ejercicios prácticos a través de tecnologías con RA, en la cual el personaje/asistente creado aparecía y brindaba ayuda a los estudiantes acerca de cómo debía realizarse cada uno de los ejercicios propuestos. Los ejercicios tenían asociada una imagen que los estudiantes debían utilizar como trigger para reproducir la escena de RA, donde aparecía el asistente animado en 3D para asistir al estudiante y brindar los tips y el feedback adecuado para resolver correctamente el ejercicio.

Se utilizaron dos cuestionarios como instrumentos de recolección de datos, que consistieron en una evaluación diagnóstica inicial y, posteriormente, un postest con un instrumento estandarizado denominado "Instructional Materials Motivation Survey" [17] para obtener resultados respecto de asociaciones entre ciertas variables educativas específicas; a su vez se realizó una entrevista a cada uno de los docentes que intervinieron en la experiencia; una observación participante; y el parcial de la asignatura, el cual consistió en una prueba de elección múltiple para un análisis de rendimiento posterior alcanzado por los estudiantes.

5 Resultados

Si bien la experiencia forma parte de un proyecto de investigación que se encuentra actualmente en curso, la misma permite tener una concepción real de la aplicación y utilización de herramientas con tecnologías de RA en una clase del nivel superior.

Del análisis comparativo respecto a la implementación de materiales con tecnologías de RA en una de las comisiones ,respecto de la otra donde se trabajó con el material tradicional, los principales resultados preliminares obtenidos pueden ser resumidos desde varias aristas.

Un primer análisis consistió en comparar el porcentaje de eficiencia respecto de los resultados sobre una sección del examen parcial en cada comisión. Dicha sección del parcial se correspondía con la unidad de la cátedra que fue implementada con material aumentado para el grupo de la experiencia, y con material tradicional para el grupo de control. Esta sección consistió en una serie de preguntas con opción de respuesta múltiple.

Los resultados muestran (ver figura 1) una mejora en el porcentaje de calificaciones en base a respuestas correctas para el grupo de la experiencia (90,2%) frente al grupo de control (84,3%).

ISBN 978 -987-633-574-4

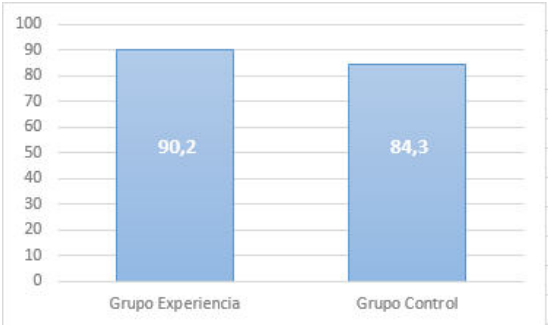


Figura 1- Porcentaje de eficiencia en cada comisión.

Tomando en cuenta al grupo que trabajó con la herramienta de RA se aplicó el instrumento estandarizado de recogida de datos para evaluar varios tópicos acerca de su implementación.

En primer lugar se consideró realizar un análisis respecto al grado de aceptación del uso de la misma (ver figura 2). En líneas generales los estudiantes recibieron y trabajaron con el material con RA de una forma muy llevadera y natural, manifestando su curiosidad, y luego interés en los contenidos aumentados que ampliaban los conocimientos que ya tenían. Calificaron el material como llamativo, al compararlo con otros materiales que venían trabajando durante las unidades anteriores. A efectos de mejor interpretación del gráfico estaba la posibilidad de seleccionar como opciones: Bastante en desacuerdo y completamente en desacuerdo, las cuales no tuvieron ningún porcentaje.

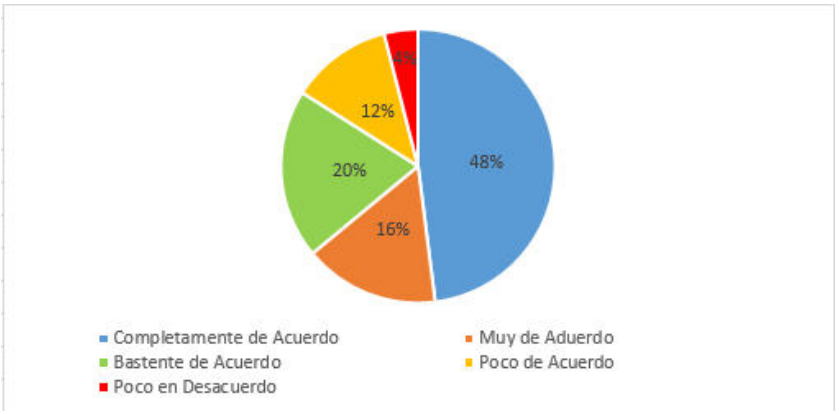


Figura 2- Aceptación del uso de la herramienta

Siguiendo el análisis anterior, se puede observar que casi a la totalidad de los estudiantes les llamó la atención el trabajar con tecnologías de RA. Al mismo tiempo, se puede destacar un alto porcentaje de estudiantes que encontraron interesante el

ISBN 978 -987-633-574-4

material con RA. Muchos de ellos expresaron tener una sensación de satisfacción de logro al haber realizado completamente los ejercicios con el material aumentado.

Un porcentaje muy alto de los estudiantes manifestó haber disfrutado el aprender los contenidos de la unidad de la cátedra con esta herramienta durante las clases en las que se expusieron los temas con el material aumentado.

A su vez, los estudiantes expresaron su seguridad de que los contenidos aprendidos en las lecciones serían de utilidad en el futuro. La figura 3 visualiza los principales tópicos observados por los estudiantes respecto a la experiencia, para lo cual se tomó como referencia aquellos que identificaron sentirse cómodos (seleccionaron completamente de acuerdo, muy de acuerdo o bastante de acuerdo), incómodos (completamente o bastante o poco en desacuerdo) y quienes no expresaron su opinión.

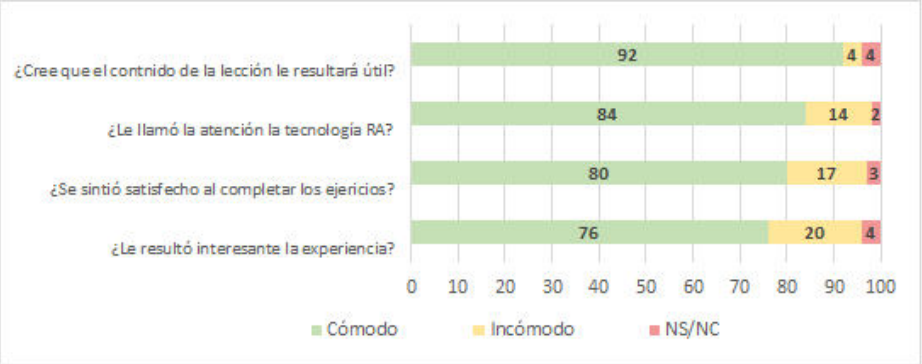


Figura 3- Porcentaje de aspectos generales de la herramienta contestado por los estudiantes.

Por último se consideró el punto de vista de los docentes respecto a su perspectiva de aplicar la herramienta en la cátedra, como el resultado de la experiencia y que fortalezas y debilidades se detectaron.

Se describió a la experiencia en su totalidad como muy interactiva, enriquecedora y positiva. Los docentes involucrados consideran que la experiencia tiene un alto potencial, no sólo durante la clase, sino fuera de ella. El hecho de que ante la necesidad de un estudiante de resolver un ejercicio y no saber cómo encararlo o avanzar en su desarrollo, exista la posibilidad de contar con un asistente virtual (robot) a través de un material creado con tecnologías con RA con el fin de ayudarles a resolverlo es altamente gratificante.

De esta manera, los estudiantes pueden avanzar de forma autónoma, y al mismo tiempo alivia la carga al docente. Esto es de gran utilidad especialmente en cátedras donde la matrícula de alumnos es alta, y el equipo docente escaso, lo cual en ocasiones, imposibilita cubrir cabalmente cada duda o consulta que surge.

Los resultados expresados respecto a la experiencia no hacen más que confirmar las reflexiones vertidas por el cuerpo docente de la cátedra.

ISBN 978 -987-633-574-4

6 Discusión y Conclusiones

Este trabajo pretende agregar nueva evidencia empírica al conjunto de investigaciones relacionadas con los efectos positivos de utilizar tecnologías de Realidad Aumentada en un marco educativo. Se realizó aquí una descripción de una serie de posibilidades que ofrecen estas tecnologías en diferentes escenarios educativos, y se analizaron tres investigaciones que vinculan la RA en la educación de nivel medio y superior.

El principal aporte a esta investigación consistió en llevar adelante una experiencia áulica de nivel universitario, donde se utilizó material didáctico con tecnologías de RA para aumentar ejercicios prácticos de una unidad de una cátedra que está relacionada a la enseñanza de la programación en el nivel superior.

Los análisis y resultados preliminares de dicha experiencia permiten visualizar resultados positivos tanto desde la percepción del estudiante como del docente. Se mencionan beneficios como un mayor interés y atención, satisfacción, interactividad, y otros factores que directa o indirectamente producen mejoras los procesos de enseñanza-aprendizaje. Incluso es posible visibilizar un aumento en el rendimiento del alumnado.

Como aporte para su mejora, se sugiere que el material empleado sea modificado para funcionar en dispositivos con sistema operativo IOs, ya que para la experiencia sólo funcionó para dispositivos con Android. Al mismo tiempo se recomienda repetir la experiencia en más unidades de la cátedra, con más trabajos prácticos aumentados, incluso en otras cátedras, a fines de obtener un panorama más amplio y completo del potencial real que se puede obtener de la utilización de esta herramienta de RA.

Ciertamente, es posible afirmar que las herramientas de RA se pueden considerar válidas como un recurso didáctico más a ser utilizado en las prácticas docentes para favorecer los procesos de enseñanza-aprendizaje, y se seguirá incursionando en el mismo en los próximos ciclos lectivos.

Referencias

1. Ramírez Otero, J.R., Solano Galindo, S.: ARprende: una plataforma para realidad aumentada en Educación Superior. Universidad del Atlántico, Colombia. (2017).
2. Cubillo Arribas, J., Martín Gutiérrez, S., Castro Gil, M., Colmenar Santos, A.: Recursos Digitales Autónomos Mediante Realidad Aumentada. Universidad Nacional de Educación a Distancia (UNED), España. (2014).
3. Martín-Gutiérrez, J., Fabiani, P., Benesova, W., Dolores Meneses, M., Mora, C. E.: Augmented reality to promote collaborative and autonomous learning in higher education. Computers in Human Behavior 51. (2014).
4. Moreno, A. J., Rodriguez, C., Ramos, M., Sola, J.: Interés y motivación del estudiantado de Educación Secundaria en el uso de Aurasma en el aula de Educación Física. (2020).

ISBN 978 -987-633-574-4

5. Moro, C., Smith, J., Finch, E.: Improving stroke education with augmented reality: A randomized control trial. Computers and Education Open. (2021).
6. Di Serio, A., Ibañez, M. B., Delgado Kloos, C.: Impact of an augmented reality system on student´s motivation for a visual art course. (2013).
7. Fabregat Gesa, R.: Combinando la realidad aumentada con las plataformas de e- learning adaptativas. (2012).
8. Cabero, J., Barroso, J.: The educational possibilities of augmented reality. New approaches in Educational Research. Vol 5. N° 1. (2016).
9. Billinghurst, M., Kato, H., Poupyrev, I.: MagicBook: Transitioning between Reality and Virtuality. (2011).
10. Cózar Gutierrez, R., Hernández Bravo, J. A., De Moya Martinez, M., Hernández Bravo. J. R.: Tecnologías emergentes para la enseñanza de las Ciencias Sociales. Una experiencia con el uso de Realidad Aumentada para la formación inicial de maestros. (2015).
11. Sánchez Bolado, J.: El potencial de la realidad aumentada en la enseñanza del español como lengua extranjera. (2016).
12. Fernández Robles, B.: La utilización de objetos de aprendizaje de realidad aumentada en la enseñanza universitaria de educación primaria. International Journal of Educational Research and Innovation. 9. (2018).
13. Leiva Olivencia, J. J., Moreno Martínez, N. M.: Tecnologías de geolocalización y realidad aumentada en contextos educativos: experiencias y herramientas didácticas. (2015).
14. Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic literature reviews in software engineering - A systematic literature review. (2008).
15. Fonseca, D., Redondo, E., Valls, F.: Motivation and Academic Improvement Using Augmented Reality for 3D Architectural Visualization. (2016).
16. López-Belmonte, J., Pozo, S., Fuentes, A., Romero, J. M.: Eficacia del aprendizaje mediante flipped learning con realidad aumentada en la educación sanitaria escolar. (2020).
17. Loorbach, N., Peters, O., Karreman, J., Steehoulder, M.: Validation of the Instructional Materials Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology. (2015).

ISBN 978 -987-633-574-4

Aportes de las herramientas digitales a STEM, durante la investigación científica, para fomentar el desarrollo de competencias científicas, tecnológicas y digitales

Silvina Manganelli[12]
[1] ITU, Instituto Tecnológico universitario, Universidad Nacional de Cuyo
[2] FCEN, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo
smanganelli@fcen.uncu.edu.ar

Resumen. Actualmente existe un amplio abanico de herramientas digitales que pueden usarse en la enseñanza de la ciencia, la tecnología, la ingeniería y la matemática. Existen puntos de encuentro entre las prácticas educativas STEM y las herramientas digitales. En las prácticas STEM que este trabajo plantea, los estudiantes utilizarán sus teléfonos celulares para realizar la recolección y análisis de datos durante la experimentación científica propuesta. Deseamos probar que una adecuada relación entre las prácticas educativas STEM y las herramientas digitales, puede servir para mejorar tanto las competencias científicas tecnológicas de los estudiantes como sus competencias digitales, necesarias para el desarrollo personal y profesional en la era digital. Para comprobar esta hipótesis, se han diseñado instrumentos con pautas para evaluar el desarrollo de competencias científicas, tecnológicas, digitales y las necesarias para afrontar los desafíos del siglo XXI, en alumnos pertenecientes a carreras del área de las ciencias exactas e ingeniería.

Palabras Claves: stem, maker, aprendizaje basado en problemas, aprendizaje basado en proyectos, competencias, smartphones, aplicaciones, competencias del siglo XXI, pensamiento computacional, matemáticas, ciencias exactas.

ISBN 978 -987-633-574-4

1. Introducción

Las prácticas STEM, han planteado la necesidad de abordar tres dimensiones de la práctica científica: la primera dimensión comprende la experimentación con fenómenos naturales y tecnológicos mediante la observación, manipulación, también la recolección y análisis de datos, la segunda dimensión comprende la elaboración de modelos científicos y matemáticos, la interacción con representaciones virtuales de entidades abstractas y la tercera dimensión comprende la argumentación y comunicación de soluciones científicas, matemáticas y tecnológicas, así como la evaluación de pruebas y argumentos aportados por los demás. En las prácticas STEM, las herramientas digitales que utilizan los estudiantes para llevar a cabo la recolección y análisis de datos, necesarias para realizar la experimentación científica propuesta, facilitan el acceso a esos datos experimentales y también enriquecen su análisis. Por ejemplo, el uso de sensores digitales en el aula. De este modo los estudiantes pueden invertir menos tiempo en el proceso de toma de datos e invertir más tiempo en el análisis y la interpretación de los datos obtenidos [1].

Para llevar a cabo la recolección de datos los alumnos cuentan con las herramientas digitales más emblemáticas de la vida actual digital diaria, sus teléfonos celulares. Estas herramientas digitales han irrumpido en todos los aspectos de la vida, incluyendo las actividades profesionales y científicas. Los celulares modernos contienen potentes cámaras digitales, micrófonos, receptores GPS/GNSS, acelerómetros, giroscopios, sensores de magnetismo, luxómetros, barómetros, termómetros, sensores de humedad, sensores biométricos, etc, por lo cual pueden convertirse en importantes aliados del trabajo de un investigador [2].

En esta comunicación se proponen y describen dos prácticas educativas, adaptadas a la modalidad de cursado actual: virtual o en línea. Estas actividades integran el uso de celulares (sensores y aplicaciones) para llevar a cabo el proceso investigativo, la recolección de datos y acompañan a toda la actividad. Se trata de dos proyectos tecnológicos que involucran metodologías activas como el Aprendizaje Basado en Problemas, el movimiento maker y el pensamiento computacional. En cada práctica participan dos espacios curriculares. Para llevar a cabo el desarrollo de cada actividad, se revisaron los descriptores, programas y planificación educativa de los espacios involucrados, con el objeto de seleccionar una temática en común, que permitiera el desarrollo de una práctica que aborde los contenidos curriculares, objetivos de aprendizaje y competencias deseadas por ambas cátedras.

1.1. Objetivos

Esta propuesta de actividades tiene como objetivo, demostrar como la integración de herramientas digitales cotidianas de la vida actual, como los smartphones, en proyectos interdisciplinares STEM, estimularán tanto el desarrollo de las competencias científicas tecnológicas de los estudiantes como el desarrollo de sus

ISBN 978 -987-633-574-4

competencias digitales necesarias para el desarrollo personal y profesional en la era digital.

También se desea demostrar como una actividad basada en el enfoque STEM fomentará en los estudiantes: a) el desarrollo de habilidades del pensamiento lógico y de la resolución de problemas y b) mejorará los resultados o rendimiento de los alumnos en matemática, en física, en materias del área de las ciencias exactas.

Finalmente se busca estimular en los estudiantes, a través de esta propuesta de actividades, el desarrollo de habilidades transversales como: la investigación, el pensamiento crítico, la resolución de problemas, la creatividad, la comunicación, la colaboración y el pensamiento computacional, competencias requeridas por los ciudadanos para asumir un papel activo en una sociedad cada vez más tecnológica, competencias para el siglo XXI [3].

2. Metodología

A continuación se describen dos actividades. La primera indaga en las bondades de una metodología de enseñanza aprendizaje basada en el enfoque STEM como herramienta optima para llevar a cabo la resolución de problemas matemáticos [4].

Se desea demostrar como una actividad basada en el enfoque STEM fomentará en los estudiantes, el desarrollo de habilidades del pensamiento lógico y de la resolución de problemas y mejorará los resultados de los alumnos en matemática.

a. **Actividad 1** Experimentación Científica en casa: *¿Quién es y qué puede hacer π?*

Esta primera actividad integra dos espacios curriculares que forman parte del ciclo básico de carreras del área de las ciencias exactas. Estos son:

Introducción a las matemáticas

- Unidad 0: Repaso de algunos conceptos de Geometría
- Subunidad: Fórmulas de área y volumen que contienen el número π

Informática

- Unidad 6: Procesadores de textos científicos: Latex, Lyx, VerbTeX
- Unidad 7: Nuevas herramientas tecnológicas aplicadas a la ciencia
- Subunidad: Herramientas aplicadas a la investigación y/o simulación

Partimos entonces con la siguiente pregunta de investigación: ¿Cómo influye en el éxito de la resolución de problemas matemáticos el uso de actividades STEM?

2.1. Exploración del problema. Inmersión.

Se invita a los alumnos a pensar las siguientes cuestiones: ¿Cómo elegir el mejor vaso para tomar algo? ¿Da lo mismo cualquier tamaño? ¿Qué ganamos y qué perdemos con esa elección? ¿Tiene importancia la forma? ¿Por qué? ¿Cuál es la importancia del material? ¿Qué se debería tener en cuenta?

ISBN 978 -987-633-574-4

Preguntas

La exploración de este tema permite a los alumnos indagar nuevos aspectos o perspectivas sobre el mismo. Esta indagación es una consecuencia natural de nuestra innata curiosidad y deseo de descubrir. Estas preguntas impulsan o activan el deseo de explorar, descubrir e investigar. Algunas preguntas posibles:

¿Porque existen diferentes vasos para diferentes líquidos? ¿Cuál de los recipientes se llena primero? ¿Cuál es la forma del agua en un vaso? ¿Cómo es el volumen del agua en un vaso? ¿Qué mide más la altura o el perímetro de un vaso? ¿Qué es más relevante para contener más líquido, el perímetro o la altura? ¿Con cuál de todos obtengo más cantidad de líquido?

Herramienta: Para registrar, publicar y compartir todas esas preguntas, reflexiones e indagaciones los alumnos utilizarán la aplicación, *Jamboard*, pizarra en la nube.

Formulación de la hipótesis

La primera tarea que realizará el alumno, será formular una hipótesis, sobre cual se presume sería el vaso ideal para contener la mayor cantidad de líquido. Cada alumno formulará su propia hipótesis.

2.2. Objetivos – Saberes - Producto final - Cronograma

La segunda tarea será recordar los objetivos y saberes de aprendizaje que plantea esta actividad.

Objetivos: Facilitar el conocimiento del valor de PI, a través de la manipulación de materiales concretos que permiten una mejor comprensión y dominio cognitivo del alumno.

Saberes: Resolver una situación problemática que requiera del conocimiento del número PI y su utilización ante una necesidad.

Herramienta: El alumno desarrollará un cronograma de las tareas previas que lo conducirán a la producción final. Se propone utilizar *Google Calendar -Time Blocking Tip,* herramienta que permite crear, gestionar y sincronizar cronogramas

ISBN 978 -987-633-574-4

individuales o colaborativos, también notifica en tiempo real al móvil o a el correo sobre las tareas pendientes en el día a día.

Experimentación

A continuación, se les solicita busquen información que consideren relevante, para responder a las preguntas formuladas.

Por ejemplo, indagar sobre: ¿Por qué es mejor usar el vaso indicado para cada tipo de líquido? ¿Qué mide más la altura o el perímetro de un vaso? ¿Qué es más relevante para contener mas liquido, el perímetro o la altura? ¿Con cuál de todos obtengo la mayor cantidad de líquido?

Mediciones

La siguiente tarea será realizar mediciones, sobre el ancho y alto de cada uno de los vasos, con estos datos el alumno calculará la capacidad y el volumen, mediciones necesarias para resolver el desafío planteado. Para llevar a cabo la recolección de estos datos y realizar los cálculos necesarios, se propone utilizar las siguientes aplicaciones: Volumen Calculadora, Conversor de unidades, Aruler (regla con Realidad aumentada), Telémetro: Smart Measure, Smart Ruler.

Registro y evaluación de mediciones

Además deberán armar una tabla para registrar las mediciones y exploraciones realizadas, la cual facilitará la elaboración de conclusiones.

Tipo de vaso	**Alto**	**Ancho**	**Perímetro**	**Área**	**Volumen (cm^3)**	**Capacidad (ml)**	**Nivel de precisión de la construcción**	**Ventajas**
Formulas			$\pi * d$	$\pi * r^2$	$\frac{h * \pi}{3}(R^2 + r^2 + R * r)$			
Caña								
….								

Emisión de juicios y conclusiones

Para volcar las observaciones, juicios y hallazgos, el alumno redactará un informe científico, que compartirá con los docentes de los espacios curriculares involucrados y con sus compañeros, con el fin de compartir y recibir observaciones y devoluciones. Se sugiere utilizar para esto la aplicación gratuita y colaborativo Zoho Writer.

Retos científicos y tecnológicos

Retos científicos

Realizar el experimento con otros recipientes, ya no solo vasos de vidrio, también de plástico, metal y cerámica.

Realizar el experimento con otros líquidos, no solo agua, por ejemplo vino ¿varían los resultados de alguna forma?

¿Y si agregáramos un cubo de hielo? ¿Aumenta el nivel de agua?
¿Y si agregáramos un rodaja de limón?

Retos tecnológicos

Aplicando pensamiento computacional: En un mundo donde math.pi y pi = atan(1) * 4 no existen, el Método de Montecarlo, se puede usar para estimar π . Se propone al alumno desarrollar un algoritmo para estimar π. Desarrollar los algoritmos correspondientes y a continuación codificar la solución más óptima en un lenguaje de programación como puede ser Python.

b. **Actividad 2** Experimentación Científica en casa: *Interactuando con Magnitudes Físicas e interpretando Unidades de Medición*

Según una investigación realizada por el Ministerio de Educación en Argentina, cuyo objetivo principal era diagnosticar el nivel de aprendizaje de los alumnos en diferentes áreas curriculares con énfasis en matemática, los rendimientos en los alumnos mejoran cuando el alumno percibe el medio ambiente físico del aula como adecuado [5].

Por lo cual se puede considerar al medio ambiente o entorno físico del estudiante como una variable que influye en el rendimiento académico del alumno. Si el alumno lo percibe como adecuado, entonces ese entorno físico se lo puede considerar como un aporte a su rendimiento escolar [6].

La modalidad de cursado actual en nivel superior y universitario es virtual o a distancia. En estas condiciones, el entorno físico del estudiante se convierte ahora en su aula física de cursado, estudio o trabajo, tomando un rol emblemático. La presente actividad invita a los alumnos a realizar una experimentación científica, una indagación, con el fin de encontrar el espacio o lugar ideal dentro de su casa, para convertirlo en su zona, pieza o habitación de estudio.

Para ello se tendrá en cuenta los factores ambientales o la ergonomía ambiental de un espacio como la temperatura, ruido, vibraciones e iluminación, que influyen en la salud, bienestar y rendimiento académico del trabajador [7], en este caso del estudiante. Para llevar a cabo esta investigación, los alumnos harán uso de los sensores de sus teléfonos y de sus aplicaciones en las investigaciones.

Esta actividad integra a dos espacios curriculares que forman parte del ciclo básico de carreras del área de las ciencias exactas. Estos son:

1. Introducción a la Física
   1.1. Unidad 1: Magnitudes Físicas y Unidades
      1.1.1. Subunidad: Mediciones e incertezas
2. Informática
   2.1. Unidad 6: Procesador de textos científicos: Lyx
   2.2. Unidad 7: Nuevas herramientas tecnológicas aplicadas a la ciencia
      2.2.1. Subunidad: Herramientas aplicadas a la investigación y/o simulación.

Se plantea la siguiente pregunta de investigación: ¿Cómo influye en el éxito de la resolución de problemas de introducción a la física, el uso de actividades STEM?

2.3. Exploración del problema. Inmersión.

El alumno investiga por un lado sobre la Ergonomía y el Medio ambiente de trabajo. También reflexiona cuales serían las características o factores ideales que

debería presentar esa habitación o entorno para ser considerada la más óptima como pieza de estudio.

Preguntas

Estas preguntas impulsan o activan el deseo de explorar, descubrir e investigar.
Algunas preguntas posibles:

- ¿El lugar ideal deberá contar con mucha iluminación?
- El lugar ideal deberá contar con una temperatura adecuada. ¿Cuál sería la temperatura adecuada para nosotros? ¿Estará muy relacionado con la temperatura corporal?
- El lugar ideal deberá contar con bajo nivel de ruido. ¿Mientras más silencioso mejor?
- ¿Qué otros factores debería tener en cuenta?

Herramienta

Para emitir y compartir las preguntas, se utilizará de nuevo la aplicación *Jamboard.*

Formulación de la hipótesis

La primera tarea que debe realizar el alumno, será formular una hipótesis, sobre cual presume sería la habitación ideal de la casa para convertirla en una pieza o zona de estudio. Cada alumno formulará su propia hipótesis

2.4. Objetivos - Producto final - Cronograma

La segunda tarea será revisar los objetivos de aprendizaje.
Objetivos: Facilitar el conocimiento e interpretación de magnitudes físicas, unidades y medidas a través de la manipulación de materiales concretos que permiten una mejor comprensión y dominio cognitivo del alumno.
Saberes: Resolver una situación problemática que requiera del conocimiento de magnitudes físicas, su utilización, manejo e interpretación ante una necesidad.
Herramienta: Nuevamente se sugiere utilizar Google Calendar - Time Blocking Tip, para desarrollar un cronograma de las tareas previas que lo conducirán a la producción final.

Experimentación

A continuación para llevar a cabo el experimento, para realizar las mediciones correspondientes, los alumnos utilizarán Arduino Science Journal, aplicación móvil que permite realizar experimentos científicos haciendo uso de los sensores del teléfono. También harán uso de otras aplicaciones, como termómetros en línea que miden la temperatura que existe en el ambiente en cualquier momento y en cualquier lugar. Las variables a medir serán: luminosidad, humedad, sonido, temperatura, vibraciones.

Mediciones

La siguiente tarea será determinar cuanta humedad, temperatura, luminosidad, sonido, vibraciones envuelve a cada habitación posible de la casa. Estas lecturas se realizarán diariamente, cada 8 hs, mañana, tarde y noche, durante al menos una semana.

ISBN 978 -987-633-574-4

Registro y evaluación de mediciones

Para el registro de las observaciones y mediciones el alumno puede utilizar el mismo Arduino Science Journal.

Emisión de juicios y conclusiones

Para volcar las observaciones y hallazgos, el alumno redactará un informe científico, que compartirá con los docentes de los espacios curriculares involucrados y con los compañeros, con el fin de compartir y recibir observaciones y devoluciones. Se sugiere utilizar nuevamente Zoho Writer.

Retos científicos y tecnológicos

Finalmente se propone al alumno varios desafíos y retos, tanto científicos como tecnológicos.

Desafíos e Interrogantes

- Intentar mover lentamente el dedo por el teléfono mientras observa la tarjeta del sensor en la aplicación Arduino Science Journal. ¿En qué momento el sensor medirá 0 lux?
- Colocar una fuente de luz puntual única en una habitación oscura, luego medir la iluminación a varias distancias de la luz. Reflexionar ¿Cómo afecta la medición al duplicar o triplicar la distancia? ¿Lo mismo ocurre con el sonido? ¿Cómo cambia la medida? ¿Qué ley se cumple?
- ¿Cómo se podrían comparan una fuente de luz artificial con la luz natural de una ventana?
- ¿Porque los teléfonos inteligentes sólo pueden detectar su temperatura interna y no cuentan con un termómetro integrado?

Retos científicos

- Hacer el experimento con más réplicas, cada 4 hs y/o durante un mes ¿se mantienen los mismos resultados?
- Hacer el experimento en diferentes estaciones del año (otoño, invierno, primavera, verano).
- Aplicar estadística para determinar si las diferencias no se deben al azar.

Retos tecnológicos

Aplicando pensamiento computacional: ¿Cuántas luces LED se necesitarán por metro cuadrado para iluminar la habitación seleccionada como pieza de estudio con el fin de obtener la misma iluminación de noche y de día?

Desarrollar los algoritmos correspondientes y a continuación codificar la solución más óptima en un lenguaje de programación como puede ser Python. También deberá determinar los lúmenes necesarios.

2.5. Producto final – Reflexión - Divulgación

Construya y comparta un resumen científico, con lo aprendido en las indagaciones efectuadas en ambas actividades propuestas. Agregue a este artículo científico, las respuestas a las preguntas formuladas en desafíos, también incorpore los resultados obtenidos en los retos, tanto los científicos como los tecnológicos. Esta documentación servirá a futuros estudiantes abordar las conclusiones obtenidas. Para la redacción del artículo utilizaremos como procesador de textos científico Latex y la aplicación VerbTeX.

ISBN 978 -987-633-574-4

3. Instrumento de Evaluación

A continuación se detalla los instrumentos de evaluación que se utilizarán para evaluar el nivel o grado de desarrollo de las competencias científicas presente en los alumnos del ciclo básico de carreras del área de las ciencias exactas e ingeniería. Estos instrumentos se aplican tanto en su modalidad de pre-test como pos-test. Se han desarrollado tres instrumentos de evaluación.

Instrumento para evaluar competencias científicas

Competencias a evaluar: explorar hechos y fenómenos, analizar problemas, formular hipótesis observar, recoger y organizar la información, utilizar diferentes métodos de análisis, evaluar métodos.

Instrumento para evaluar competencias del siglo XXI

Competencias a evaluar: pensamiento crítico, colaboración, resolución de problemas, creatividad, pensamiento computacional.

Instrumento para evaluar competencias del siglo XXI

Competencias digitales como: trabajar en un entorno digital, ser responsable en la era digital, producir, procesar, explotar y difundir documentos digitales, organizar la búsqueda de información en la era digital, trabajar en red, comunicar y colaborar.

Clasificaremos el grado de desarrollo de estas competencias presentes en los alumnos en 5 niveles: Inferior, Medio Bajo, Medio, Media Alto, Superior

4. Conclusiones

En estas prácticas STEM, para llevar a cabo, la investigación científica, los estudiantes, utilizaron las aplicaciones y los sensores internos de su teléfono. Estas herramientas facilitaron el acceso a los datos experimentales, enriquecieron su análisis y favorecieron los tiempos de recolección de datos, ya que los estudiantes pudieron invertir menos tiempo en el proceso de toma de datos e invertir más tiempo en el análisis y la interpretación de los datos obtenidos.

Estas prácticas resultarán beneficiosas para los alumnos, tanto para el desarrollo de las competencias STEM como para el desarrollo de las competencias digitales, ya que el uso de las herramientas digitales propuestas y la manipulación de materiales concretos permitirán una mejor comprensión y dominio cognitivo del alumno, dando como resultado experiencias que resultarán ser análogas a las prácticas reales que lleva a cabo un investigador.

Ambas competencias STEM y digitales son necesarias para la ciudadanía del siglo XXI, ya que las prácticas educativas STEM buscan alfabetizar y dotar de competencias STEM al conjunto de los futuros ciudadanos (vayan a convertirse o no en profesionales STEM), para hacer una sociedad más capaz de involucrarse y tomar partido en los retos científico-tecnológicos propuestos, una sociedad que cada vez será más tecnológica y digital.

Respecto al empleo de los sensores internos de los teléfonos celulares y sus aplicaciones, nos encontramos en un momento apasionante y oportuno para incorporar estas tecnologías de uso cotidiano en las investigaciones. Los teléfonos celulares, que tan inesperadamente y en forma acelerada han irrumpido la vida diaria, tienen potencialidades para ir más allá y convertirse en importantes herramientas de trabajo científico. Aprender a hacer un uso más eficiente de estas herramientas puede revolucionar la manera en que los investigadores trabajan.

Por otro lado se podemos observar que son varios los beneficios que aportan las actividades STEM a la enseñanza matemática y a la resolución de problemas tanto del área de la matemática como de la física. Los estudiantes destacan que esta experiencia les sirve para razonar y valorar distintas alternativas al enfrentarse a un problema, ya que a menudo los resolvían de manera mecánica y automática sin evaluar otras opciones. Las actividades STEM permitirán al alumno organizar la información disponible en el enunciado y discernir entre datos relevantes y superfluos, este es uno de los aspectos más señalados, por gran parte del alumnado, ya que muchas veces suelen tener problemas en entender qué se les está preguntando y qué datos son los necesarios para resolver el problema.

Esto se traduce entonces como una mejora en la comprensión de los contenidos implicados en la resolución de problemas, en la comprensión de enunciados y en la organización y análisis de los datos.

5. Agradecimientos

La autora agradece la ayuda financiera recibida de la Secretaría de Investigación, Internacionales y Posgrado de la UNCUYO, para llevar a cabo el Proyecto Bienal SIIP Tipo 1 2019-2021, T003 “STEAM un nuevo enfoque didáctico para la formación científica de alumnos pertenecientes a carreras del área de las ciencias exactas, necesarias para afrontar los desafíos del siglo XXI” (aprobado por resolución Nº 3922/2019-R). También a las unidades académicas ITU y FCEN Uncuyo, involucradas en el desarrollo del proyecto, por sus prestaciones y excelente predisposición para hacer posible la ejecución de las actividades propuestas.

6. Referencias

[1] López, V., Couso, D., Simarro, C. (2018). Educación STEM en y para el mundo digital. Cómo y por qué llevar las herramientas digitales a las aulas de ciencias, matemáticas y tecnologías. RED. Revista de Educación a Distancia, 5XX. Consultado el (28/07/2021) en https://revistas.um.es/red/article/view/410011

[2] Denis, D., Cruz Flores, D., Ferrer-Sánchez, Y., & Felipe Tamé, F. (2021). Potencialidades de los celulares inteligentes para investigaciones biológicas. Parte 1: Sensores integrados. Revista del Jardín Botánico Nacional, 42, 77-91. Recuperado de http://www.rjbn.uh.cu/index.php/RJBN/article/view/542

[3] Zanafria Zepeda, Romero. (2018). Competencias del siglo XXI en proyectos co-tecnocreativos. Revista Mexicana de bachillerato a distancia. Vol X (Nro 19). Recuperado de http://revistas.unam.mx/index.php/rmbd/article/view/64889/56919

[4] Fernández-Blanco, T.,González-Roel, V., Ares, A. (2020). Estudio exploratorio de las steam desde las matemáticas DOI: http://dx.doi.org/10.17346/se.vol0.375. Saber & Educar

[5] Ministerio de Cultura y Educación. Secretaría de Programación y Evaluación Educativa. (1999). Características del alumno y rendimiento escolar en matemática, alumnos del 7º año-escuela urbana, operativo nacional de evaluación 1995-1997. Ciudad Autónoma de Buenos Aires, Argentina. Biblioteca Nacional de Maestros. Recuperado de: http://www.bnm.me.gov.ar/giga1/documentos/EL001003.pdf

[6] Bernal García, Y., Rodríguez Coronad, C. (2017). Factores que Inciden en el Rendimiento Escolar de los Estudiantes. Maestría en educación, Universidad Cooperativa de Colombia, Facultad de Educación.

[7] Marín Martínez, Amina. (1993). Los Factores ergonómicos para el aumento de la productividad. Licenciatura en Higiene y Seguridad. Recuperado de http://www.bidi.uson.mx/TesisIndice.aspx?tesis=4431

ISBN 978 -987-633-574-4

SIMA. Un sistema integral modular para la gestión administrativa de la Educación Superior

Eduardo E. Mendoza[1], Juan P. Méndez[1],
Diego F. Craig[2], Verónica K. Pagnoni[3],
[1]Ministerio de Educación de la Provincia de Corrientes - Dirección de Sistemas
[2]Ministerio de Educación de la Provincia de Corrientes - Dirección de Nivel Superior
[3]Instituto Superior de Formación Docente 'Bella Vista' - Corrientes
equipotecnico@dgescorrientes.net

Abstract. En el presente artículo se expone acerca del desarrollo e implementación de un Sistema Integral Modular Administrativo (SIMA) que conforma una serie de funcionalidades diversas concernientes a la informatización de los Institutos Superiores de Formación Docente y Técnica (ISFD) dependientes de la Dirección de Nivel Superior de la Provincia de Corrientes. Se presentan objetivos y resultados alcanzados en el desarrollo del proyecto, una caracterización general de la arquitectura y funcionalidades de la plataforma, como así también se expresan conclusiones y futuras líneas de acción.

Palabras clave: Desarrollo de software, diseño modular, educación superior, gestión educativa, gestión administrativa

1. Contexto

Este proyecto se realiza en el marco de las líneas de acción de la Dirección de Nivel Superior de la Provincia de Corrientes. El área TIC, tiene como objetivo agilizar y fortalecer la gestión de la información que circula hacia el interior y exterior de los institutos superiores.

El Sistema Integral Modular Administrativo (SIMA) nace como una iniciativa informática de una unidad educativa específica. Desde el 2018 se convierte en una acción a nivel jurisdiccional y se continúa su desarrollo dentro del Proyecto Informatización de Institutos de Educación Superior de La Provincia de Corrientes. De esta manera, SIMA ha crecido en funcionalidades y complejidad, multiplicando su presencia en más unidades educativas y posibilitando nuevos desarrollos con una visión más amplia.

ISBN 978 -987-633-574-4

2. Introducción

Un Sistema de Información y Gestión Educativa (SIGED) se puede definir como el conjunto de funcionalidades de gestión educativa que se utilizan para diseñar, registrar, explotar, generar y diseminar información estratégica de forma integral, encuadrados por una infraestructura legal, institucional y tecnológica determinada [1].

El uso de las tecnologías posibilita la automatización de procesos que se realizaban manualmente o con poca sistematización. De esta manera, la transformación digital puede impulsar la innovación en la gestión educativa [2].

Según [3] se deben considerar las necesidades del sistema educativo en su totalidad para desarrollar SIGED que posibilite gestionar los procesos importantes de forma eficiente valiéndose de las tecnologías digitales. Estos autores indican que la transformación digital de un SIGED conlleva una serie de ventajas en la gestión del educativo, tales como:

1. La disponibilidad de información oportuna y de calidad para el diseño de políticas y la asignación de recursos.
2. El ahorro de tiempo resultante de aquellas tareas administrativas que pasan de realizarse de manera manual a implementarse usando tecnologías.
3. Ahorros presupuestarios debido al uso más eficiente de los recursos.

En [4] brindan algunos datos sobre los avances de la digitalización de la gestión educativa: en Uruguay y Espíritu Santo (Brasil) la asistencia de los estudiantes se comprueba por medio de aplicaciones digitales permitiendo la transición del registro en papel al digital, mejorando la calidad, la confiabilidad y la oportunidad de los datos; en Bogotá, Colombia; Espíritu Santo, Brasil; Mendoza, Argentina; Uruguay se han digitalizado procesos tales como la gestión de reemplazos docentes, la entrega de títulos y certificaciones, y la gestión de reparaciones de edificios en sistemas implementados; respecto al intercambio de información, algunos casos mencionados por los autores son la consulta de los registros de recursos humanos mediante aplicaciones en Santa Fe (Argentina), la comunicación entre docentes y padres por medio de aplicaciones en Mendoza (Argentina) y Uruguay, y la consulta de calificaciones por medio de apps en Santa Fe (Argentina) y Uruguay.

ISBN 978 -987-633-574-4

La industria del desarrollo del software ha cambiado en los últimos años con la aparición de internet y herramientas que permiten crear sistemas de forma colaborativa y por ende de manera más rápida. En este marco, las metodologías tradicionales de desarrollo de software fueron sustituidas o adaptadas sobre todo debido a la demanda de los usuarios [5].

En [6] indican que un sistema de software complejo se debe dividir en piezas más simples o módulos. La organización modular permite aplicar la separación de intereses en dos sentidos, por un lado trabajando con los detalles de un módulo en forma aislada, y por otro desarrollar las características de todos los módulos y sus relaciones para integrarlos.

La participación del usuario en todas las etapas de desarrollo de los sistemas se ha vuelto crucial, en la identificación de requerimientos, en la definición de los objetos que pertenecen al dominio del problema y en la implementación eficiente del sistema.

Finalmente, se considera importante destacar que en estos tiempos de cambios vertiginosos es fundamental llevar adelante un desarrollo de software que se adapte a ellos, estableciendo formas de trabajo que favorezcan la optimización de los procesos.

3. Resultados y Objetivos

Los objetivos establecidos desde el departamento TIC para llevar adelante el proyecto son:

- Estudio de la normativa vigente.
- Seleccionar paradigma, marcos de trabajo y herramientas para los procesos de desarrollo de software.
- Definir de un equipo que desarrolle acciones I+D.
- Desarrollar e implementar de forma modular SIMA en diferentes Institutos Superiores de la provincia.
- Capacitar al personal administrativo y docente en el manejo de SIMA.

Se lograron avances en los siguientes tópicos:

- Se realizó una recopilación y análisis exhaustivo de las normativas que rigen la vida institucional de los institutos superiores de la provincia. Sin embargo, se debe considerar que la normativa se actualiza constantemente por lo cual este objetivo se cumplimentará cada vez que esto suceda.
- Se decidió trabajar considerando la programación modular para favorecer la reutilización de código y la independencia entre los módulos, debido a que se deben ajustar constantemente para cumplimentar la normativa. También se decidió dar un rol importante al usuario, con quien los desarrolladores definen los módulos a programar y sus características; asimismo los usuarios son los que aportan a la mejora de los módulos ya implementados.
- El equipo está en constante formación, cuenta con un coordinador, dos programadores y editores de documentación.
- Se utilizan los módulos correspondientes a la Formación Docente Continua en 50 organizaciones (institutos, direcciones, coordinaciones) dependientes del Ministerio de Educación. En tanto, los módulos referidos a la Administración de Formación Inicial en 74 instituciones entre sedes centrales y anexos.
- Se realizan constantemente reuniones, relevamientos, capacitaciones, acompañamiento, etc., desde los usuarios finales hasta los directivos de los establecimientos, para asegurar el mejor grado utilización y aceptación del Sistema.

4. Arquitectura de la plataforma

SIMA es un sistema que se basa en cuatro pilares:

ISBN 978 -987-633-574-4

Figura 1. Pilares de SIMA (producción propia)

Los institutos superiores de la provincia se rigen por normativas emitidas por la Dirección de Nivel Superior del Ministerio de la Provincia, las que conforman un cúmulo de documentos que enmarcan las acciones de estas instituciones. Es por ello, que es fundamental su estudio y análisis, para establecer en SIMA funcionalidades que ayuden a su cumplimiento.

La participación del usuario es clave tanto en la implementación de SIMA como en su desarrollo. Se mantiene una comunicación fluida con los directivos y usuarios directos (administrativos, docentes, estudiantes), la que sirve para recopilar información sobre las prioridades a informatizar, las falencias que se deben subsanar y las mejoras que se pueden implementar. Asimismo, cada vez que se programa una nueva funcionalidad se realiza la capacitación y acompañamiento correspondientes para favorecer su comprensión y un adecuado manejo del sistema.

La implementación de un SIGED debe abarcar cierta estructura de infraestructura tecnológica, es decir, el hardware y software necesario para soportar el buen desenvolvimiento del sistema. Considerando las funcionalidades que conciernen a la condición estructural de la infraestructura tecnológica establecidas en [3], SIMA cuenta con:

ISBN 978 -987-633-574-4

- Conectividad: esta característica debe considerarse tanto en el centro de operaciones como en los nodos, SIMA está montado sobre servidores propios en la Dirección de Nivel Superior que aseguran su buen funcionamiento, en tanto, las instituciones se conectan al sistema usando sus propios recursos de hardware y software. Se debería a futuro realizar un relevamiento de necesidades de infraestructura en los Institutos y gestionar la adquisición de equipos que garanticen una conectividad adecuada.
- Tecnología para procesamiento y desarrollo: En las tareas de programación y desarrollo se cuenta con equipos de última generación. Para el procesamiento y almacenamiento de los datos se trabaja con 4 servidores, uno destinado para los procesos y datos de formación continua, otro para las funcionalidades administrativas y bases de datos de los institutos, un tercero usado como respaldo de datos y el último es utilizado para la gestión de correo.
- Ciberseguridad e integridad de datos: Se utilizan métodos de encriptación en el tratamiento de datos. Asimismo, se usa un generador de backup automático que realiza copias diarias de todas las bases de datos. Estas copias y el servidor de respaldo garantizan el acceso a las funcionalidades de SIMA y bases de datos prácticamente sin interrupciones. Además, para salvaguardar la integridad de las bases de datos los institutos superiores y formación continua trabajan de manera separada.
- Documentación y mantenimiento de sistemas e interoperabilidad: Se mantiene una actualizada documentación de todas las implementaciones realizadas. Así mismo, es una prioridad el mantenimiento y adecuación de los procesos considerando el feedback de los usuarios. Como ya se mencionó antes, se presta especial atención a la interoperabilidad de los módulos debido a los constantes cambios que se realizan, para asegurar la integridad del sistema.

ISBN 978 -987-633-574-4

5. Descripción general de la plataforma

Acceso general desde sima.mec.gob.ar. El Ministerio de Educación de Corrientes cuenta con el dominio https://www.mec.gob.ar/ , sobre el mismo se fueron implementando los módulestablecióos que se iban desarrollando, como por ejemplo http://sima.mec.gob.ar/**catalogo/** o http://sima.mec.gob.ar/**certificados/** o http://simaeducativa.mec.gob.ar/**institutos/**. Al ir aumentando la cantidad de módulos se generó un sub-dominio específico que abarca a todos los existentes e irá brindando acceso a los que se vayan sumando, este dominio es el http://sima.mec.gob.ar/

Formación Docente Continua

La Formación Docente Continua se opera desde tres niveles de gestión, la Dirección de Nivel Superior, los Institutos Superiores que ofrecen capacitación a los docentes de la provincia y los propios docentes que son los usuarios finales del módulo.

La Provincia de Corrientes cuenta con el Programa Provincial de Formación Docente Continua "Corrientes Educa Virtual" http://dgescorrientes.net/cev/ y sobre el mismo es que funciona este módulo, ofreciendo los siguientes servicios:

- Gestión de todas las propuestas formativas que ofrecen, tanto las reparticiones gubernamentales, como los Institutos Superiores de Gestión Pública y de Gestión Privada.
- Inscripción On-line, con amplias posibilidades de validación y control.
- Catálogo dinámico de propuestas formativas futuras, en desarrollo y culminadas.
- Historial de formación de cada docente de la provincia.
- Exportación e importación de datos de los cursantes hacia y desde los entornos virtuales de enseñanza y aprendizaje E-ducativa y Moodle.

ISBN 978 -987-633-574-4

- Emisión de certificaciones oficiales con validación mediante códigos QR.

Es de destacar que todos los módulos funcionan en entorno web responsive.

Figura 2. Acciones Administrativas de Formación Continua

Administración de Formación Inicial

La Formación Inicial, se refiere a los trayectos formativos de profesorados y tecnicaturas superiores. La informatización abarca las siguientes funcionalidades:

- Gestión institucional: admisión de los ingresantes a todas las carreras, las inscripciones a las materias, los registros de exámenes, emisión de constancias y de certificados analíticos al finalizar, entre otras funciones específicas.
- Gestión de perfiles de usuarios: que determina permisos y restricciones para todos los módulos. Gracias a esto, las áreas de gestión que funcionan dentro de las instituciones, cuentan con la capacidad de ingresar, editar o sólo observar información en función de sus necesidades específicas.
- Gestión de la información relativa al movimiento de docentes: ingreso, desarrollo profesional y acceso a cargos/horas en función de un completo sistema de padrones y valoraciones basado en la normativa existente.

ISBN 978 -987-633-574-4

Los módulos disponibles para la realización de las gestiones descritas son: Ingresantes, Interinatos y Suplencias, Perfiles Profesionales, Turnos, Cooperadora, Alumnado, Bedelía, Docentes, Coordinación, Valoración, Secretaría y Ajustes.

Figura 3. Servicios Informáticos disponibles para Administración de Formación Inicial

Conclusiones y líneas futuras

A través de este desarrollo I+D se pudo realizar un proceso de diseño y desarrollo de una plataforma para la integración de TIC orientada a la gestión administrativa de instituciones educativas de Nivel Superior de la Provincia.

El grado de cumplimiento de los objetivos propuestos es más que satisfactorio, en 3 años se han informatizado una gran cantidad de procesos administrativos logrando su implementación en 74 instituciones educativas, mejorando de esta manera la calidad y eficiencia en la producción de información oportuna y actualizada; agilizando tareas administrativas que antes se realizaban de manera manual.

Por la caracterización de arquitectura y funcionalidades realizadas se puede afirmar que SIMA se constituye como un SIGED en construcción. Son tareas pendientes: la mejora de infraestructura tecnológica de las

instituciones para lograr una mejor conectividad, la optimización de interfaces de usuario y el ajuste de las herramientas que redunden en favorecer la accesibilidad web para todos los usuarios.

Como visión de futuro se espera proseguir con la detección, evaluación, análisis e informatización de procesos de gestión que generen información útil para la toma de decisiones; avanzar en la implementación de SIMA en otras instituciones educativas de la Provincia; además, se pretende evaluar y analizar los resultados obtenidos capitalizando los saberes y experiencias adquiridos en instancias de transferencia de conocimientos.

Referencias

[1] E. Arias Ortiz, J. Eusebio; M. Pérez Alfaro, M. Vásquez, P. Zoido. Del papel a la nube: Cómo guiar la transformación digital de los Sistemas de Información y Gestión Educativa (SIGED). Washington, D.C.: BID. 2019.

[2] C. Pombo, R. Gupta, M. Stankovic. Servicios sociales para ciudadanos digitales: Oportunidades para América Latina y el Caribe. Washington D.C.: BID. 2018.

[3] E. Arias Ortiz, J. Eusebio, M. Pérez Alfaro, M. Vásquez, P. Zoido. Los Sistemas de Información y Gestión Educativa (SIGED) de América Latina y el Caribe: la ruta hacia la transformación digital de la gestión educativa. Washington, D.C.: BID. 2021.

[4] G. Elacqua, S. Cavalcanti e I. Brant. Em busca de uma maior eficiência e equidade dos recursos escolares: Uma análise a partir do gasto por aluno em Pernambuco. Washington D.C.: BID. 2019.

[5] E. G. Maida, J. Pacienzia. Metodologías de desarrollo de software . Tesis de Licenciatura en Sistemas y Computación. Facultad de Química e Ingeniería “Fray Rogelio Bacon”. Universidad Católica Argentina. 2015.

[6] C. Ghezzi, M. Jazayeri, D. Mandrioli, D. (2002). Fundamentals of Software Engineering. Prentice-Hall.

ISBN 978 -987-633-574-4

Revisión sistemática de metodologías educativas implementadas durante la pandemia por COVID-19 en la Educación Superior en Iberoamérica

Omar Spandre[1], Paula Dieser[1], Cecilia Sanz [2,3]
[1] *Maestría en Tecnología Informática Aplicada en Educación. Facultad de Informática, UNLP*
[2]*Instituto de Investigación en Informática LIDI – CIC. Facultad de Informática, UNLP*
[3]*Comisión de Investigaciones Científicas de la Provincia de Buenos Aires*
spandreomar@gmail.com, pauladieser@gmail.com, csanz@lidi.info.unlp.edu.ar

Resumen. A partir de la interrupción de las clases presenciales, a causa de la pandemia provocada por COVID-19, las Instituciones de Educación Superior adaptaron sus cursos a un formato virtual para atender a los estudiantes durante la contingencia. Este trabajo analiza, mediante una revisión sistemática de un corpus de 24 artículos que estudian los modelos pedagógicos adoptados por dichas Instituciones en Iberoamérica. Los trabajos revisados han sido publicados entre los años 2020 y 2021, y analizan las implicancias de reorganizar los procesos de enseñanza y aprendizaje, a la luz del marco teórico del modelo Comunidad de Indagación. Los resultados muestran que un alto porcentaje de las investigaciones ponen el foco en la Presencia docente, particularmente en el diseño educativo y de organización, con mucho menor énfasis se estudia la Presencia cognitiva, y hay escasa producción sobre la dimensión Presencia social y el análisis de indicadores que refuercen el aprendizaje.

Palabras clave: Educación a distancia, Educación virtual, COVID-19, Educación superior, Modelo de comunidad de indagación.

1 Introducción

Este trabajo analiza la crisis emergente en la educación, y en particular en la educación superior, a partir de la suspensión de las clases presenciales, producida por la pandemia a causa de COVID-19, mediante una revisión sistemática (RS) de un corpus de 24 artículos publicados entre los años 2020 y 2021 en Iberoamérica [1-24].

El nuevo coronavirus SARS-CoV-2 y su enfermedad potencial COVID-19, fue determinado como una pandemia de niveles sin precedentes, que hoy, más exactamente, ha ingresado a una nueva fase avanzada y de mayor letalidad [25].

El reto de los sistemas educativos en los últimos meses ha sido mantener la vitalidad de la educación y promover el desarrollo de aprendizajes significativos. Para ello, ha contado con dos aliados claves: sus docentes y la virtualidad, en términos más precisos, los docentes a través de la virtualidad [7].

Las estimaciones del Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (IESALC), muestran que el cierre temporal de las Instituciones de Educación Superior (IES) había afectado, aproximadamente, a 23,4

millones de estudiantes de educación superior y a 1,4 millones de docentes en América Latina y el Caribe ya antes del fin de marzo de 2020. Esto representaba, aproximadamente, más del 98% de la población de estudiantes y profesores de este nivel educativo en la región [26].

Ante esta emergencia sanitaria es importante mantener a salvo la salud de cada persona, pero aun así también es un punto importante el continuar con la educación. Para ello las IES han adoptado ciertas herramientas o plataformas [10].

En su mayoría sin estar preparados para la educación en línea, pero con coraje y empeño los profesores se esforzaron por comprender lo que significaba enseñar a distancia utilizando un entorno de aprendizaje completamente en línea, luchando por crear contenido que fuera atractivo y relevante, o experimentando con la evaluación digital. Al elegir jugar seguro y evitar riesgos importantes, la mayoría simplemente se limitó a replicar sus experiencias tradicionales en el aula, brindando conferencias en línea a través de sistemas de conferencias web, como Zoom, Skype, Microsoft Teams, Google Meet y WhatsApp, y a prácticas evaluativas basadas en exámenes en línea. Esta simplificación excesiva de las metodologías de enseñanza a distancia y en línea, ha dado como resultado un enfoque excesivamente basado en la entrega de contenidos, lo que devalúa el apoyo y la retroalimentación adecuadas, que son de suma importancia para asegurar el desempeño de los estudiantes [14].

Como han señalado los expertos, la mayoría de estas prácticas pueden caracterizarse mejor como una enseñanza remota de emergencia, que se define como "un cambio temporal de la entrega de instrucción a un modo de entrega alternativo debido a circunstancias de crisis" [27].

Algunos dispositivos y herramientas implementados para enfrentar la crisis, dan cuenta del alcance y utilidad de las tecnologías digitales y han marcado tendencia entre las principales decisiones adoptadas en las IES como la realidad virtual, los aprendizajes basados en video juegos, el e-learning, la inteligencia artificial, la educación on-line y mediada por tecnología móvil, además de recursos considerados como soporte a la viabilidad de éstas, como por ejemplo las impresoras digitales, entornos virtuales de enseñanza y aprendizaje (EVEA) y pizarras digitales interactivas, cuyo fin es el modelamiento dinámico y transformador de la gestión en las IES [17].

Todo este arsenal de herramientas innovadoras adaptado para enfrentar la crisis, o las estrategias pedagógico didácticas implementadas por las IES en muchas ocasiones, no atienden los efectos inmediatos, ni el impacto que producirá esta transición hacia la educación a distancia de emergencia. Otros impactos no menos importantes para los distintos actores, menos visibles y no documentados todavía, incluyen ámbitos como el socioemocional, el laboral, el financiero y, obviamente, sobre el funcionamiento del sistema en su conjunto [18].

Los aspectos antes mencionados motivaron este trabajo, en el que se realiza una revisión sistemática (RS), para analizar la selección de artículos, en el marco del modelo de Comunidad de indagación (CoI) de Garrison y Anderson [28], así a partir de sus dimensiones de: Presencia docente, Presencia cognitiva y Presencia social, se revisan las estrategias pedagógico-didácticas consideradas en los artículos que conforman el corpus de revisión.

De aquí en más este artículo se organiza de la siguiente manera: en la sección 2 se describe el modelo CoI y los tres tipos de presencia que este modelo enuncia. En la sección 3 se presenta la metodología empleada en la RS, en particular características de

los estudios seleccionados, aspectos del proceso de búsqueda y los criterios de inclusión y exclusión adoptados. Más adelante en la sección 4, se presenta una descripción de los artículos incluidos, se describen las estrategias pedagógico-didácticas adoptadas por las IES ante la interrupción de las actividades presenciales y se muestran los resultados del análisis a la luz de las dimensiones del Modelo de CoI y sus efectos sobre la comunidad educativa en las IES, así como los impactos, percepciones y problemáticas emergentes en la población estudiantil, a causa de las soluciones implementadas. La sección 5 concluye con reflexiones acerca de los modelos adoptados en las IES en Iberoamérica, sus implicancias y un posible contexto futuro postpandemia.

2 Consideraciones acerca del Modelo de CoI

El modelo de CoI (Community of Inquiry) es un marco teórico desarrollado por Garrison y Anderson [28].
Para que el aprendizaje en línea en una CoI sea posible, es necesaria la interrelación de tres elementos o presencias: cognitivos, docentes y sociales [29].
Presencia cognitiva. Indica hasta qué punto los estudiantes son capaces de construir significado a través de reflexión continua en una comunidad de investigación crítica [28], [30], a través de una comunicación sostenida [31]. El modelo propuesto identifica cuatro fases inmutables y no secuenciales en la presencia cognitiva: activación, exploración, integración y resolución [28], [30].
Presencia social. Es la capacidad de los participantes para proyectarse social y emocionalmente y como personas, para promover la comunicación directa entre individuos y para hacer la representación personal explícita. La presencia social marca una diferencia cualitativa entre una comunidad de investigación colaborativa y el proceso de meramente descargar información [30].
Presencia docente. Se define en el modelo CoI como el acto de diseñar, facilitar y orientar los procesos de enseñanza y aprendizaje para obtener los resultados previstos de acuerdo con las necesidades y capacidades de los estudiantes [32].

3 Metodología

3.1 Características de los estudios seleccionados

Se requirió que todos los estudios indagaran sobre la implementación de soluciones que dieran respuesta al contexto educativo de emergencia, a causa de la suspensión de las clases presenciales, considerando estudiantes de ES como población de interés, usando Tecnologías de la Información y la Comunicación (TIC) o Tecnologías del Aprendizaje y el Conocimiento (TAC) y haberse desarrollado en, al menos, un país iberoamericano. De la lectura y análisis del corpus seleccionado se intenta responder a las siguientes preguntas de investigación:

ISBN 978 -987-633-574-4

P1: ¿Cuáles son las estrategias pedagógico-didácticas adoptadas por las IES ante la interrupción de las actividades presenciales, a causa de la pandemia por COVID-19?
P2: ¿Cuál fue el alcance de dichas estrategias y que implicancias tienen en el marco de las presencias social, cognitiva y docente del modelo de CoI?
P3: ¿Cuál es la valoración de cada una de las presencias en la búsqueda de soluciones ante el paradigma de educación remota de emergencia?

3.2 Estrategia de búsqueda

El período de búsqueda se limitó a los años 2020 y 2021, lapso en el cual los centros de ES aludieron a un cierre total o parcial de sus puertas, dejando imposibilitado temporal o permanentemente la educación presencial [3], debido a la propagación del nuevo coronavirus SARS-CoV-2 y su enfermedad potencial COVID-19.
Se desarrolló una revisión bibliográfica y webgráfica de los artículos publicados, indagando en revistas y eventos académicos de Iberoamérica, preferentemente en español e inglés, como por ejemplo Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología (TE&ET), Congreso TE&ET, Revista Iberoamericana de Educación a Distancia RIED. Revista Virtualidad, Educación y Ciencia (UNC), Revista Question/Juventudes (FPyCS-UNLP), Jornadas de Educación a Distancia (UNLP), Repositorio Institucional de la Universidad Nacional de La Plata, Servicio de Difusión de la Creación Intelectual (SEDICI), Informes Ministerio de Educación de la Nación, Revista Estado y Políticas Públicas/Propuesta Educativa (FLACSO). IEEE-RITA, ACM, Journal Computers & Education (Elsevier), Journal NAER, American Journal of Distance Education (AJDE) y pertinentes, utilizando los motores de búsqueda Google Académico, eric.ed.gov, worldwidescience.org. y https://dialnet.unirioja.es/. Además, se incorporaron estudios sugeridos por expertos en la temática, y se realizó una búsqueda manual sobre las tablas de contenido de las revistas de Educación a Distancia (RED); Revista Electrónica de Investigación Educativa (REDIE) y Scientific Electronic Library Online (SciELO).
Los estudios incluyen experiencias llevadas a cabo en países de América Latina y España. Se analizaron trabajos que, editados en un país europeo, incluían problemáticas latinoamericanas y estadísticas de la región. Con respecto a la cadena empleada en el proceso, se utiliza un conjunto de términos en español e inglés, como indica la Tabla 1.

Tabla 1. Cadena de términos utilizada para la búsqueda

	Español	Inglés
A1	educación a distancia	distance education
A2	educación virtual	e learning
A3	aprendizaje electrónico	e learning
B1	cuarentena	quarantine
B2	pandemia	pandemic
B3	coronavirus	coronavirus emergency
C1	educación superior	higher education
D1	interacción	interaction
E1	paradigma de comunidad de indagación	community of inquiry paradigm
E2	modelo de comunidad de indagación	community of inquiry model

ISBN 978 -987-633-574-4

Las cadenas de búsqueda se corresponden a las siguientes expresiones booleanas:
(A1 OR A2 OR A3) AND (B1 OR B2 OR B3) AND (C1) AND (D1)
(A1 OR A2 OR A3) AND (B1 OR B2 OR B3) AND (C1) AND (D1) AND (E1 OR E2)
A partir de esta estrategia, se procedió a la selección de documentos, los cuales fueron evaluados en una primera etapa por medio de la lectura de título, resumen y fecha. Fueron seleccionados 127 artículos, según los criterios explicitados en relación a las preguntas de investigación planteadas para este trabajo de investigación y siguiendo los criterios de inclusión y exclusión mencionados en la Tabla 2, lo que determinó la lectura exhaustiva y completa del corpus seleccionado. El resultado fue la elección de 24 artículos presentados en este trabajo para su análisis y síntesis

Tabla 2. Criterios de inclusión y exclusión

CRITERIOS DE INCLUSIÓN	CRITERIOS DE EXCLUSIÓN
Estudios empíricos y reflexiones teóricas que indagan sobre del uso de EVEA y herramientas sincrónicas y asincrónicas en contextos educativos de ES durante la pandemia de COVID-19	Estudios empíricos y reflexiones teóricas que No incluyan el uso de EVEA y herramientas sincrónicas y asincrónicas en contextos educativos y NO estén ubicados en el período afectado
Estudios empíricos que indagan sobre la implementación de soluciones y modelos educativos basados en TIC y TAC en ES.	Estudios empíricos que NO indagan sobre la implementación de soluciones y modelos educativos basados en TIC y TAC en ES.

4 Resultados

4.1 Descripción de los artículos incluidos que conforman el corpus de la RS

En la Tabla 3 se detallan los 24 artículos seleccionados que estudian estrategias adoptadas durante la interrupción de actividades presenciales y que se los puede analizar a la luz del modelo de CoI.

Tabla 3. Artículos seleccionados que estudian estrategias adoptadas. Elaboración propia.

dimensión	Categorías	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Presencia social	Afecto		X		X					X					X							X			
	Comunicación		X		X					X					X							X			
	Cohesión		X		X					X					X							X			
Presencia cognitiva	Activación		X		X					X	X	X	X		X									X	
	Exploración		X							X		X	X		X									X	
	Integración		X							X		X	X		X									X	
	Resolución		X							X		X	X		X									X	
Presencia docente	Diseño	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Discurso		X		X				X	X	X	X	X	X	X	X			X	X		X		X	X
	Enseñanza		X		X				X	X	X	X	X	X	X	X			X	X		X		X	X

ISBN 978 -987-633-574-4

Investigaciones que ponen foco en el docente, [1], [7], [11], [18], [22], están orientados a las estrategias implementadas y a los recursos pedagógicos y didácticos utilizados ante la emergencia, se trata de encuestas y entrevistas que analizan con metodologías de investigación cuantitativas y cualitativas, los modelos adoptados. Las poblaciones estudiadas van desde 26 hasta 1237 docentes en universidades de Iberoamérica.
Investigaciones que incluyen una metodología de análisis documental, [5], [14], la primera de ellas se lleva a cabo en un escenario de estudio de educación universitaria en el primer semestre de 2020, en Perú, considerando más de veinte textos académicos que incluyen artículos, textos e informes relacionados con la educación universitaria. El segundo artículo es la presentación de un número especial de la Revista de Educación a Distancia (RED), ese número reúne cerca de una docena y media de artículos que representan una rica variedad de temas y enfoques. Los autores representan también un trasfondo cultural diverso, provenientes de países como Brasil, México y Colombia
Investigaciones que abordan IES [6], [10], [16], [17], incluyen una investigación con enfoque cualitativo en una muestra de 10 instituciones de ES en Colombia, una descripción de las metodologías adoptadas en 6 universidades de Ecuador, un resumen que proporciona datos interesantes sobre las tendencias de la ES mundial en el contexto de la pandemia que incluye 424 universidades de todo el mundo y por último, un estudio sobre contenidos web de 25 Universidades, de Brasil, Chile, Perú, México y Colombia.
Investigaciones que involucran estudiantes de IES [4], [12], [13], [15], [19], [21], [23], se llevaron a cabo mediante encuestas e incluyen metodologías descriptivas con análisis mixtos en poblaciones que van desde 41 a 548 estudiantes de diversas carreras en universidades de Iberoamérica.
Textos que incluyen reflexiones teóricas [2], [3], [8], [9], [20], [24], refieren a artículos de posición y reflexiones argumentativas a partir de encuestas y testimonios. Algunos de los principales temas abordados en estos artículos son, la tensión producida por la brecha creada a partir de los modelos adoptados, su implicancia en los estudiantes y sus familias, las herramientas de comunicación utilizadas y el imaginario social sobre la presencialidad versus la educación remota de emergencia.
Fueron seleccionados trabajos de México (34%), España (29%), Perú (13%) Argentina (8%), Cuba (8%), Colombia (4%), Ecuador (4%), como se ilustra en la Figura 1.

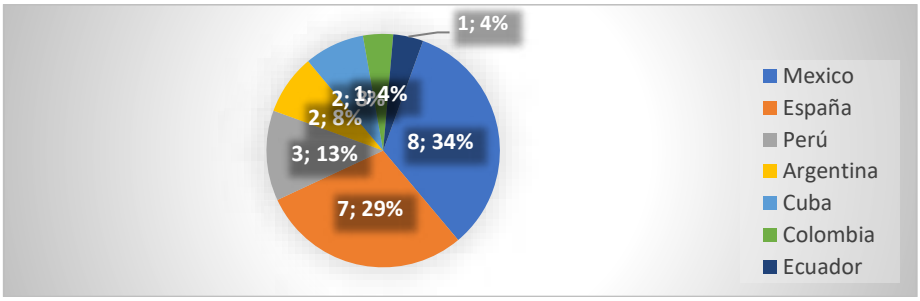


Figura 1. Artículos seleccionados según el país de origen de la publicación. Elaboración propia

4.2 Estrategias pedagógico-didácticas adoptadas por las IES ante la interrupción de las actividades presenciales.

Para dar respuesta a la primera pregunta de investigación se sintetizan las estrategias pedagógico-didácticas adoptadas por las IES, en el período mencionado y las acciones llevadas a cabo para enfrentar la crisis, según el corpus seleccionado.
La RS ha permitido identificar una serie de aportes con respecto a los modelos adoptados, siendo en su mayoría estudios que analizan los recursos pedagógico-didácticos implementados a causa de la interrupción de las clases presenciales.
Además, un caso pone de manifiesto los efectos inmediatos de la crisis, qué impactos está teniendo, y cómo el sector está respondiendo a los enormes desafíos planteados; al mismo tiempo, también incluye principios en los que debería basarse la planificación de la salida de la crisis [18].
Por otra parte, otro conjunto de artículos analiza la brecha que se hace presente a partir de la interrupción de la presencialidad, que no solo ha inducido a reinventar la docencia y reorganizar los procesos de enseñanza y aprendizaje, sino que, además, llegó a profundizar las condiciones estructurales de una población estudiantil en desventaja.
En referencia a los trabajos que investigan mediante encuestas y entrevistas a docentes y estudiantes, se estudia la percepción de estos grupos en relación con las habilidades y competencias necesarias para integrar herramientas de aprendizaje y comunicación, llegando a la conclusión que se adoptó una réplica del sistema de educación presencial a un sistema de educación a distancia, sin tener en cuenta la esencia de esta última.
Durante el confinamiento, en los niveles no universitarios se aportaron soluciones muy provisionales que, aunque dejarán elementos de reflexión para adoptar determinadas innovaciones, finalizarán gran parte de ellas una vez superada la crisis. Sin embargo, en la universidad probablemente será diferente. Las modalidades a distancia, digitales, en línea y flexibles van a ser aprovechadas una vez superada la pandemia [8].
Otra categoría encontrada en esta RS es el análisis comparativo entre diferentes escenarios, y pone el foco entre estudiantes universitarios, sobre la acción instruccional recibida en un escenario de enseñanza en línea antes de la cuarentena, y en un escenario de enseñanza remota de emergencia durante la pandemia [15]. Una reflexión interesante que aborda esta dimensión académica es que esta crisis emergente puso en descubierto las debilidades formativas en competencias digitales de los estudiantes y docentes [11].

4.3 Análisis de la RS en el marco de la comunidad de Indagación

En relación a la segunda y tercera preguntas de investigación, se valoran los modelos adoptados y se muestran los resultados obtenidos a la luz del modelo de CoI.
En los 24 artículos seleccionados se tiene en cuenta la dimensión Presencia docente y en especial la categoría diseño educativo y organización.
En el corpus seleccionado en esta RS, sólo 8 artículos trabajan la dimensión Presencia cognitiva, con énfasis en la categoría Activación.
Con respecto a la dimensión Presencia social, solamente 5 artículos de esta RS, ponen foco en lo que en el modelo de CoI está integrada por tres indicadores: la comunicación afectiva, la comunicación abierta y la cohesión del grupo.

ISBN 978 -987-633-574-4

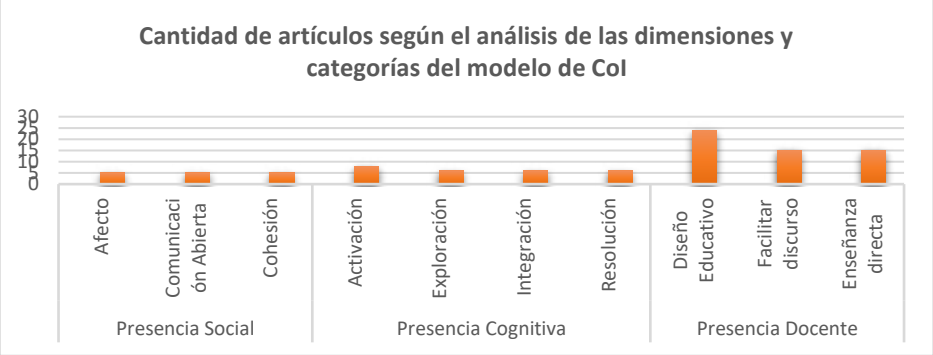


Figura 2. Análisis de la RS en el marco del modelo de CoI. Elaboración propia.

5. Discusión y Conclusiones

En los trabajos incluidos en el corpus de revisión se identifican investigaciones, que en su mayor parte hacen foco en la dimensión Presencia docente y en particular en la categoría diseño educativo y de organización.

En menor medida los trabajos revisados abordan la dimensión Presencia cognitiva, en cuanto a cómo los momentos del desarrollo de la propuesta educativa producen significado y tratan los contenidos desde una perspectiva crítica. No obstante, los que alcanzan a analizar este aspecto, enfatizan las diferencias de formación digital en docentes y estudiantes, poniendo en tensión la calidad de la educación en este contexto.

Se pone de manifiesto en las investigaciones incluidas en esta RS, la escasa producción sobre la dimensión Presencia social, y el análisis que haga foco en el afecto, la cohesión del grupo y la comunicación abierta, los cuales son indicadores que refuerzan el aprendizaje y mantiene una dinámica de relaciones sociales positiva.

Queda en evidencia, según esta investigación, la necesidad de repensar las estrategias pedagógico didácticas implementadas por las IES, en relación a la dimensión del modelo de CoI Presencia social, con vistas a un contexto post pandemia, donde muchas de estas metodologías adoptadas durante el confinamiento, seguirán vigentes.

Las derivaciones de este estudio, brindan un marco de referencia para futuros trabajos, que investiguen las experiencias llevadas a cabo por IES y sus implicancias, en un contexto futuro de propuestas semipresenciales post pandemia, como así también, las estrategias que favorezcan las interacciones y los procesos de enseñanza y aprendizaje colaborativos mediados por entornos virtuales.

Las conclusiones de este artículo, serán considerados en el proceso de avance de la tesis “Estudio de las interacciones entre profesores y estudiantes en entornos digitales, en el marco de continuidad pedagógica en contexto de cuarentena” para alcanzar el grado de Magister en Tecnología Informática Aplicada en Educación, de la Facultad de Informática de la UNLP. Además, abren el camino para profundizar en el diseño de propuestas educativas que tengan en cuenta las tres dimensiones, para considerar su incidencia en el proceso educativo.

Referencias

1. Amaya, A., Cantú, D. & Marreros, J. G. (2021). Análisis de las competencias didácticas virtuales en la impartición de clases universitarias en línea, durante contingencia del COVID-19. RED. Revista Educación a Distancia, 21(64). http://dx.doi.org/10.6018/red.426371
2. Cabero-Almenara, J.; Llorente-Cejudo, C. (2020). COVID-19: transformación radical de la digitalización en las instituciones universitarias. Campus Virtuales, 9(2), 25-34. (www.revistacampusvirtuales.es)
3. Canaza-Choque, F. A. (2020). Educación Superior en la cuarentena global: disrupciones y transiciones. Revista Digital De Investigación en Docencia Universitaria, 14 (2), 1-10.
4. Cano, S.; Collazos, C. A.; flórez-Aristizabal, L.; Moreira, f.; Ramírez, M. (2020). Experiencia del aprendizaje de la Educación Superior ante los cambios a nivel mundial a causa del COVID-19. Campus Virtuales, 9(2), 51-59. (www.revistacampusvirtuales.es)
5. Chiparra, W. E. M., Vásquez, K. M. C., Casco, R. J. E., Pajuelo, M. L. T., Jaramillo-Alejos, P. J., & Morillo-Flores, J. (2020). Disruption Caused by the COVID-19 Pandemic in Peruvian University Education. *International Journal of Higher Education*, *9*(9), 80-85.
6. Díaz-Guillen, P.A., Andrade Arango, Y., Hincapié Zuleta, A.M., y Uribe, A.P. (2021). Análisis del proceso metodológico en programas de educación superior en modalidad virtual. RED. Revista Educación a Distancia, 21(65). https://doi.org/10.6018/red.450711
7. Expósito, C. D., & Marsollier, R. G. (2020). Virtualidad y educación en tiempos de COVID-19. Un estudio empírico en Argentina.
8. García Aretio, L. (2021). COVID-19 y educación a distancia digital: preconfinamiento, confinamiento y posconfinamiento. RIED. Revista Iberoamericana de Educación a Distancia, 24(1), pp. 09-32. doi: http://dx.doi. org/10.5944/ried.24.1.28080
9. Gómez, Á. I. P. Capítulo II Repensar el sentido de la educación en tiempos de pandemia. La formación del pensamiento práctico, el cultivo de la sabiduría. *Reconstruyendo la educación superior a partir de la pandemia por COVID-19*, 32.
10. Indio Toala, J. M., León Tigua , M. X., López Farfán , F. A., & Muñiz Jaime, L. P. (2020). Educación virtual una alternativa en la educación superior ante la pandemia del covid-19 en Manabí. *UNESUM-Ciencias. Revista Científica Multidisciplinaria. ISSN 2602-8166*, *5*(1), 1-14. https://doi.org/10.47230/unesum-ciencias.v5.n1.2021.328
11. Cesar Ponce Gallegos, J., Angelica Toscano de la Torre B. and Silva Sprock, A. "Distance Education. An Emerging Strategy for Education in the pandemic COVID-19" 2020 XV Conferencia Latinoamericana de Tecnologias de Aprendizaje (LACLO), 2020, pp. 1-11, doi: 10.1109/LACLO50806.2020.9381137.
12. Lovón, M., & Cisneros, S. (2020). Repercusiones de las clases virtuales en los estudiantes universitarios en el contexto de la cuarentena por COVID-19: El caso de la PUCP. Propósitos y Representaciones, 8 (SPE3), e588. Doi: http://dx.doi.org/10.20511/pyr2020.v8nSPE3.588
13. Martin, M. V., & Vestfrid, P. (2020). Reinventar la enseñanza en tiempos de COVID-19. In *III Jornadas sobre las Prácticas Docentes en la Universidad Pública (Edición en línea, junio de 2020)*.
14. Moreira Teixeira, A. y Zapata-Ros, M. (2021). Introducción / presentación al número especial de RED "Transición de la educación convencional a la educación y al aprendizaje en línea, como consecuencia del COVID-19. Revista Educación a Distancia (RED), X(X). https://doi.org/10.6018/red.462271
15. Niño, S., Castellanos-Ramírez, J. C., y Patrón, F. (2021). Contraste de experiencias de estudiantes universitarios en dos escenarios educativos: enseñanza en línea vs. enseñanza remota de emergencia. Revista Educación a Distancia (RED), 21(65). https://doi.org/10.6018/red.440731
16. Ordorika, I. (2020). Pandemia y educación superior. *Revista de la educación superior*, *49*(194), 1-8.
17. Paredes-Chacín, A., Inciarte, A. y Walles-Peñaloza, D. (2020). Educación superior e

investigación en Latinoamérica: Transición al uso de tecnologías digitales por COVID-19. Revista de Ciencias Sociales (Ve), XXVI (3), 98-117.
18. Pedró, F. (2020). COVID-19 y educación superior en América Latina y el Caribe: efectos, impactos y recomendaciones políticas. *Análisis Carolina*, *36*(1), 1-15.
19. Pérez-López, E., Vázquez Atochero, A., y Cambero Rivero, S. (2021). Educación a distancia en tiempos de COVID-19: Análisis desde la perspectiva de los estudiantes universitarios. RIED. Revista Iberoamericana de Educación a Distancia, 24(1), pp. 331-350. doi: http://dx.doi.org/10.5944/ried.24.1.27855
20. Ramos, M. L. R., & Ruelas, M. R. Capítulo 6 La realidad de las personas con discapacidad frente a la COVID-19 en educación superior. *La pandemia de la COVID-19-19 como oportunidad para repensar la educación superior en México*, 119.
21. Roig-Vila, R., Urrea-Solano, M., y Merma-Molina, G. (2021). La comunicación en el aula universitaria en el contexto del COVID-19 a partir de la videoconferencia con Google Meet. RIED. Revista Iberoamericana de Educación a Distancia, 24(1), pp. 197-220. doi: http://dx.doi.org/10.5944/ried.24.1.27519
22. Sánchez Mendiola, M., Martínez Hernández, A. M., Torres Carrasco, R., de Agüero Servín, M., Hernández Romo, A. K., Benavides Lara, M. A., Rendón Cazales, V. J. y Jaimes Vergara, C. A. (2020). Retos educativos durante la pandemia de COVID-19: una encuesta a profesores de la unam. Revista Digital Universitaria (rdu) Vol. 21, núm. 3 mayo-junio. doi: http://doi.org/10.22201/codeic.16076079e.2020.v21n3.a12
23. Sureima, C. F., María del Carmen, M. G., Omara Margarita, G. O., Virgen, C. S., Ada María, D. A. F., & Ibis, R. G. (2021, March). El aula virtual como entorno virtual de aprendizaje durante la pandemia de COVID-19. In *aniversariocimeq2021*
24. Vialart Vidal, M. N. (2020). Estrategias didácticas para la virtualización del proceso enseñanza aprendizaje en tiempos de COVID-19. *Educación Médica Superior*, *34*(3).
25. Naciones Unidas [NU]. (2020). COVID-19 y educación superior: El camino a seguir después de la pandemia. Recuperado de https://www.un.org/es/impacto-académico/ COVID-19 y-educación-superior-el- camino-seguir-después-de-la-pandemia
26. IESALC, «Global University Network for Innovation. Publications. Report "COVID-19 y educación superior: De los efectos inmediatos al día después. Análisis de impactos, respuestas políticas y recomendaciones",» 6 abril 2020. [En línea]. Available: http://www.guninetwork.org/files/COVID-19-060420-es-2.pdf. [Último acceso: 05 julio 2021].
27. Hodges, C. B. , Moore, S. , Lockee, B. B. , Trust, T. , & Bond, M. A. (2020, March 27). The difference between emergency remote teaching and online learning. Educause Review. https://bit.ly/34tYI9r
28. Garrison, D.R., & Anderson, T. (2003). E–learning in 21st century: A framework for research and practice. London: Routledge Falmer.
29. Garrison, Randy; Cleveland-Innes, Martha y Fung, Tak S. (2010). "Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework", Th e Internet and Higher Education, vol. 13, núms. 1-2, pp. 31-36. doi: 10.1016/j.iheduc.2009.10.002.
30. Garrison, Randy; Anderson, Terry y Archer, Walter (2000). "Critical inquiry in a textbased environment: computer conferencing in higher education", Internet and Higher Education, vol. 11, núm. 2, pp. 1-14. doi: 10.1016/S1096-7516(00)00016-6.
31. Gunawardena, C. N., Lowe, C. E., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. Journal of Educational Computing Research, 17(4), 397–431.
32. Gutiérrez-Santiuste, E., Rodríguez-Sabiote, C., & Gallego-Arrufat, M-J. (2015). Cognitive presence through social and teaching presence in communities of inquiry: A correlational–predictive study. Australasian Journal of Educational Technology, 31(3), 349-362.

ISBN 978 -987-633-574-4

Identificación de brechas digitales en estudiantes de Relaciones Laborales. Una aproximación desde la virtualidad en 2021

Viviana R. Bercheñi y Sonia I. Mariño
Departamento de Informática
Facultad de Ciencias Exactas y Naturales y Agrimensura,
Universidad Nacional del Nordeste, 9 de Julio 1449
viviber@hotmail.com simarinio@yahoo.com,

Abstract La pandemia causada por COVID-19 originó en la sociedad diversos y abruptos cambios. Uno de ellos es el provocado en el sistema educativo que debió rediseñar sus estrategias sin precedentes transformando un modelo educativo presencial o hibrido a un modelo virtual sincrónico y asincrónico. El artículo indaga en torno a las brechas competenciales en dos asignaturas de grado distintas de la disciplina Informática. Se aplicó una metodología descriptiva transversal. Se diseñó un instrumento en-línea para relevar las percepciones de 45 participantes anónimos. Se procesaron los datos para su tratamiento estadístico. Los resultados relevan una alteración de las brechas identificadas en relación del período 2021 en relación a que durante el período 2020 la brecha predominante era la competencial, observando que, en el 2021, la brecha predominante pasó a ser la de uso. Analizando los resultados expuestos, se podría inferir que, transcurrido más de un año desde la declaración de la emergencia sanitaria, las brechas que se identificaron fueron básicamente de uso (en el sentido de relacionar la cantidad de integrantes por núcleo conviviente y la cantidad de dispositivos disponibles, y la calidad de la conectividad) y en menor medida se observó que solo el 6% del total de los alumnos declararon manejar inadecuadamente las herramientas que plantea el aula virtual. Estos datos permiten diseñar estrategias virtuales de aprendizaje superadoras como participación en foros de actividades, resolución de formularios de revisión teórico práctico, tareas, construcción de glosarios, entre otros, con el acompañamiento de clases sincrónicas y grabaciones asincrónicas de las exposiciones docentes, tratando de evitar el desgranamiento y deserción en espacios de Educación Superior.

Keywords: brechas digitales, educación superior, equidad, políticas redistributivas.

Introducción

La sociedad de la información y del conocimiento impactó significativamente en el modo de ser, hacer y conocer. Un cambio representativo deriva de las acciones y decisiones que se debieron desarrollar vertiginosamente desde las políticas públicas ante la pandemia del COVID-19. En este contexto las TIC desenvuelven un papel fundamental que permite continuar los procesos educativos.

Las universidades debieron afrontar nuevos desafíos atendiendo a los distintos roles que desempeña en la sociedad actual: docencia, investigación, extensión,

transferencia y compromiso contextual, focalizando sus esfuerzos en la pertinencia regional con proyección a la internalización.

En la sociedad y universidad moderna e innovadora, los nuevos modelos educativos remiten a las competencias. Competencia, etimológicamente procede del término latino "competere" [1]. Sevillano [2], citado en [3], define que la competencia "supone valores, actitudes y motivaciones, además de conocimientos, capacidades, habilidades y destrezas, todo formando parte del ser integral que es la persona, una persona inserta en un determinado contexto, en el que participa e interactúa, considerando también que aprende de manera constante y progresiva a lo largo de toda su vida".

En la literatura se identifican diversas competencias, denominadas como: Competencias Genéricas o Transversales, Intermedias, Generativas o Generales y las Competencias Específicas (Técnicas o Especializadas). Las competencias digitales o competencias en el uso de las tecnologías de la información y la comunicación (TIC) son competencias genéricas o transversales.

Levano-Francia et al. [4] mencionan que las "competencias digitales son entendidas a manera de concepto que ha generado diversas líneas de investigación que a luz de los nuevos avances tecnológicos en el rubro de las TICs". Estos ámbitos corresponden tanto a los educativos, laborales, sociales, culturales entre otros. En [5] se establece que "ser competente en el uso y apropiación de la tecnología mejora la competitividad y productividad de la población, ya que se desarrollan habilidades que permiten la solución de problemas a través del uso de la misma".

En Díaz Arce & Loyola Illescas [6] se expone una revisión de la literatura dirigida a diferenciar definiciones de competencias digitales, alfabetización informacional y alfabetización digital. En Ordóñez-Olmedo et al. [7], se estudian las competencias básicas digitales de 759 estudiantes universitarios españoles en el periodo 2012 – 2019. En Mariño y Bercheñi [8] y Bercheñi y Mariño [9] superando la limitación referida a la distinción de edades y sexo, variables contempladas en la presente investigación. Llanque Quispe [10] propone competencias para el aprendizaje permanente mediadas por TIC.

En Orosco et al. [11] se indagó en las competencias digitales en estudiantes no universitarios. Aplicar estudios como el expuesto en estudiantes del nivel secundario podría anticipar las acciones a diseñar para atender a nuevos ingresantes universitarios.

Por lo sintetizado en párrafos previos, indagar en torno a las competencias digitales de distintos colectivos permite identificar las brechas existentes, ajustar y rediseñar acciones con miras a aportar la apropiación de las TIC para lograr un desempeño efectivo y eficiente en los estudiantes universitarios. Proporciona información para el diseño de acciones de políticas públicas que tiendan a mitigar o reparar las consecuencias del impacto del pase de la presencialidad a la virtualidad.

La Organization for Economic Cooperation and Development [12] y la CEPAL [13], reconocen que la brecha digital se puede tratar desde diversas perspectivas. El concepto se modificó considerando que las tecnologías evolucionaron en un mundo complejo. En Bossolasco et al. [14]; Navarro et al. [15]; CEPAL [16] se relaciona el concepto con la apropiación social de las TIC, es decir, abordar un sentido amplio desde el estudio, lo laboral, lo cultural, entre otros aspectos.

García Peñalvo et al. [17] [18] señalan tres brechas digitales observables entre los

jóvenes estudiantes:

- Brecha de Acceso: La restricción viene dada cuando el individuo no tiene acceso a la tecnología: computadoras, dispositivos móviles de altas prestaciones, conectividad adecuada. En esta tipología, la falta de acceso puede tener origen económico o geográfico.
- Brecha de Uso: Se presenta cuando en los hogares hay conectividad adecuada, pero menos dispositivos que las personas que conviven, viéndose obligados a restringir el uso por horarios.
- Brecha Competencial: Es complementaria a las anteriores y se refiere a la falta de competencias adecuadas para utilizar todos los beneficios de las herramientas digitales y evitar sus riesgos o malas prácticas.

El objetivo de este artículo es identificar y dar a conocer el perfil digital de los estudiantes, información que podría anticipar acciones del plantel docente en las asignaturas siguientes del plan de estudio. En el artículo se sintetiza una indagación en torno a ciertos aspectos de las brechas digitales competenciales para inferir el cómo estas inciden en el acceso o uso en una EVA. Para lograr este cometido se aplicó una encuesta a estudiantes de dos asignaturas.

La relevancia del estudio se sitúa en disponer de conocimiento en torno a la brecha competencial y de uso, en un determinado contexto permite diseñar programas educativos ajustados a la realidad a la cual están destinados, considerando que impacta significativamente en las posibilidades de desarrollo actual y futuro [9].

Método

El diseño de investigación corresponde a un estudio transversal, de naturaleza descriptiva, exploratoria e interpretativa. Constó de las siguientes fases:

Fase 1. Revisión de la literatura en torno a brechas digitales, brechas competenciales y en particular aquellos estudios referidos a la educación superior pública.

Fase 2. Selección de las asignaturas objeto de indagación. Las asignaturas de grado elegidas corresponden a la carrera de Licenciatura en Relaciones Laborales de la Facultad de Ciencias Económicas de la Universidad Nacional del Nordeste. Las mismas no tratan específicamente temas de informática aplicada y no se corresponden a una carrera de la disciplina Informática. A continuación, se caracterizan sintéticamente:

- Historia Socio Económica General. La asignatura es de carácter cuatrimestral y se dicta en el primer año de la carrera. Su objetivo general es enfocar el estudio de la historia socioeconómica mundial y Argentina de forma interrelacionada y desde una perspectiva dinámica que comprende a los conceptos de procesos e impactos, revalorizando el instrumento estadístico como enriquecedor de su formación.
- Economía II: La asignatura es de carácter cuatrimestral y se dicta en el tercer año de la carrera. Su objetivo es el estudio del enfoque microeconómico, el comportamiento de las familias o unidades de consumo y empresas y el funcionamiento de los mercados de bienes y de factores en los que ellas operan. Utiliza modelos formales que mediante supuestos simplificadores explican y

ISBN 978 -987-633-574-4

predicen, el comportamiento de los individuos en su doble rol de consumidores y productores para arribar a conclusiones generales.

Fase 3. Diseño del instrumento de identificación de brechas digitales. Para generar una identificación preliminar cuantitativa orientada a inferir la existencia y el carácter de las brechas tecnológicas en estudiantes de las asignaturas mencionadas se aplicó como instrumento una encuesta diseñada a efectos de esta indagación con preguntas cerradas y abiertas basadas en [8] [9]. Para la captura y pre-procesamiento de los datos, se utilizó un formulario de Google. La muestra se conformó por los alumnos cursantes de las asignaturas objeto de estudio, quienes respondieron anónimamente la encuesta. El instrumento relevó el género y la edad, y constó de las siguientes preguntas:

- ¿Cuántos miembros viven en su mismo hogar y cuantas PC están disponibles?
- ¿Todos los miembros de su hogar utilizan la PC para actividades escolares o laborales?
- ¿Tiene buena conectividad en su hogar? ¿Se han presentado problemas de acceso al sistema de educación virtual institucional?
- Qué tipo de conexión utiliza
- Presenta experiencias educativas previas en relación con el uso de herramientas de comunicación asincrónicas.
- ¿Considera que maneja perfectamente las herramientas que le plantea el aula virtual?
- Presenta experiencias educativas previas en relación con el uso de herramientas de comunicación sincrónicas. Ej. Videoconferencias
- Además, se incluyeron las siguientes preguntas abiertas:
- Mencione debilidades de la educación mediada por TIC en el marco de esta situación nacional y mundial
- Mencione fortalezas de la educación mediada por TIC en el marco de esta situación nacional y mundial
- Mencione oportunidades de la educación mediada por TIC en el marco de esta situación nacional y mundial.

Fase 4. Análisis de la información para determinar tipologías de brechas digitales en estudiantes de universidades públicas a través del análisis comparativo cuantitativo de los datos expuestos. En el caso de identificación de brechas competenciales, se pretendió establecer aspectos integrados al EVA, para generar propuestas superadoras.

Resultados

En el ciclo lectivo 2021, como estrategia virtual de aprendizaje superadora, se incorporó al aula virtual, las clases grabadas de los temas a desarrollar por semana. Así, el horario de clases se destinó a resolver cuestiones concretas a partir del trabajo con el video correspondiente a la unidad.

Albalá Genol & Guido [19] mencionan que la brecha socioeducativa y la virtualización de la educación, denotan la necesidad de conocer al alumnado para proponer soluciones educativas mediadas por TIC en entornos complejos. Con fines

de identificar la existencia de brechas digitales competenciales existentes en las asignaturas elegidas, se implementó una encuesta en el primer cuatrimestre de 2021. Se sintetizan los hallazgos de una muestra de tamaño equivalente a n= 45. Del total de encuestados el 55% corresponden al sexo femenino. La edad mínima es de 18 y máxima de 44, siendo la moda 18 años, es decir, los estudiantes que se inician tienen mayor predisposición a responder.

En la Tabla I se representa -en promedio- la cantidad de miembros por unidad conviviente y la cantidad de PC disponibles en cada hogar. Se puede observar la existencia de menos de dos computadoras cada cuatro personas que conviven en el mismo hogar.

En referencia a la calidad en la conectividad que tienen acceso los estudiantes para el desarrollo de sus actividades académicas. Desde una perspectiva porcentual, manifestaron contar con una conectividad regular o mala en el 90% de los encuestados, realidad que impacta significativamente de manera desfavorable, impidiendo en muchos casos, el normal desarrollo del proceso de aprendizaje. Se indagó sobre la accesibilidad al sistema de educación mediado por aulas virtuales institucionales percibida por los alumnos, ascendiendo en el 74% de los estudiantes quienes manifestaron siempre y nunca, mientras que el 26% expresaron que nunca.

Tabla I: Cantidad de integrantes y PC por núcleo conviviente.

Carreras	*Por núcleo conviviente, promedio*	
	Integrantes	*Nro PC*
Promedio	4	1.66
Mediana	4	1
Desvío Estándar	1.77	2.29

Fuente: Elaboración propia, abril 2021.

Del total de alumnos encuestados, en más de un 49%, los docentes utilizaron plataformas digitales complementarias para desarrollar los contenidos, independientemente de los espacios digitales institucionales. La Tabla II expone la existencia de experticia en el manejo de herramientas disponibles en el aula virtual institucional, revelándose que el 94% se manifiesta positivamente. En la mayoría de los casos, los alumnos han contado con este tipo de experiencias relevadas. Se determinó que el 8% de los alumnos de ambas materias, carecían de experiencias educativas previas en el manejo de herramientas de comunicación sincrónicas. Respecto a las experiencias laborales previas en relación al uso de herramientas de comunicación sincrónica y asincrónica, en ambos casos cerca del 70% de los alumnos relevados, carecían de la misma (Tabla III).

Tabla II: Experticia en el manejo de herramientas que propone el aula virtual.

Experticia en herramientas TIC	*Muy bueno*	*Regular/ Malo*
Alumnos	94%	6%

Fuente: Elaboración propia, abril 2021.

ISBN 978 -987-633-574-4

Tabla III: Experiencias laborales previas en relación al uso de herramientas de comunicación sincrónica y asincrónica.

Experiencias laborales	*Sincrónicas*		*Asincrónicas*	
	Con	*Sin*	*Con*	*Sin*
Alumnos.	27%	73%	33%	67%

Fuente: Elaboración propia, abril 2021.

En referencia a las fortalezas, debilidades y oportunidades de la educación mediada por TIC e inferidas del procesamiento de los datos se da cuenta que:

- Las fortalezas explicitadas en la encuesta, giraron en torno a: i) Mayor adaptación al cambio, ii) Continuación de actividades académicas, evitando la pérdida del ciclo lectivo, mayor comunicación con el equipo de asignatura, bajo ausentismo, iii) Preservación de la salud y el tiempo disponible, iv) Mayor adiestramiento en utilización de TIC, que incide positivamente en futuros desempeños laborales.
- Entre las debilidades detectadas se mencionaron: i) Inclusión, dada la disponibilidad de recursos tecnológicos, ii) Conexión a internet o problemas eléctricos en algunos casos, iv) “Proceso educativo poco humano, a pesar de tener múltiples sesiones para transmitir dudas, o se carece de "contacto" humano para completar ese proceso educativo”, v) Mayor posibilidad de distraerse, vi) Dificultad en la concentración.
- Al consultar a los estudiantes sobre las oportunidades de la educación mediada por TIC, respondieron: i) Nuevas alternativas de capacitación. ii) Disminución de costos de traslado, iii) Ampliación de las posibilidades de acceder a la oferta educativa propuesta, iv) Refuerzo de contenidos a través de tutorías asincrónicas y acceso a recursos, v) Mayor adaptabilidad de la carrera a la disponibilidad de tiempos personales, vi) Adquisición de mayor entrenamiento para el futuro laboral en estas competencias, vii) Socialización, las charlas compartidas con el docente.

El escenario de pos-pandemia seguramente replanteará a la docencia su verdadera capacidad de adaptarse a cambios tecnológicos, optimizando el uso de TIC, para reuniones, clases, exámenes o encuentros virtuales entre pares o con alumnos, y eso exigirá un poder de resilencia de alumnos y docentes, que, como nunca antes, enfrentarán juntos los desafíos que plantea un paradigma impuesto para lo cual no se tuvo tiempo de contar con recursos materiales, financieros y humanos pertinentes para semejante hazaña quijotesca.

Analizando los resultados expuestos, se podría inferir que, transcurridos ya más de un año desde la declaración de la emergencia sanitaria, las brechas que se identificaron fueron básicamente de uso (en el sentido de relacionar la cantidad de integrantes por núcleo conviviente y la cantidad de dispositivos disponibles, y la calidad de la conectividad) y en menor medida se observó que solo el 6% del total de los alumnos declararon manejar inadecuadamente las herramientas que plantea el aula virtual. Si se comparan los hallazgos de 2021 con respecto a los expuestos en [8] [9].

En relación a la brecha competencial identificada y su vinculación con el diseño de estrategias en entornos virtuales de aprendizaje para evitar el desgranamiento escolar, En [20] y. [21] mencionan cuestiones en torno a las competencias digitales para continuar procesos de aprendizaje en tiempos de pandemia, y dan cuenta de algunas herramientas para mediar procesos de aprendizajes. En las asignaturas relevadas, durante el primer cuatrimestre del período 2021, las actividades se realizaron a través

de intervenciones sincrónicas y asincrónicas. Las primeras a través de Google Meet y Zoom. Para mediar las actividades asincrónicas, se trabajó con canales en Youtube habilitados a tal fin. Se utilizó además el aula virtual de la Universidad, (basada en la plataforma Moodle y Cisco Webex) para asegurar el acceso a la bibliografía necesaria, el material y las presentaciones digitales con audios explicativos de temas vinculados al diseño y desarrollo de los contenidos. Se activaron además foros de actividades, links de temáticas vinculadas a unidades específicas, formularios Google para la revisión de contenidos teórico prácticos, y grabaciones de las clases brindadas por los docentes de manera sincrónica.

Estos datos permiten re-diseñar estrategias superadoras en torno a los EVA y evitar el desgranamiento y deserción en espacios de Educación Superior.

Conclusiones

En el artículo se identifican y analizan algunas cuestiones en torno a brechas digitales para determinar la brecha competencial relacionada con el uso de EVA, los datos se capturaron a partir de las percepciones de los estudiantes de dos asignaturas de la carrera Licenciatura en Relaciones Laborales, en el Nordeste Argentino. La experiencia se relevó a inicios del ciclo lectivo 2021 desarrolladas en la modalidad a distancia en el contexto de la pandemia causada por COVID-19.

En referencia a la brecha de acceso se considera pertinente relevar cuestiones referentes a la residencia de los estudiantes y el proveedor de Internet, variables que implícitamente se detectaron como influyentes en las respuestas sintetizadas en las dos primeras preguntas.

La brecha de uso podría estar signada por la disponibilidad de dispositivos de acceso, así como de otros complementos que soportan la educación mediada por TIC y que los sujetos debieron adquirir para afrontar esta modalidad de educación virtual. Cabe aclarar que antes de esta situación sanitaria, los estudiantes asistían a los laboratorios o espacios de la facultad para consultar repositorios, realizar actividades en plataformas entre otras actividades. Algún indicio sobre la existencia de esta brecha se observa en la evaluación de la conectividad en sus domicilios particulares.

Se destacan las numerosas menciones positivas a las grabaciones de las clases disponibles y accesibles desde el aula virtual, que los independiza en la asistencia. Sin embargo, se menciona dificultad en la compresión en torno a ciertos temas dados en estas asignaturas, esto último podría relacionarse con la ausencia a las clases previstas y que aun cuando estos contenidos contribuyen en su formación profesional, no se tratan de temas disciplinares para este grupo de estudiantes.

Respecto a la brecha competencial y en concordancia con [17] [18], en el artículo, se destaca que las experiencias previas en relación con los estudios y desempeño laboral en menor medida contribuyeron a disminuir las brechas competenciales e incidieron en la finalización exitosa del cursado y aprobación de las asignaturas consideradas en este estudio conforme se observa en la Experticia en el manejo de herramientas que propone el aula virtual y las experiencias educativas y laborales previas en el uso de las mismas. Esto último puede justificarse dado que no todos los estudiantes se desempeñan laboralmente o han contado con experiencias previas de

interacción en la modalidad virtual. Con miras a profundizar en esta brecha, se sostiene la necesidad de relevar cuestiones vinculadas con una diversidad de objetivos ligados al aprendizaje significativo en nuevos escenarios educativos como los expuestos en [22].

Referencias

[1] M. Sanchez Diaz, Las competencias desde la perspectiva informacional: apuntes introductorios a nivel terminológico y conceptual, escenarios e iniciativas, Ci. Inf., Brasília, v. 37, n. 1, p. 107-120, jan./abr. 2008

[2] E. López Gómez, En torno al concepto de competencia: un análisis de fuentes Profesorado. Revista de Currículum y Formación de Profesorado, vol. 20, núm. 1, enero-abril, 2016, pp. 311-322

[3] M. L. Sevillano, (Dir.) (2009). Competencias para el uso de herramientas virtuales en la vida, trabajo y formación permanentes. Madrid: Pearson, Prentice Hall.

[4] L. Levano-Francia, S. Sanchez Diaz, P. Guillén-Aparicio, S. Tello-Cabello, N. Herrera-Paico, & Z. Collantes-Inga, Competencias digitales y educación. Propósitos y Representaciones, 7(2), 569-588. 2019, https://dx.doi.org/10.20511/pyr2019.v7n2.329

[5] D. González Campos, F. Olarte Dussán, & J. Corredor Aristizabal, La alfabetización tecnológica: de la informática al desarrollo de competencias tecnológicas. Estudios pedagógicos (Valdivia), 43(1), 193-212. 2017, https://dx.doi.org/10.4067/S0718-07052017000100012

[6] D. Díaz-Arce & E. Loyola-Illescas, Competencias digitales en el contexto COVID 19: una mirada desde la educación. Revista Innova Educación, 3(1), 120-150. 2021, https://doi.org/10.35622/j.rie.2021.01.006

[7] E. Ordóñez-Olmedo, E. Vázquez-Cano, S. Arias-Sánchez & E. López-Meneses, Las Competencias en el uso de las Tecnologías de la Información y la Comunicación en el alumnado universitario. Pixel-Bit. Revista de Medios y Educación, 60, 153-167. 2021, https://doi.org/10.12795/pixelbit.74860

[8] S. I. Mariño & V. R Bercheñi. "Identificación de brechas digitales en pandemia: dos experiencias de grados superiores en la disciplina Informática." Mendive. Revista de Educación 18.4 910-922. 2020 http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S18157696202000040091O

[9] V. Bercheñi y S. I. Mariño, Identificación de brechas digitales en pandemia: Experiencias en carreras de grado de Facultades de la Universidad Nacional del Nordeste. Argentina, SISCOM, 2021.

[10] E. Llanque Quispe, Competencias para el aprendizaje permanente con el uso de TIC. Rev Inv Tec, La Paz, v. 3, n. 2, dic. 2015. http://www.revistasbolivianas.org.bo/scielo.php?script=sci_arttext&pid=S2306-05222015000100007&lng=es&nrm=iso

[11] F. Orosco, J. R., Gómez Galindo, W., R. Pomasunco Huaytalla, E. Salgado Samaniego, & R. C. Álvarez Casabona, Competencias digitales en estudiantes de educación secundaria de una provincia del centro del Perú. Revista Educación, 45(1), 51-69, 2021. https://dx.doi.org/10.15517/revedu.v45i1.41296

[12] OCDE Organization for Economic Cooperation and Development, 2001

[13] CEPAL https://www.cepal.org

[14] M. L. Bossolasco; E. E. Enrico; B. A Casanova. y, R. J. Enrico. Análisis de brechas de accesibilidad, uso y apropiación de las TIC en aspirantes al nivel superior universitario. Revista Virtu@lmente, 5(1), 38-49. 2017.

[15] CEPAL, La educación en tiempos de la pandemia de COVID-19, https://www.cepal.org/es/publicaciones/45904-la-educacion-tiempos-la-pandemia-covid-19
[16] D. A. Navarro, et al. "La brecha digital una revisión conceptual y aportaciones metodológicas para su estudio en México." Entreciencias: diálogos en la sociedad del conocimiento 6.16 (2018): 49-64. DOI: https://doi.org/10.22201/enesl.20078064e.2018.16.62611
[17] F. J. García-Peñalvo, A. Corell V. Abella-García, M. Grande, "La evaluación online en la educación superior en tiempos de la COVID-19." Education in the knowledge society, vol. 21 pp 26, 2020, DOI: https://doi.org/10.14201/eks.23086
[18] F. J. García-Peñalvo, El sistema universitario ante la COVID-19: Corto, medio y largo plazo. En: Universídad. https://bit.ly/2YPUeXU.
[19] M. A. Albalá Genol, J. I. Guido La brecha socioeducativa derivada del Covid-19: posibles abordajes desde el marco de la justicia social, Revista Latinoamericana de Estudios Educativos (México), vol. L, núm. Esp.-, 173-194, 2020, DOI: https://doi.org/10.48102/rlee.2020.50.ESPECIAL.101
[20] M. del C. Crespo Argudo & M. C. Palaguachi Tenecela. Educación con Tecnología en una Pandemia: Breve Análisis. Revista Scientific, 5(17), 292–310, 2020. https://doi.org/10.29394/Scientific.issn.2542-2987.2020.5.17.16.292-310
[21] F. H. Velasco Tutivén, J. E. Lecaro Castro, G. Y. Correa Pachay, F. A. García Quinto, N. del R. Mota Villamar, C. A. Moreno Pérez & J. M. Tulcán Muñoz, La brecha digital en el proceso de aprendizaje durante tiempos de pandemia. Ciencia Latina Revista Científica Multidisciplinar, 5(3), 3096-3107, 2021. https://doi.org/10.37811/cl_rcm.v5i3.515
[22] S Cueva Gaibor & D Abraham "La tecnología educativa en tiempos de crisis." Conrado 16.74: 341-348. 2020 de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1990-86442020000300341&lng=es&tlng=es.

ISBN 978 -987-633-574-4

Una guía de Accesibilidad Web para portales educativos. La revisión de usuarios

Verónica K. Pagnoni[1], Sonia I. Mariño[2]
1 Instituto Superior de Formación Docente Bella Vista, Bella Vista, Corrientes, Argentina
2 Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina
vero_pagnoni@hotmail.com, simarinio@yahoo.com

Abstract. La educación inclusiva debe asegurar el acceso a los contenidos. Una alternativa/ cuestión es disponer de sitios web accesibles. Las plataformas educativas surgieron y se afianzan como una estrategia que elimina barreras, asegurar la accesibilidad es imprescindible. El estudio se encuadra en una investigación cuantitativa descriptiva, que caracteriza la Accesibilidad Web de las páginas evaluadas de una plataforma para formadores, a través de la percepción de sus usuarios. Se diseñó una guía de revisión manual basada en otras propuestas y siguiendo la WCAG 2.0. Se describen las fases que comprendió el diseño de la evaluación. La muestra ascendió a 101 usuarios del ciclo lectivo 2019. Los resultados dan cuenta de que la mayoría de los inconvenientes corresponden a los principios Perceptible y Comprensible, y la importancia de incluir en los estudios el género, edad y dispositivos utilizados para conectarse y determinar los potenciales problemas de visión. Finalmente, los hallazgos sostienen la necesidad de continuar estudios como los expuestos y proponer mejoras orientadas a los usuarios de las TI para fidelizarlos.

Keywords: Accesibilidad Web, Participación de usuarios, Ambiente de aprendizaje, Inclusión digital, Diseño inclusivo, Discapacidad visual.

1. Introducción

En el ámbito de la educación, se ha impuesto con fuerza en los últimos años el modelo de "educación inclusiva", como superador del modelo integrador preexistente. Si bien no existe un acuerdo respecto de los significados de estos conceptos. Los teóricos catalogan a la educación inclusiva como la atención educativa de calidad a todos los estudiantes, pero en la práctica dependerá de los contextos particulares [1]. De este modo, la aplicación de la educación inclusiva es compleja, especialmente en los contextos donde son preponderantes modelos que se contraponen a sus principios [2].

Considerando las bases de la educación inclusiva y las dimensiones de brecha digital, se debe resaltar que los ambientes educativos virtuales pueden favorecer la inclusión de las personas en condición de discapacidad, pero también pueden provocar su exclusión [3].

La accesibilidad puede ser considerada en un sentido amplio como el conjunto de

ISBN 978 -987-633-574-4

características de que debe disponer un entorno, producto o servicio, para ser utilizado en condiciones de comodidad, seguridad e igualdad por parte de todas las personas [4]. La accesibilidad está relacionada al Diseño Universal, que consiste en la posibilidad que tiene cualquier persona de tener a su alcance los bienes y servicios ofrecidos en la sociedad [5].

La Accesibilidad Web, es entendida como la capacidad de acceso y contenidos digitales por todas las personas, sin importar las discapacidades que puedan poseer y de las características de su contexto [6].

El Consorcio World Wide Web (W3C) definió las Pautas de Accesibilidad para el Contenido Web (WCAG) [7]. Las WCAG se componen de principios, pautas y criterios de conformidad que permiten clasificar la accesibilidad de un contenido web en tres niveles: A, AA y AAA [8]-

En la actualidad existe una gran necesidad de desarrollar plataformas y diseñar materiales educativos accesibles que se correspondan a las pautas establecidas por el Diseño Universal y las normas W3C [9], [10], [11].

2. Metodología

El estudio se encuadra en una investigación cuantitativa descriptiva, se busca caracterizar la Accesibilidad Web de las páginas evaluadas a través de la percepción de los usuarios. Se aplicaron criterios y procedimientos sistemáticos [12] como los expuestos por el WCAG 2.0. El diseño de la evaluación, constó de las siguientes fases:

Fase 1: Profundización de los aspectos teóricos referentes a accesibilidad web:

- Definición de destinatarios: se abordó la accesibilidad web considerando la discapacidad visual de los potenciales usuarios, centrada en el perfil docente.
- Relevamiento de proyectos similares desarrollados en el dominio de la AW en la educación superior, sintetizados en la Introducción.

Fase 2: Definición y aplicación de una metodología para el abordaje empírico del tema:

- Estudio y elección de estándares referentes a la accesibilidad web: Se seleccionó la norma WC3 en su versión WCAG 2.0.
- Estudio y elección de herramientas para la medición de la accesibilidad web. Se optó por una Guía de revisión manual diseñadas para la AW: a fines de describir la muestra, se relevaron ciertas características de los usuarios participantes tales como: edad, género, tipo de dispositivo que utiliza para conectarse y si cuenta con algún problema visual.
- Configuración del equipo software y hardware: cada usuario utilizó su dispositivo.
- Selección de las páginas web a evaluar: Para realizar las evaluaciones validaciones se consideraron tres páginas representativas de la plataforma educativa elegida. Como se mencionó en [13], el docente que realiza una formación, debe acceder al contenido de cada una de estas páginas, para las clases

ISBN 978 -987-633-574-4

virtuales y realizar las actividades propuestas en la formación. La Página 1 es la página inicial, contiene la bienvenida a los cursantes, un menú para ordenar y filtrar las propuestas académicas y un listado de formaciones disponibles. La Página 2 permite que un cursante inicie sesión y acceda a las inscripciones de los cursos y a la plataforma virtual. La Página 3 corresponde a la pantalla de inicio del aula virtual.

- Evaluación de las páginas seleccionadas utilizando las Guía de revisión manual diseñada a partir de los aportes de otros autores.
- Procesamiento de los datos: Se utilizó el software SPSS v.22.

Fase 3 Análisis de los resultados y propuestas de mejoras

3. Resultados

Los resultados se exponen considerando la Guía de revisión manual diseñada a partir de aportes de otros autores y su validación con una muestra de docentes en el año 2019.

3.1 Propuesta de guía de revisión.

La guía elaborada se compone de dos secciones. La primera caracteriza a los participantes, y la segunda una serie de interrogantes para evaluar cada una de las páginas. Para capturar la percepción de los potenciales usuarios de la plataforma, se diseñó una guía basada en los criterios planteados por [14], [15], [16], y [17]. La propuesta de evaluación se sustenta en preguntas simples cuyas respuestas pueden resolverse navegando por el contenido web. Los interrogantes abarcan los principales criterios de la WCAG 2.0 siendo las posibles opciones de respuesta: Si, Medianamente y No. En este estudio se consideraron los niveles de conformidad A (menos exigente), AA y AAA (más exigente) para establecer los pesos a cada incumplimiento de un criterio. Las preguntas se seleccionaron teniendo en cuenta el contenido de las páginas web a evaluar, como se expone en la Tabla 1.

Tabla 1. Criterios a evaluar e interrogantes

Criterio y nivel de conformidad	Pregunta
Criterio de conformidad 1.1.1 Contenido no textual - Nivel A	1) Si existen imágenes ¿si no pudieran visualizarse, se mantiene la información de la página?
Criterio de conformidad 1.4.1 Uso del Color - Nivel A	2) Si la página cuenta con textos o elementos con color ¿si estos carecieran de ese color se mantiene la información que transmiten?
Criterio de conformidad 1.4.3 Contraste (mínimo) - Nivel AA	3) El contraste entre la fuente (de los links, el texto en general, el texto de los botones) y el fondo ¿es suficiente para comprender con claridad lo que está escrito?
Criterio de conformidad 1.4.4 Cambiar el tamaño del texto - Nivel AA	4) El tamaño de la fuente (de los links, el texto en general, el texto de los botones) con la página visible a un 200% (sin agrandar o achicar) ¿ayuda a su visualización y comprensión?

ISBN 978 -987-633-574-4

Criterio de conformidad 2.4.2 Título de la página - Nivel A	5) ¿El título de la página es claro y comprensible, considerando el contenido de la misma?
Criterio de conformidad 2.1.2 Sin trampa de teclado – Nivel A	6) Si se utiliza la tecla TAB para moverse entre los elementos de la página ¿se puede acceder a todos?
Criterio de conformidad 2.4.7 Enfoque visible - Nivel AA	7) Cuando se presiona la tecla TAB para recorrer los elementos de la página ¿se puede ver alguna señal visual que muestre qué elemento está activo?
Criterio de conformidad 2.4.9 Propósito del vínculo (sólo vínculo) - Nivel AAA	8) Considerando el nombre de cada link ¿se comprende sin ambigüedades a dónde llevan estos enlaces?
Criterios de conformidad 1.3.1 Información y relaciones - Nivel A	9) En un formulario ¿cada dato a ingresar está acompañado de un texto que lo identifica?
Criterio de conformidad 2.2.1 Tiempo ajustable - Nivel A	10) ¿El tiempo ofrecido para completar los datos es suficiente?
Criterio de conformidad 3.3.2 Etiquetas o Instrucciones - Nivel A	11) ¿Los textos que acompañan a los espacios en blanco para introducir datos son suficientemente claros para comprender exactamente qué dato se espera que se coloque?
Criterio de conformidad 3.3.2 Etiquetas o Instrucciones - Nivel A	12) ¿Es fácil comprender cómo enviar los datos, considerando el color y texto del botón de envío?
Criterio de conformidad 3.3.3 Sugerencia de error - Nivel AA	13) En el caso de que se requiera la introducción de datos ¿el mensaje de incorporación de datos erróneos orienta al usuario para saber cuál fue el error cometido?

3.2 Validación Guía de revisión manual propuesta. Experiencia en el año 2019

Para determinar la percepción visual del perfil docente de la plataforma seleccionada como objeto de estudio, el estudio descriptivo se basó en las respuestas de 101 docentes, usuarios de la plataforma y que conforman la muestra. Los datos revelan que quienes respondieron mayoritariamente pertenecen al género femenino (80, 2%), y que el 74,3% cuentan con menos de 40 años. Del total, 88 profesores encuestados (87,1%) utilizan computadora y 85 (84,2%) teléfonos digitales para conectarse. En tanto en su mayoría (69 casos que representan el 68,3%) se valen de ambos dispositivos indistintamente. Se determinó, que casi un 40% de los docentes que respondieron a la encuesta tienen alguna discapacidad visual. Cabe destacar que los de 38 profesores que poseen alguna discapacidad visual, en su mayoría sufren de miopía (26 casos), y en la mitad de éstos se presenta conjuntamente con astigmatismo.

Las Tablas 2, 3 y 4 ilustran la sistematización de los datos correspondientes a la percepción de los usuarios. La información se capturó al aplicar la guía propuesta (Tabla 1) a las Páginas 1, 2 y 3, respectivamente. Se muestran en estas tablas el identificador de las preguntas utilizadas para la evaluación, las respuestas obtenidas de los encuestados expresadas en valores absolutos y porcentajes.

ISBN 978 -987-633-574-4

Tabla 2. Evaluación de la Página 1

Pregunta	Cantidad de respuestas					
	Si		Medianamente		No	
	Absoluto	%	Absoluto	%	Absoluto	%
1)	78	77,2	15	14,9	8	7,9
2)	70	69,3	25	24,6	6	5,9
3)	83	82,2	16	15,8	2	2
4)	78	77,2	19	18,8	4	4
5)	89	88,1	12	11,9	0	0
6)	89	88,1	12	11,9	0	0
7)	89	88,1	12	11,9	0	0
8)	81	80,2	20	19,8	0	0

Tabla 3- Evaluación de la Página 2

Pregunta	Cantidad de respuestas					
	Si		Medianamente		No	
	Absoluto	%	Absoluto	%	Absoluto	%
1)	82	81,2	17	16,8	2	2
2)	78	77,2	19	18,8	4	4
3)	82	81,2	18	17,8	1	1
4)	79	78,2	20	19,8	2	2
5)	89	88,1	11	10,9	1	1
6)	89	88,1	12	11,9	0	0
7)	89	88,1	12	11,9	0	0
8)	81	80,2	20	19,8	0	0
9)	101	100	0	0	0	0
10)	101	100	0	0	0	0
11)	86	85,1	14	13,9	1	1
12)	89	88,1	12	11,9	0	0
13)	0	0	0	0	101	100

Tabla 4- Evaluación de la Página 3

Pregunta	Cantidad de respuestas					
	Si		Medianamente		No	
	Absoluto	%	Absoluto	%	Absoluto	%
1)	82	81,2	17	16,8	2	2
2)	78	77,2	19	18,8	4	4
3)	82	81,2	18	17,8	1	1
4)	79	78,2	20	19,8	2	2
5)	89	88,1	11	10,9	1	1
6)	89	88,1	12	11,9	0	0
7)	89	88,1	12	11,9	0	0
8)	81	80,2	20	19,8	0	0

ISBN 978 -987-633-574-4

4. Discusiones

El análisis de los datos relevados concernientes a género, edad, dispositivo que utiliza para conectarse y si posee un problema visual, da cuenta que: i) La edad de la mayoría de los usuarios consultados oscila entre los 21 y 30 años, con un 33% del total, mientras que un 80% tiene menos de 50 años. ii) Un 53% de las personas encuestadas son de sexo femenino. iii) Todos los usuarios usan la PC, además un 80% el celular y un 7% la tablet, para acceder a la plataforma analizada. iv) El 60% de las personas consultadas tienen algún tipo de problema visual.

Estos datos demuestran que muchas personas tienen una visión natural imperfecta (60% de la muestra), según [18] esta discapacidad debe considerarse dado que es fácilmente corregida al utilizar anteojos. Se debe contemplar aquí que a pesar de que estas falencias se pueden mejorar con una tecnología de apoyo visual muchas personas carecen de anteojos adecuados. Así, también el uso del celular para navegar por el contenido web (un 80% de los encuestados utiliza tecnología móvil) dificulta aún más la posibilidad de ver y comprender la información. Es por ello el contenido web debe estar preparado para una persona con estas deficiencias, posea o no una tecnología correctiva.

Los resultados considerando las respuestas “medianamente” y “no”, permitieron observar que:

- En las Páginas 1 y 3 el criterio que más aportó a la inaccesibilidad fue el 1.4.1 Uso del Color del Nivel de conformidad A, referido a que el color no debe ser usado como el único medio visual de transmitir información. El cumplimiento de este criterio ayuda a los usuarios con visión parcial a menudo experimentan una visión limitada de los colores, así como a los que tienen inconvenientes para distinguir entre colores [19]. Asimismo, en la Página 3 este criterio aporta considerablemente a la inaccesibilidad.
- En la Página 2 el criterio que menos se cumple es el 3.3.3 Sugerencia de error del Nivel de conformidad AA con un 48,13% del total. Éste se refiere a la existencia de sugerencias al detectarse errores en la entrada de datos. En particular, ayuda a los usuarios ciegos o con problemas de visión a que comprendan más fácilmente la naturaleza del error de entrada y cómo corregirlo[20].
- En la Página 1 también se debe considerar en las páginas la falta de cumplimiento del criterio 1.1.1 Contenido no textual correspondiente al Nivel de conformidad A, el cual hace referencia a que todo contenido no textual tiene una alternativa textual que cumple el mismo objetivo. Es importante el acatamiento de este criterio, dado que facilita a las personas con deficiencias visuales puesto que las tecnologías de asistencia pueden leer texto en voz alta, presentarlo visualmente o convertirlo a Braille [21].
- En las páginas 1 y 3 se incumple considerablemente el criterio 1.4.4 Cambio de tamaño del texto, cuyo éxito es indispensable para asegurar que el texto procesado visualmente pueda ser aumentado de tamaño para que quienes sean personas disminuidas visuales puedan leer sin necesidad de usar ayudas técnicas [22].

Los hallazgos comunicados en este artículo, contribuyen a la AW desde una mirada de la responsabilidad social, con miras a asegurar el acceso universal de los

contenidos [23]. En particular, a los docentes del nivel superior no universitario, formadores de formadores, quienes son transmisores de información y conocimiento.

5. Conclusiones

El artículo expuso el análisis de tres páginas web de un EVEA sustentado en la percepción visual de una muestra de usuarios, contemplando los principios de la norma WCAG 2.0. El análisis de la muestra seleccionada da cuenta que ninguna de las páginas evaluadas cumplimenta totalmente los principios de Accesibilidad Web.

La revisión de los antecedentes mencionados, sustentó la definición de una guía para determinar la percepción visual de usuarios en torno a accesibilidad web en contextos de educación superior no universitaria. El mismo integra las fortalezas de otras guías analizadas. En particular, esta guía aporta practicidad y especificidad, el usuario puede contestar las preguntas de manera simple recorriendo cada página.

Los datos relevados dan cuenta de la importancia de incluir en los estudios el género, edad y dispositivos utilizados para conectarse y determinar los potenciales problemas de visión a fin de proponer estrategias superadoras. Un análisis derivado de este estudio, permite apreciar que si bien la mayoría de los usuarios encuestados se pueden considerar personas jóvenes, un 60% de ellos padecen problemas de visión. Los hallazgos en torno a la investigación permitieron determinar la presencia de problemas visuales y su representación, motivo que podría originar futuras intervenciones con la finalidad de fidelizar a estos usuarios.

Las páginas evaluadas no cumplimentan el Nivel de Conformidad A, y sus obstáculos más graves se corresponden a los principios Perceptible y Comprensible, los cuales son indispensables para asegurar la AW del contenido.

6. Referencias

[1] I. García Cedillo, S. Romero Contreras, C. L. Aguilar Orozco, K. A. Hernández & D. C. Rodríguez Ugalde. Terminología internacional sobre la educación inclusiva. Revista. Actualidades Investigativas en Educación. Vol.13 N°1. San José Jan./Apr. 2013. ISSN 1409-4703.

[2] C. Duk & F. J. Murillo. El mensaje de la educación inclusiva es simple, pero su puesta en práctica es compleja. Revista Latinoamericana de Educación Inclusiva, 12(1), 11, 2018.

[3] S. J. Hernández Otálora, O. M. Quejada Durán & G. M. Díaz Guía metodológica para el desarrollo de ambientes educativos virtuales accesibles: una visión desde un enfoque sistémico. Digital Education Review - N 29. 2016.

[4] O. Palma, X. Soto, C. Barría, X. Lucero, D. Mella, Y. Santana & E. Seguel, Estudio cualitativo del proceso de adaptación e inclusión de un grupo de estudiantes de educación superior con discapacidad de la Universidad de Magallanes. Magallania (Punta Arenas), 44(2), 131-158. 2016.

[5] M. F. Chamorro Cristaldo, Web 2.0, accesibilidad e inclusión social aplicada a las bibliotecas. ACADEMO Revista de Investigación en Ciencias Sociales y Humanidades, 2(1). 2015 Recuperado a partir de

https://revistacientifica.uamericana.edu.py/index.php/academo/article/view/15.

[6] Y. Stable Rodríguez & C. A. Sam Anlas. National Libraries and Web Accessibility. Situation in Latin America. Revista Interamericana de Bibliotecología, 41(3), 253-265. 2018.

[7] L. F. Londoño Rojas, V. Tabares Morales, M. R. Bez & N. D. Duque Mendez. Análisis comparativo de guías para el desarrollo web accesible. Ciencia e Ingeniería Neogranadina, 28(1), 101-115. 2017.

[8] N. Duque, J. Flores, & N. Castaño. Accesibilidad en sitios web colombianos. *Ingeniería E Innovación, 2*(1), 34-41. 2014.

[9] Abid Ismail, K.S. Kuppusamy, Web accessibility investigation and identification of major issues of higher education websites with statistical measures: A case study of college websites, Journal of King Saud University - Computer and Information Sciences, 2019.

[10] C M. Carvajal. Evaluación de accesibilidad web de las universidades chilenas. Formación universitaria, 13(5), 69-76. 2020.

[11] M.C. Roma, La accesibilidad en los entornos educativos virtuales: Una revisión sistemática. Revista Científica Arbitrada de la Fundación MenteClara, Vol. 6 (219). 2021.

[12] AGESIC. (2009). Guía para diseño de portales estatales. Obtenido de https://www.agesic.gub.uy/innovaportal/file/549/1/Guia_Completa_simple_faz.pdf.

[13] V. K. Pagnoni & S. I. Mariño (2020 inédito). Accesibilidad Web centradas en revisiones manuales. Estudio de un EVA de formación docente continua.

[14] F. García, De la convergencia tecnológica a la convergencia comunicativa en la educación y el progreso. ICONO 14-Revista de Comunicación y Nuevas Tecnologías – ISSN: 1697 - 8293(7), 2006.

[15] S. L. Mora, (2006). Accesibilidad en la Web: ¿Qué hace el atributo alt? Recuperado el diciembre de 2016, de http://accesibilidadenlaweb.blogspot.com.ar/2006/03/qu-hace-el-atributo-alt_17.html

[16] J. R. Hilera, L. Fernández, E. Suárez & E. T. Vilar. Evaluación de la accesibilidad de páginas web de universidades españolas y extranjeras incluidas en rankings universitarios internacionales. Revista Española de Documentación Científica, 36(1), enero-marzo 2013.

[17] V. Tabares, N. D. Duque, J. Flórez, N. Castaño & K. J. Ruiz. Evaluación de accesibilidad en sitios web educativos. Revista Vínculos, 29-40. 2014.

[18] S. L. Mora, (2021). Déficit visual. Recuperado el 2020, de http://accesibilidadweb.dlsi.ua.es/?menu=deficit-visual-introduccion

[19] W3C. (2016a). Entendiendo WCAG 2.0. Uso del color. Obtenido de https://www.w3.org/TR/UNDERSTANDING-WCAG20/visual-audio-contrast-without-color.html

[20] W3C. (2016b). Entendiendo WCAG 2.0. Sugerencia de error. Obtenido de https://www.w3.org/TR/UNDERSTANDING-WCAG20/minimize-error-suggestions.html

ISBN 978 -987-633-574-4

[21] W3C. (2016c). Entendiendo WCAG 2.0. Contenido no textual. Obtenido de https://www.w3.org/TR/UNDERSTANDING-WCAG20/text-equiv-all.html
[22] W3C. (2016d). Entendiendo WCAG 2.0. Tamaño del texto. Obtenido de https://www.w3.org/TR/UNDERSTANDING-WCAG20/visual-audio-contrast-scale.html
[23] S. I. Mariño, P. L. Alfonzo & M. V. Godoy. Accesibilidad Web. Un aporte de responsabilidad social universitaria. 2020, Obtenido de DOI 10.22533/at.ed.6832020032

ISBN 978 -987-633-574-4

Catalogación de Aplicaciones Realidad Aumentada para enseñanza-aprendizaje

Mario A. Vincenzi[1], María J. Abásolo[12]

1 Universidad Nacional de La Plata, Facultad de Informática
calle 50 y 120, 1900 La Plata, Argentina
2 Comisión de Investigaciones Científicas de la Pcia. de Bs.As.
marioavincenzi@gmail.com, mjabasolo@lidi.info.unlp.edu.ar

Resumen. La aplicación de la Realidad Aumentada (RA) en la educación abarca diferentes áreas temáticas con incidencia en todos los diferentes niveles educativos. En este artículo se presenta una catalogación de aplicaciones móviles educativas basadas en RA disponibles en la web. Se elabora para ello una ficha que reúne aspectos generales, técnicos, didácticos y específicos de RA. Se describe una herramienta utilizada para la catalogación la cual permite la consulta on line, y se presenta la búsqueda de las aplicaciones y resultado de la catalogación.

Keywords: realidad aumentada, enseñanza-aprendizaje, aplicaciones, catalogación

1 Introducción

En la educación, la Realidad Aumentada (RA) constituye un soporte tecnológico que puede enriquecer cualquier experiencia de aprendizaje.Gavilanes, Abásolo & Cuji [1] presentan un resumen de diferentes artículos de revisión sobre RA en educación para determinar cuáles son los grupos destinatarios, áreas de aplicación; metodologías, tipo de aplicaciones, tecnologías, software utilizados; ventajas, desventajas señaladas de la aplicación de esta tecnología. Entre los puntos a remarcar se evidencia la motivación de los alumnos por el uso de esta tecnología RA, aunque también se presenta como desventaja, que esa misma motivación con el tiempo pueda disminuir provocada por el factor de novedad.

En la bibliografía se encuentran diversas experiencias de la RA en los diferentes niveles educativos, desde el preescolar al universitario, con objetivos y temáticas didácticas que abarcan: aprendizaje de lectura en nivel inicial [2]; practicar idiomas [3], literatura[4], química [5] en nivel primario; introducir en la música a alumnos del secundario [6]; matemática nivel secundario, en particular a estudiantes con discapacidad [7]; la práctica e investigación en medicina en el nivel universitario [8].

Gavilanes et al. [9] plantean el uso de un modelo productor-consumidor en el ámbito universitario, donde los propios estudiantes del área informática son entrenados para realizar objetos de aprendizaje con RA para ser utilizados en la docencia universitaria de estudiantes de otras áreas. Las experiencias han despertado un alto grado de

ISBN 978 -987-633-574-4

aceptación y motivación por parte tanto de los estudiantes productores como de los consumidores.
Entre las actividades educativas con RA que se pueden desarrollar en el aula están los juegos educativos los cuales ayudan a presentar contenidos de diferentes formas activando los distintos sentidos del estudiante mientras aprende en forma lúdica. Puede citarse [10] que presentan una aplicación educativa de RA para el primer ciclo de educación primaria, la cual ofrece un simple juego de preguntas y respuestas donde las respuestas están asociadas a marcadores, representando recursos como personajes 3D, audios, imágenes, etc. El debate actual sobre gamificación se centra en el análisis de variados modelos empleados en la educación para incrementar la motivación y participación de los estudiantes [11]. Los libros aumentados, que mediante la utilización de un dispositivo con cámara capturando el libro real, enriquecen la lectura tradicional con la visualización de modelos 3D y otra información virtual. Se puede destacar MagickBook [12] como una de las primeras experiencias de libros aumentados, y más recientemente la tesis de Gazcón [13] propone la extensión de los libros aumentados a partir de libros en papel existentes, incorporando distintos contenidos aumentados que puedan obtenerse de la reconstrucción automática de objetos y sean compartidos por los lectores de manera colaborativa
El material didáctico, permite apoyar en los procesos pedagógicos en la presentación de contenidos interactivos, en distintas modalidades de formación tales como mlearning y elearning [14]. Por su parte, en las aplicaciones de geolocalización [15], los estudiantes desempeñan un papel activo relacionándose con su entorno geográfico real. Navegadores con RA y geolocalización - como Layar, Wikitude, CamOnApp, ARACAMA3D, Smart Reality - utilizan el receptor GPS y la brújula del dispositivo móvil para determinar la posición del usuario y su orientación, pudiendo así obtener la imagen del escenario real a través de la cámara del dispositivo, y es allí donde el software agrega sobre esa imagen, la información vinculada para reproducirse en la pantalla., teniendo varias capas de información que pueden ser utilizadas según haya sido la elección del usuario. Actualmente, hay en desarrollo investigaciones que permitirán a usuarios sin conocimientos específicos de programación, tener la capacidad de crear procedimientos o series de pasos a realizar en entornos físicos mediante el uso de navegadores de RA [16].

En este artículo se presenta la búsqueda y análisis de un conjunto de aplicaciones RA educativas, con el objeto de realizar una recopilación y catalogación de algunas de las aplicaciones de RA actuales disponibles para utilizarse en el ámbito educativo de diferentes niveles, y convertirse en una guía a divulgar entre la comunidad docente. En la sección 2 se define el formato de la ficha que utilizaremos para catalogar las aplicaciones encontradas. En la sección 3 se detalla la metodología de búsqueda de las aplicaciones de RA para catalogar. En la sección 4 se describe la herramienta de consulta desarrollada para la búsqueda de las aplicaciones catalogadas. En la sección 5 se analizan los resultados de las aplicaciones RA educativas encontradas. Por último la sección 6 presenta las conclusiones.

ISBN 978 -987-633-574-4

2 Catalogación

2.1. Antecedentes

Se revisaron trabajos relacionados de diferentes autores e instituciones, donde se establecen criterios a tener en cuenta para catalogar aplicaciones. Según el relevamiento en sitios especializados en RA y educación, se detectó la falta de documentación directamente relacionada al área de exploración.

Sanchez Zuaín & Duran [17] establecen una taxonomía para clasificar aplicaciones web. Al tratarse de aplicaciones web, se deben considerar características de interés, observando requisitos de contenido, interfaces con el usuario, navegación y avances de la aplicación. Se dirige a la definición de lo que se desea producir con la identificación de requisitos, previa identificación de los objetivos y del tipo de aplicación web.

Gaetán, Buccella & Cechich [18] confeccionaron un esquema de 21 categorías de 3 grupos,basados en el Framework USCS para la taxonomía de componentes de sistemas de información geográfica. Para lograr la conformación de esta tabla analizaron la información relevada en los catálogos web de componentes SIG y para definir la funcionalidad de los componentes estandarizados, donde adecuaron la taxonomía de servicios geográficos ISO/IEC 19119.

Por su parte, el estándar IEEE-LOM (Learning Object Metadata) [19] define la estructura de metadatos para facilitar la búsqueda, evaluación, adquisición y uso de objetos de aprendizaje por parte de los alumnos, docentes o sistemas automatizados, así como el intercambio de los mismos y su uso compartido, permitiendo así el desarrollo de catálogos e inventarios.está dividido en nueve categorías, con subcategorías que definen los objetos de aprendizaje con mayor detalle: General, Ciclo de vida, Meta-Metadatos, Requisitos técnico, Características pedagógicas, Derechos de uso, Relaciones, Anotaciones, Clasificación

2.2. Diseño de ficha de catalogación de aplicaciones RA

En esta sección se presenta el diseño de una ficha de catalogación a utilizar para un repositorio de aplicaciones de RA educativas. Este formato de ficha se conformó a partir de criterios recogidos y propios, teniendo en cuenta las características mencionadas en este trabajo en lo referido a la tecnología de RA utilizada en las aplicaciones educativas.

Se definieron 19 tópicos distribuidos en 4 categorías: general, técnicos, realidad aumentada, y didácticos (Figura 1).

A partir de la definición de características descriptivas de una aplicación de RA educativa se conforma la ficha (tabla 1), que permitirá registrar los datos analizados de cada una de las aplicaciones encontradas y establecer una validación cualitativa y cuantitativa de cada aplicación.

ISBN 978 -987-633-574-4

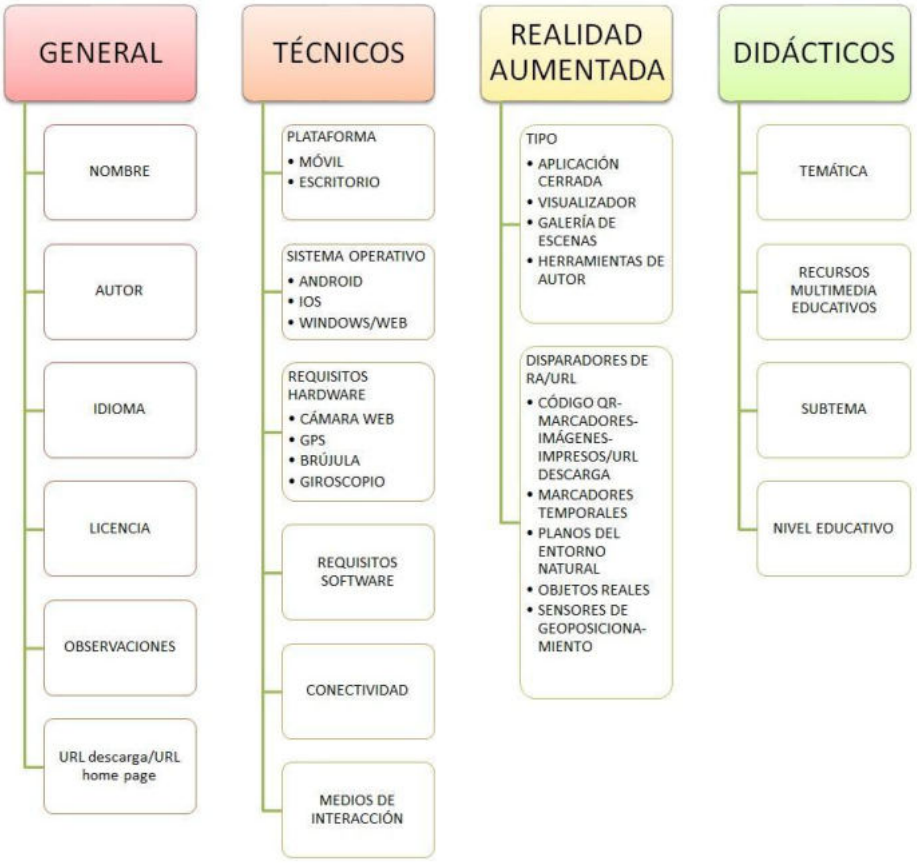


Figura 1 Esquema descriptivo de características de aplicaciones de RA educativas. Elaboración propia

3 Metodología de búsqueda de aplicaciones RA educativas

En esta sección se detalla la metodología para la búsqueda de las aplicaciones y también los criterios utilizados en el proceso.

Al momento de comenzar con la búsqueda de las aplicaciones para este estudio, se determinó que estarán acotadas -principalmente- a los dispositivos móviles, ya que estos concentran a la mayoría de los recursos utilizados por la tecnología de RA, también teniendo en cuenta que el uso de celulares entre los estudiantes y los docentes es de gran incidencia. A su vez, dentro de estos dispositivos móviles la búsqueda se concentró en la plataforma Android, ya que la gran mayoría de los teléfonos inteligentes actuales utilizan este sistema operativo como se ha manifestado en este trabajo. Por otra parte, para esta recopilación se excluyeron de la búsqueda aquellas aplicaciones de escritorio y web, herramientas de authoring, galerías de escenas y visualizadores asociados a las herramientas de *authoring*, limitándose a

ISBN 978 -987-633-574-4

Tabla 1 Ficha de catalogación aplicaciones educativas RA

Ficha descriptiva de aplicación de RA educativa						
General						
Nombre	*Nombre de la aplicación a analizar*					
URL	*Sitio de descarga – Sitio de la aplicación*					
Autor	*Nombre del desarrollador de la aplicación*					
Idioma	*Idiomas en los que se encuentra disponible la aplicación*					
Licencia	*Se describe la forma de poder adquirirlo (Paga-Gratuita)*					
Observaciones	*Características especiales de la aplicación*					
Técnicos						
		Sistema operativo				
Plataforma	O Móvil	O Android	O IOS			
	O Escritorio	O Windows	O Linux			
Requisitos de Hardware	O Cámara Web	O GPS				
	O Brújula	O Giroscopio				
Requisitos de Software	*Software preinstalado para su correcto funcionamiento*					
Conectividad	O SI	O NO				
Medios de interacción	O Mouse	O Táctil	O Gestual	O Voz	O Tangible	
Realidad Aumentada						
Tipo	Aplicación cerrada: *sin posibilidad de cambios a menos que se actualice la versión*					
	Visualizador: *aplicación que permite reproducir escenas, generalmente presentes en una galería de escenas*					
	Galerías de escenas: *Repositorio de escenas creadas de forma cerrada por una empresa o de forma abierta por usuarios. Para ser visualizada puede ser necesario una aplicación visualizadora*					
	Herramienta de Autor: *Aplicación que permite la creación de escenas de RA. Puede estar asociada con una galería de escenas donde los autores pueden publicar y compartir sus creaciones*					
Disparadores de RA/URL	O Marcadores Impresos	O Código QR	O Imágenes			
	O Marcadores Temporales					
	O Planos del entorno natural					
	O Objetos reales					
	O Sensores de geo posicionamiento					
Didácticos						
Temática	Descripción del área o áreas donde se puede utilizar la aplicación (ciencias sociales, naturales, exactas, cultura general, etc.)					
Sub tema	Área más específica de aplicación (Historia, geometría, química, etc.)					
Nivel educativo	O Especial	O Inicial	O Primaria	O Secundaria	O Superior	O Universitario
Recursos multimedia educativo	o Enlace	o Texto	o Imagen	o Imagen interactiva	o Audio	
	o Video	o Modelo 3D estático	o Modelo 3D animado	o Modelo 3D interactivo	o Actividades	

Fuente: elaboración propia

a las aplicaciones cerradas.

Los dispositivos utilizados para el proceso de búsqueda fueron un celular Motorola G6 Play y una Tablet Lenovo YOGA 8.

Se realizó una búsqueda en Google Play Store[1], utilizando como palabras claves "Realidad Aumentada" en la categoría Educación, arrojando un resultado de 250

[1] https://play.google.com/store/apps/category/GAME_EDUCATIONAL?hl=es_419

ISBN 978 -987-633-574-4

aplicaciones. Además, se utilizó “Augmented reality” y en este caso se registraron 252 resultados. No todas las aplicaciones encontradas fueron posibles de instalar porque Google Play compara los requisitos de hardware y software del dispositivo a utilizar, como así también las características propias del desarrollador; para luego compararlas con las restricciones y dependencias que se expresan en el archivo de manifiesto de la aplicación y los detalles de publicación.

Como criterios principales de inclusión para la búsqueda de las aplicaciones, se acotó a las que son gratuitas y con valoraciones de 4 o más estrellas, con la intención de filtrar aplicaciones con mal funcionamiento, las encontradas fueron 62 aplicaciones[2].

4 Herramienta de catalogación y consulta de las aplicaciones RA

En esta sección se presenta el sistema de catalogación y consulta desarrollado (Figura 2), el cual puede utilizarse como herramienta de consulta on line[3].

Las categorías definidas fueron seleccionadas a partir de la ficha desarrollada en el capítulo anterior: temática, subtema, requisitos de hardware, requisitos de software, nivel educativo, conectividad, medios de interacción y recursos multimedia.

Se desarrolló desde una planilla de cálculos la carga de la información de manera organizada según se estableció en la ficha anteriormente presentada. Se aplicó a dicha planilla de cálculos los filtros que ofrece el complemento *Awesome Table* para posibilitar realizar una búsqueda guiada, según los criterios pretendidos. Cada aplicación encontrada es visualizada en un renglón con su imagen, su nombre de la aplicación -desde donde se puede acceder al sitio de Google Play para su descarga, al estar hipervinculado- de la misma manera, además muestra el tipo de disparador de RA

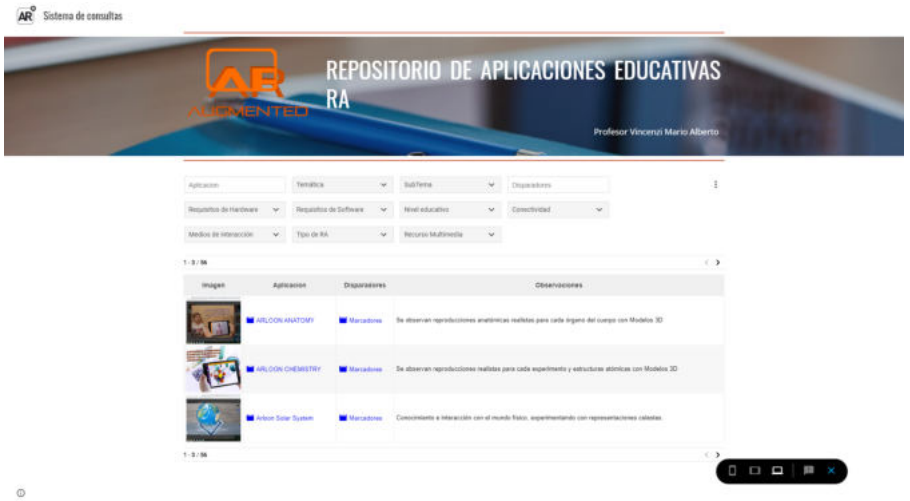


Figura 2 *Sistema de consulta de aplicaciones educativas*

[2] Última actualización: noviembre 2019

[3] https://sites.google.com/view/repositorio-educativo-ra/p%C3%A1gina-principal

ISBN 978 -987-633-574-4

que posee la aplicación con la posibilidad de acceder a ellos y las observaciones que reflejan la finalidad de dicha aplicación.

A partir de la utilización de los filtros del sistema de consultas, se puede observar cada una de las categorías de la aplicación: temática, subtema, requisitos de hardware, requisitos de software, nivel educativo, conectividad, medios de interacción y recursos multimedia. Los filtros pueden utilizarse para afinar la búsqueda, estos permiten desplegar un menú de opciones donde se pueden seleccionar las sub-categorías ya definidas de acuerdo a su característica propia. Estos filtros se pueden combinar, para así obtener un resultado más acotado y preciso sobre las aplicaciones deseadas.

5 Resultados de la catalogación de las aplicaciones RA educativas

Se catalogaron según la ficha diseñada un total de 56 aplicaciones. En la figura 3 se muestra la cantidad de aplicaciones catalogadas por temática, subtema y nivel educativo y en la tabla 2 pueden verse el resto de clasificaciones.

En relación a los aspectos didácticos se analizaron los niveles educativos, temáticas y los recursos multimedia utilizados. Según se muestra en el gráfico de la figura 3, el porcentaje más alto está representado por el 46,43% del nivel secundario, le sigue el 35,71% del nivel educativo primario, luego con el 10,71% el nivel inicial y finalmente el 7,14% corresponde al universitario y sin representación en los niveles especial (o diferencial) ni el superior (o terciario).

La temática en su mayoría, está representada por las ciencias naturales con el 66,07%, luego las temáticas ciencias sociales y exactas representan cada una el 10,71%, cultura presenta un 5,36% del total de aplicaciones, seguida por arquitectura con el 3,57%; por último, educación física y varios con sólo un 1,79%. A su vez, los subtemas brindan la mayor variedad y representación entre las aplicaciones, química con el 13,85% representa la mayoría, luego le sigue zoología y anatomía con un 10,77%. A continuación, astronomía con el 9,23%, matemáticas con el 6,15%, física junto a geografía, biología, idiomas y enciclopedia con el 4,62%; geometría, geología, historia, arquitectura, medicina representan el 3,08% cada una y por último mecánica, educación física, alfabeto, números, colores, formas y ecología con el 1,54% cada una.

Los recursos multimedia que utilizan las aplicaciones se trata en su gran mayoría de modelos 3D (animados y estáticos), seguido de texto y audio, y algunos casos que poseen actividades y con videos.

En relación a los aspectos técnicos se señaló previamente que la búsqueda se limitó a aplicaciones móviles y se analizaron requisitos de hardware y software, conectividad y medios de interacción.

Los requisitos del software hacen referencia a versiones superiores a Android 4 en la mayoría de las aplicaciones, dependiendo de las actualizaciones que ofrece el desarrollador. También podemos decir que la gran mayoría no requiere de conexión a internet para descargar contenido.

Se puede decir que el uso de la cámara es común a todas las aplicaciones, la brújula y acelerómetro solo en un caso y el giroscopio en dos de ellas.

ISBN 978 -987-633-574-4

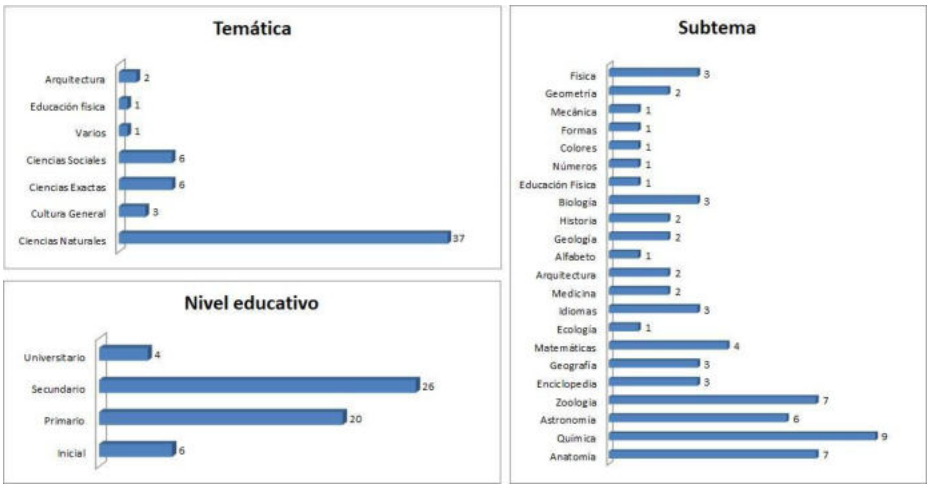


Figura 3 Cantidad de aplicaciones catalogadas por temática, subtema y nivel educativo

Tabla 2 Cantidad de aplicaciones según las categorías definidas

Requisitos Hardware	Brújula	1
	Cámara	56
	Acelerómetro	1
	Giroscopio	2
Requisitos Software	Android 1.0 y posteriores	1
	Android 2.2 y posteriores	5
	Android 2.3 y posteriores	2
	Android 4.0 y posteriores	18
	Android 4.1 y posteriores	14
	Android 4.4 y posteriores	8
	Android 5.0 y posteriores	2
	Android 5.1 y posteriores	2
	Android 6.0 y posteriores	1
	Android 7.0 y posteriores	1
	Android 8.0 y posteriores	1
	No especifica	1
Conectividad	Si	3
	No	53
Medios de interacción	Tangible	17
Disparadores	Marcadores	49
	Plano Natural	7
Tipo de RA	Aplicación cerrada	56
Recursos multimedia	Actividades	5
	Audio	8
	Modelos 3D animados	19
	Modelos 3D estáticos	39
	Modelos 3D interactivos	4
	Texto	12
	Video	2

En relación a los aspectos propiamente de RA el tipo de RA como mencionó previamente se limitó a las aplicaciones cerradas. Los disparadores en su mayoría son marcadores (tanto específicos de RA como imágenes impresas) y solo tres casos utilizan la integración de objetos en el plano natural de la escena real. Los marcadores también son utilizados en algunos casos como medio de interacción tangible.

ISBN 978 -987-633-574-4

6 Conclusiones

Con el objetivo de reunir aplicaciones de RA educativas que puedan ser útiles a la comunidad de docentes a la hora de planificar la enseñanza-aprendizaje mediada con tecnología, se realizó una búsqueda de aplicaciones móviles basadas en RA disponibles en Google Play y se procedió a realizar una catalogación de las mismas. Para esto se indagaron diferentes instrumentos de catalogación para proponer una ficha que incluye campos generales, técnicos, didácticos y específicos de RA.

Las aplicaciones catalogadas se pueden acceder mediante una herramienta de consulta on line, la cual está disponible a la comunidad. El porcentaje más alto son aplicaciones destinadas a nivel secundario, seguidas de las destinadas a nivel educativo primario. La temática en su mayoría se centra en las ciencias naturales. Se trata en general de aplicaciones que hacen uso de modelos 3D tanto animados como estáticos. En relación a los requisitos del software se encontró que la mayoría de aplicaciones no requiere de conexión a internet para su uso, lo cual representa una ventaja a la hora de poder ser usadas en la práctica en los establecimientos educativos.

En relación a los aspectos propiamente de RA el tipo de RA se utilizan marcadores impresos (tanto específicos de RA como imágenes) como disparadores y medios de interacción. Por otra parte, sólo una de las aplicaciones encontradas hace uso de la geolocalización.

Queda como trabajo futuro la actualización de la base de datos de aplicaciones, incluyendo no sólo aplicaciones cerradas sino también herramientas de *authoring* asociadas con aplicaciones visores de galerías de escenas.

Por otra parte, se espera avanzar en el trabajo de campo con docentes que hagan uso de la herramienta de consulta, para poder seleccionar y probar aplicaciones de RA educativas que encuentren de utilidad para el desarrollo de sus clases.Se prevé la generación de una comunidad virtual que en ámbitos educativos favorecerá la difusión del uso de esta tecnología entre la comunidad docente. Se pretende probar el uso de la herramienta diseñada dentro de la comunidad docente, permitiendo que se retroalimente entre los usuarios alcanzando un crecimiento colaborativo.

References

1. W. Gavilanes, M. Abasolo y B. Cuji, «Resumen de revisiones sobre Realidad Aumentada en Educación,» *Espacios,* vol. 39, nº 15, p. 14, 2018.
2. A. Vara López, «Las narrativas digitales en Educación Infantil: una experiencia de investigación e innovación con booktrailer, cuentos interactivos digitales y Realidad Aumentada» *Diablotexto Digital*, vol. 3, pp. 111-131, 2018.
3. L. Pombo y M. Marques, «Learning with the Augmented Reality EduPARK Game-Like App: Its Usability and Educational Value for Primary Education», *Intelligent Computing. CompCom. Advances in Intelligent Systems and Computing.* vol. 997, pp. 113-125, 2019.
4. E.B. Ginés Rojas, «Programa basado en la realidad aumentada para mejorar la producción de cuentos en estudiantes del 3er. grado de educación primaria de la Institución Educativa N° 88240 "Paz y amistad" Nuevo Chimbote - 2017», Tesis,

Universidad Nacional del Santa, Facultad de Educación y Humanidades, Chimbote, Perú, 2019.

5. A. Ewais y O. D. Troyer, «A Usability and Acceptance Evaluation of the Use of Augmented Reality for Learning Atoms and Molecules Reaction by Primary School Female Students in Palestine,» *Journal of Educational Computing Research*, vol. 57, nº 7, pp. 1643-1670, 2019.
6. J. Cardoso Gomes, M. Figueiredo, L. Amante y C. Gomes, «Augmented Reality in Informal Learning Environments: A Music History Exhibition». En *Interface Support for Creativity, Productivity, and Expression in Computer Graphics*, pp. 281-305, 2019.
7. R. Kellems, G. Cacciatore, y K. Osborne, «Using an Augmented Reality–Based Teaching Strategy to Teach Mathematics to Secondary Students With Disabilities», *Career Development and Transition for Exceptional Individuals*, vol. 42, nº4, 2019.
8. M. Walsh, y O. Khan, O, «P.105 The "Comprehensive 3D Skull Base Lab"-- enhancing resident education with virtual/augmented reality and 3D printing at Northwestern University» en *Canadian Journal of Neurological Sciences*, vol. 41,2019.
9. W. L. Gavilanes López, B. Cuji, J.Salazar Mera, M.J.Abásolo, «Methodology for the Production of Learning Objects Enriched with Augmented Reality by University Students,» ICL 2019, AISC 1134, pp. 492–502, 2020.
10. T. Castellano Brasero y L. Santacruz Valencia, «EnseñAPP: aplicación educativa de realidad aumentada para el primer ciclo de educación primaria», *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, vol. 21, pp. 7-14, 2018.
11. D. Claros Perdomo, E. Millán Rojas y A. Gallego Torres, «Uso de la realidad aumentada, gamificación y m-learning», *Revista Facultad de Ingeniería*, vol. 29, nº54, 2020.
12. M. Billinghurst, H. Kato e I. Poupyrev, «The Magicbook: Moving seamlessly between reality and virtuality», *IEEE Computer Graphics and Applications*, vol. 21, nº3, pp. 6-8, 2001.
13. N. F. Gazcón, Libros aumentados: Extensión de concepto, exploración e interacciones. Tesis doctoral. Universidad Nacional del Sur, Bahía Blanca, 2015.
14. A.T. Korucu, A. Alkan, «Diferencias entre m-learning (aprendizaje móvil) y e-learning, terminología básica y uso del m-learning en educación», *Procedia - Social and Behavioral Sciences,*vol. 15, pp. 1925-1930, 2011
15. J. Fombona Cadavieco y E. Vázquez Cano, «Posibilidades De Utilización De La Geolocalización Y Realidad Aumentada En El Ámbito Educativo», *Educación XX1*, vol. 20, n.º 2, 2017.
16. Becerra, M., Ierache, J., & Abasolo, M. J. «Towards Ubiquitous and Actionable Augmented Reality Browsers by using Semantic Web Technologies», In *Short Papers of the 9th Conference on Cloud Computing, Big Data & Emerging Topics* (p. 80), 2021
17. S. Sánchez Zuaín y E. Durán, «Identificación de requisitos para aplicaciones Web mediante el uso de una taxonomía basada en la catalogación de las aplicaciones», in 46 Jornadas Argentinas de Informática e Investigación Operativa (JAIIO)-43 Conferencia Latinoamericana de Informática (CLEI) XVIII Simposio Argentino de Ingeniería de Software (ASSE), Córdoba, Argentina, pp. 93-100, 2017.
18. G. Gaetán, A. Buccella y A. Cechich, «Un esquema de clasificación facetado para publicación de catálogos de componentes SIG», in Proceeding CACIC, Chilecito, La Rioja, Argentina, 2008.
19. Sicilia, M.A., García-Barriocanal, E., Pagés, C., Martínez, J.J. y Gutiérrez, J.M. «Complete metadata records in learning object repositories: some evidence and requirements», International Journal of Learning Technology, vol. 1, num. 4, pp. 411-424, 2005

ISBN 978 -987-633-574-4

Elementos de gamificación como complemento en una propuesta educativa

Angela Belcastro[1], Rodolfo Bertone[2]
[1] Universidad Nacional de la Patagonia San Juan Bosco, Departamento de Informática, Facultad de Ingeniería.
[2] Instituto de Investigaciones de Informática. Facultad de Informática. Universidad Nacional de la Plata.
angelab@ing.unp.edu.ar
pbertone@lidi.info.unlp.edu.ar

Abstract. En este trabajo se describen algunos resultados de un proyecto de investigación con elementos resultantes de revisión de bibliografía, descripción de recursos educativos creados y utilizados en la cursada 2020 en "Sistemas y Organizaciones", materia de la UNPSJB, de segundo año, primer cuatrimestre; y resultados. Estas actividades educativas, se crearon con herramientas gratuitas disponibles en la web, una de ellas apoya al docente y a los alumnos, en la creación de juegos educativos y brinda al docente, información en tiempo real del uso que cada alumno hace de estas actividades. La otra actividad descripta, fue diseñada con una herramienta que permite incorporar en un video, preguntas intermedias, abiertas y cerradas, y brinda al docente la posibilidad de realizar una devolución de resultados a cada alumno. Se destacan algunos resultados de la experiencia que promueve el aprendizaje significativo, y presenta elementos de la propuesta asociados a técnicas de gamificación.

Keywords: aprendizaje significativo, gamificación, sistemas, organizaciones.

1 Introducción

El aprendizaje puede realizarlo uno mismo; se produce dentro de cada persona. La enseñanza, por lo general, se produce con la intervención, al menos de una persona más; no es algo que ocurra dentro de la cabeza de un solo individuo. La noción de aprendizaje se asocia, en algunos casos a lo que el estudiante realmente adquiere de la enseñanza, y en otros, a los procesos que el alumno usa para adquirir el contenido (tarea). En la teoría constructivista del aprendizaje, una tarea central de la enseñanza es permitir al estudiante realizar las tareas del aprendizaje, enseñar consiste en permitir la acción de estudiar; consiste en enseñarle cómo aprender. Al enseñar, y promover el aprendizaje, en el enfoque constructivista, se intenta encontrar un equilibrio entre los resultados del aprendizaje (lo que se aprende), los procesos (como se aprende), y las condiciones prácticas (cuando, cuanto, con quienes, donde se aprende), que son elementos que intervienen en el aprendizaje. [1]

ISBN 978 -987-633-574-4

De acuerdo con McCombs & Vakili (2005), el concepto “centrado en el aprendizaje del alumno o centrado en el aprendiz” desemboca en principios y enfoques instruccionales, asociados a:

- Lo que se sabe acerca de la persona que aprende, el aprendiz: sus experiencias, perspectivas, intereses, necesidades, estilos cognitivos, los cuales deben tomarse en cuenta al diseñar el currículo y ser motivo de apoyos y adaptaciones curriculares pertinentes.
- Los procesos de aprendizaje mismos: la recuperación del mejor conocimiento disponible, basado en la teoría y la investigación educativa, acerca de cómo aprende la gente, así como de las prácticas y enfoques de enseñanza más efectivos para promover altos niveles de motivación, aprendizaje y desempeño para todos los aprendices en diversos contextos y condiciones. [2]

El aprendizaje significativo (AS), apunta a la comprensión, a organizar elementos de información, relacionándolos dentro de una estructura de significación. El modelo de AS de Anderson, el más utilizado en la enseñanza constructivista, consta de los niveles: 1- Articulación de conocimientos con saberes previos. 2- Estructuración, implica formación de nuevas estructuras conceptuales a nuevas formas de conocer. Se logra a través de esquemas, mapas, metáforas y guiones, en otros recursos. 3- Ajuste o actuación, acopla el conocimiento y la tarea (competencia). Éste se logra con la práctica y da como resultado un aprendizaje experto. [3][4][5][6][7]

Pese a que el aprendiz y el aprendizaje tienen un lugar protagónico, se reconoce la necesidad y la relevancia de la enseñanza como actividad de soporte realizada en distintos niveles (diseño, planificación, gestión, interacción, evaluación), sin la cual la actividad constructiva del alumno no ocurriría como debiese en función de la valoración sociocultural de los saberes a aprender, y de las intenciones educativas.

Al enseñar se busca propiciar que los estudiantes construyan su propio conocimiento, tomen decisiones respecto a su trayecto formativo, impulsen habilidades del pensamiento de alto nivel, aprendan a trabajar de modo colaborativo, se apropien de tecnologías de avanzada y adquieran competencias o saberes tales que les permitan afrontar el mundo complejo, incierto y cambiante que les toca vivir. Los estudiantes deben estar motivados por aprender, para poder hacerlo; los juegos simples, con entrenamientos de temas bajo estudio, pueden ser de ayuda para ello, complementando el conjunto de actividades consideradas en la propuesta educativa.

El juego es una actividad esencial y permanente del ser humano, que le ha permitido asimilar la cultura, fortalecer destrezas para conocer, comprender y actuar sobre el mundo; es una fuente educativa, de diversión y de placer, que brinda vivencias y situaciones que favorecen el desarrollo del ser humano.

El aprendizaje es más eficaz cuando se entiende como un proceso de construcción intencional de significado a partir de la información y la experiencia.

Werbach y Hunter (2012) definen gamificación, como la adhesión de elementos y técnicas propias del desarrollo de los juegos a contextos que no están ideados para ser lúdicos.

Según Ibar (2014) las definiciones sobre gamificación constan de tres partes principales: 1. Elementos de juegos: aquellos elementos comunes a todos los juegos (estrategias, avatares, puntuaciones, potenciadores, etc.) 2. Técnicas de desarrollo: el diseño de los juegos, la ingeniería detrás de los mismos. 3. Contextos: los espacios de no juego donde podemos desarrollar estrategias de gamificación.

Kapp (2012) señala que gamificar es la aplicación de mecánicas, estéticas y estrategias asociadas comúnmente a los juegos para motivar, promover y resolver problemas. [8][9][10][11]

La gamificación o ludificación sugiere poder utilizar elementos del juego, y el diseño de juegos, para mejorar el compromiso y la motivación de los participantes. La vivencia del juego puede transformar la actitud de un usuario pasivo, en uno activo, que avanza en él, voluntariamente. Hay diversos recursos educativos de apoyo al docente, algunos incluyen técnicas de gamificación y apoyan la evaluación en tiempo real con retroalimentación y otros respaldan el trabajo en equipo. Tanto socrative (https://www.socrative.com/), como kahoot (https://kahoot.com/), son sistemas SRS, los estudiantes pueden participar empleando smartphones o a través de Internet, permiten crear actividades multiple-choice interactivas basadas en la lógica de un juego, ofrecen un importante nivel de personalización, ayudan al docente a generar una partida basada en la temática de estudio. Permiten incorporar imágenes y videos, y resolver los acertijos en equipo. Ayudan a estimular a los alumnos a participar, y pueden funcionar como instrumento de evaluación. Socrative es una herramienta online que apoya la autoevaluación y la evaluación en línea, con ella los alumnos reciben retroalimentación en tiempo real. Permite al docente preparar cuestionarios y carreras de mente, es intuitiva, gráfica y tiene una versión gratuita, permite medir la participación del alumnado, con resultados individuales y globales de la clase.

Existen plataformas que permiten a los profesores editar y compartir contenido multimedia en la web, o brindan herramientas gratuitas para crear recursos educativos con evaluaciones interactivas con ludificación. La plataforma learning app (https://learningapps.org/) permite crear aulas y ejercicios interactivos, con juegos educativos que se ejecutan en la web, registra el uso de cada ejercicio en tiempo real, y lo pone a disposición del docente. EdPuzzle (https://edpuzzle.com/) permite a los docentes generar recursos educativos en formato audiovisual, insertando en videos, preguntas intermedias, y actividades, que el alumno puede realizar en la web, desde cualquier lugar; recibiendo la devolución del docente. [13][14][15][16]

2 Desarrollo

Las acciones desarrolladas, asociadas a la confección y uso de recursos educativos que complementan actividades con trabajo colaborativo en la asignatura “Sistemas y Organizaciones (SyO)” en la UNPSJB, sede Comodoro Rivadavia, en 2020, son:

- **a.** Exploración y selección de fuentes relevantes de información provenientes de investigaciones actuales, y análisis de contenidos de fuentes de datos seleccionadas. Los temas centrales considerados, han sido: herramientas de apoyo a la enseñanza, constructivismo, AS.
- **b.** Análisis de contenidos de los informes científicos seleccionados en la búsqueda de investigaciones actuales.
- **c.** Creación de propuesta educativa mediada por TICs, aplicando la Teoría Constructivista del aprendizaje (PETC), con evaluación formativa, para promover la creatividad, con actividades de distintas características, trabajo colaborativo, cooperativo e individual; para la materia SyO, del primer

ISBN 978 -987-633-574-4

cuatrimestre de segundo año de las carreras: “Analista Programador Universitario (APU)” y “Licenciatura en Informática (LI)”, de la facultad de Ingeniería, UNPSJB, sede Comodoro Rivadavia. Las actividades surgieron a partir de un proyecto de investigación, en algunas de ellas, se crearon recursos educativos con herramientas gratuitas para la construcción de juegos educativos interactivos. Se elaboraron otros materiales educativos disponibles en el aula virtual, como presentaciones, videos, prácticos, y el “trabajo integrador en equipos (TI)”, que incluye un índice con: destinatarios, objetivos y resultados de aprendizaje, herramientas, material de base y algunas competencias a reforzar, metodología, se pide, y, evidencias y criterios de evaluación.

d. Preparación de dispositivos de medición de resultados.

Se describen en este trabajo, resultados de revisión bibliográfica, algunas actividades diseñadas y utilizadas en la cursada 2020, y sus resultados. Actividades que incluyen elementos de gamificación, y ayudaron a complementar otros desarrollos de los alumnos, con trabajo cooperativo y colaborativo, diseñados para propiciar el despliegue del alumno, de distintos niveles de procesamiento: recuerdo, comprensión, transferencia y experticia en temas centrales de la materia.

En 2020, la materia SyO, se dictó bajo la modalidad virtual, respetando las políticas definidas frente a la pandemia. Fue necesario reemplazar una de las actividades colaborativas en la que se desarrollaba una tarea auténtica que generó transferencia en base a investigaciones sobre metacognición y evaluación por competencias, con publicaciones en congresos. En ella, los alumnos debían contactarse con gerentes de empresas reales y desarrollar entrevista presencial. Se contemplaron riesgos de contagio, o de realizarse las entrevistas en forma virtual, de acercar a los alumnos a personas que pueden haber sufrido grandes pérdidas, a una situación poco motivadora para lograr AS de las organizaciones como sistemas. Se consideraron otras estrategias, para 2020, con trabajo colaborativo e investigación, creación de trabajo escrito con formato científico, mapas conceptuales, elaboración de poster, e integración de contenidos. [17][18][19]

3 Descripción de recursos educativos y actividades

Se describen recursos educativos creados con edPuzzle y con learning app, en 2020.

3.1 Actividad creada con Edpuzzle en 2020

Este recurso fue preparado con la herramienta Web EdPuzzle que apoya el desarrollo de evaluación formativa, permite al docente editar un video incluyendo interrogantes cerrados y abiertos, observar las respuestas, y enviar devoluciones a cada alumno. Dicha actividad, se denominó “Actividad anticipativa”, fue preparada como optativa, considerada para promoción, y para los restantes estudiantes. En ella se seleccionó el video titulado: “El impacto de la tecnología en los negocios. Ricoh”, https://www.youtube.com/watch?v=84Bz8nUjbc8. Dispone de preguntas de tipo multiple choice, ubicadas en momentos intermedios diferentes del video seleccionado, y de una pregunta abierta, con la consigna: “Analice otros tipos de cambios que se

han desarrollado en los últimos ocho años o que están dándose, que afectaron o afectan el trabajo de los gerentes, que son atípicos, y que no necesariamente han sido ocasionados por aspectos tecnológicos. (A) Incluya enlace completo y fecha de acceso, exclusivamente de: A-1 "revista empresarial o tecnológica prestigiosa (forbe, mercado, infotechnology, apertura)", o A-2 (sitio de una empresa, que destaca una experiencia propia no asociada a productos que proporciona, sino a experiencias que vivió), o A-3 (diario prestigioso). Y (B) Explique brevemente con sus palabras cual fue la situación que se les presentó, y qué efectos produjo, qué debieron hacer los gerentes, que antes no hacían".

Esta actividad invita al alumno a interpretar lo observado y a desarrollar actividades de investigación en internet, documentando las referencias seleccionadas y analizando su confiabilidad, vinculándose con el medio, y anticipándose al análisis del tema: "Cambios que están afectando a los gerentes". Fue confeccionada en 2020, en el marco del proyecto Nº 10/E 143, está disponible en: https://edpuzzle.com/join/wukifse. Las participaciones meritorias de los alumnos se incorporaron en un video, accesible para alumnos de las cursadas 2020 y 2021, citando apellido y nombre y aporte.

3.2 Juegos educativos interactivos creados con learningapp

La herramienta permite observar estadísticas de uso y brinda la posibilidad de obtener el recurso en formato SCORM, o ejecutarlo desde la web. Puede emplearse para solicitar a los alumnos la creación de juegos simples de un tema específico.

Fig. 1. Juego de la memoria, seguido del juego de tipo "asignar a una tabla".

Los juegos creados fueron depurados y utilizados también en la cursada 2021. Ellos ayudan a promover distintos tipos de actividades cognitivas, contemplando otras actividades, ya que se emplean como complemento, junto a otras prácticas realizadas en la cursada. La metodología propuesta consistió en realizar lectura comprensiva del material didáctico brindado, (en algunos temas, materiales en pdf y video, y en otros solamente pdf), consultar todas las dudas a la cátedra, y luego, realizar los juegos educativos asociados al tema. Algunos juegos eran optativos para quienes cursaban para obtener concepto. Entre los tipos de juegos creados, encontramos:

- El juego de la memoria, como el de la figura 1, que promueve un nivel de procesamiento que apoya el "recuerdo", asociando nociones claves y su definición. Se utilizó para nociones iniciales de "las organizaciones", "los sistemas, características y propiedades", "dato, información y conocimiento", y "el proceso de toma de decisiones".
- De tipo "asignar en una tabla", que se observa a la derecha en la figura 1, preparado para los temas: "Organigrama" y "Clases de decisiones". Estas

ISBN 978 -987-633-574-4

actividades interactivas se realizaron antes del desarrollo de la práctica del tema. El objetivo de la actividad de organigrama fue identificar nociones claves de la etapa de organización del proceso de administración, los distintos niveles de autoridad, y aspectos de su representación. Se promueve la "transferencia", ya que después de desarrollar estas ejercitaciones, aplicaron los conocimientos adquiridos al resolver casos de aplicación en equipo en foros del aula virtual (AV). Quienes no aprobaron la actividad en equipo en primera instancia, desarrollaron actividad individual. En actividades en foros, se brindó inicialmente la metodología de trabajo, los indicadores de desempeño que se utilizarían al evaluar, contemplando la participación de cada alumno, y considerando el problema a resolver.

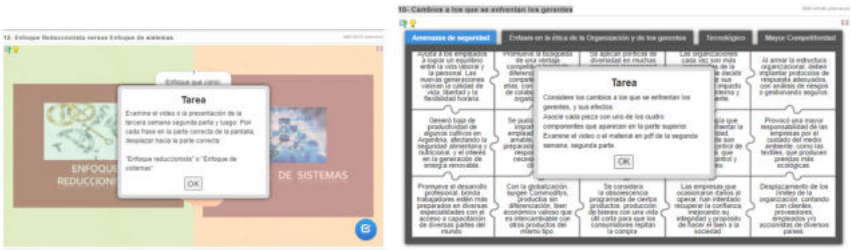


Fig. 2. "Ejercicio de clasificación", y luego "Rompecabezas con preguntas".

- Ejercicio de clasificación para entrenamiento del tema: "enfoque reduccionista versus enfoque de sistemas". El mismo queda representado en la figura 2.
- Rompecabezas con preguntas, como vemos en la figura 2. Empleados para entrenamiento de los temas: "medidas de desempeño de las organizaciones", y "cambios a los que se enfrentan los gerentes". El desarrollo de esta actividad se propuso al alumno, después del desarrollo de la actividad anticipativa creada con edPuzzle, y de realizar lectura comprensiva de material didáctico junto al análisis del video que incluye ejemplos con investigaciones realizadas por los alumnos.
- Carrera de taconeo, de la figura 3, empleado para entrenamiento en el tema: "Globalización. Cuarta Revolución Industrial". Este ejercicios interactivos tienen una vinculación con una entrada creada en 2020, del blog de SyO https://syoaprendizaje.blogspot.com/2020/03/globalizacion-y-cuarta-revolucion.html, diseñado en proyecto de investigación previo, con material de diferentes tecnologías de la industria 4.0.
- Crucigrama, al realizar entrenamiento en el tema: "como se vuelven globales las empresas".
- Emparejar elementos, para el tema: "importancia de los valores en las organizaciones".

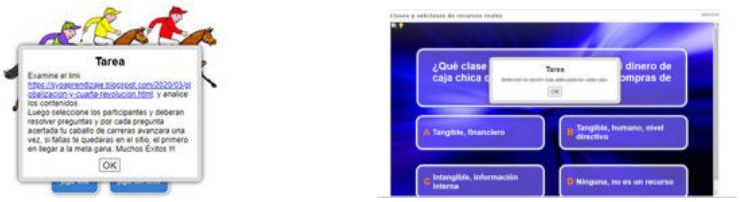


Fig. 3. Carrera de Taconeo inicialmente, seguido del Juego del millonario.

ISBN 978 -987-633-574-4

- ¿Quieres ser millonario?, con múltiple choice, como se presenta en la figura 3, de "Clases y subclases de recursos reales", actividad requerida para obtener concepto, como paso previo al desarrollo de práctica del tema.
- Ejercicios de respuesta abierta, integradores, que abarcan más de un tema, observado previamente. Para entrenamiento en los temas: "las organizaciones, sus valores y su administración", y "sistemas de información, enfoque de actividades y calidad de la información" y "las organizaciones". En este tipo de ejercicios el nivel de procesamiento es mayor, el alumno debe cargar la respuesta para cada concepto, y en algunos casos, responder verdadero o falso.

4 Resultados

Del total de inscriptos (43), un 44,53% aprobó por promoción, obtuvo concepto o aprobó en segunda cursada, un 16,28% promocionó, un 5% desarrolló la segunda cursada, un 44% estuvo ausente, y un 11,62% abandonó en el primer parcial. Debían aprobar actividades, y podían iniciarlas si tenían aprobada la actividad previa, habían aplicado temas iniciales que se observaron gradualmente. Entre las condiciones para lograr AS, encontramos la motivación por aprender, la secuencia de contenidos y los conocimientos previos de cada alumno. Desde fines de abril, trabajamos con tres tipos de actividades: (de alumnos que iban al día), (recuperaban por primera vez) y, (recuperaban por segunda vez). Todos los docentes de cátedra participaron en prácticas. Se realizaron conferencias, consultas y otras actividades. El 68,4% de los alumnos que obtuvieron concepto o promoción, trabajaron con entusiasmo, aprobando en primera instancia actividades de las primeras unidades. El promedio de notas de promoción fue 8,86, y el de notas de concepto, 8.

En el diagnóstico se consultaron aspectos del perfil del estudiante, se propició el recuerdo de nociones claves ya observadas, y se solicitaron observaciones de las actividades que habían realizado hasta el momento, entre ellas encontramos: "Las actividades en general, creo que son entretenidas y motivan a querer avanzar con ellas", "Por el momento se me ha hecho interesante las actividades virtuales didácticas", "hasta ahora me van gustando mucho las actividades del módulo interactivo, ya que son muy didácticos y a la vez te ayudan a pensar". El 51,16% de los inscriptos desarrolló el diagnóstico en la cuarta semana del cuatrimestre.

La herramienta learning app, provee estadísticas de uso en tiempo real, y resultados de actividades, en los cuales incluye apellido y nombre del alumno seguido del nombre de la actividad, luego indica si abrió, desarrolló o solucionó el juego, y presenta la fecha y hora. Un 30,23% de los alumnos inscriptos desarrolló la totalidad de los ejercicios de learning, un 9,3% realizó 15 ejercicios; un 9,3%, 10 de los ejercicios; y un 4,65% realizaron 8, 7 y 6 juegos interactivos. Las estadísticas indican cuanto tiempo tardó el alumno en resolver el ejercicio.

Se confeccionó dispositivo en google doc con cuestionario de fin de cursada, para propiciar el ejercicio de metacognición y recabar información de retroalimentación para mejora y medición de resultados con autoevaluación, con las dimensiones: "A-Metacognición", "B- Estilos de aprendizaje", "C- Trabajo colaborativo", y "D-Retroalimentación". Centraremos la atención, en este trabajo, en aspectos asociados a

actividades de learning y edPuzzle. El 37,2% de los inscriptos completó este cuestionario. En la primera dimensión, se solicita al alumno que analice la imagen del proceso de metacognición (con la escalera de metacognición), reflexione, y luego responda lo solicitado. Se citan a continuación, preguntas y sólo algunas respuestas, las asociadas a las actividades de learning y anticipativa:

• ¿Podés transmitirnos una reflexión sobre lo aprendido?: "Aprendí a desarrollar de mejor manera los escritos (buscar fuentes confiables, destacar puntos claves del tema que se solicita, relacionar temas)", y "Podría volver a aplicarlo. Logré comprender los conocimientos. He utilizado métodos diversos que la materia ofrecía. He aprendido a investigar mejor y a reconocer fuentes confiables".

• Reflexión sobre las habilidades que puso en juego al cursar la materia. Un 93,8% de los alumnos eligieron "Redacción de trabajo escrito", un 81,3%, "Armado de mapas conceptuales" y un 75% "entrenamiento con análisis de material y desarrollo de ejercicio interactivo". Las habilidades de "realización de investigación con redacción de fundamentación", y "Lectura comprensiva", fueron elegidas por el 68,8% de los alumnos, junto a otras habilidades.

• Métodos de estudio, un 68,8% seleccionó "desarrollo de ejercicios interactivos", un 37,5% "lectura comprensiva regularmente, a medida que se incorporaban los materiales", y un 25%, "participación en la actividad de EdPuzzle".

En la dimensión B, el estilo de aprendizaje (EA) predominante fue el visual en un 68,8%, el auditivo en un 37,5%, y ningún alumno con EA kinestésico.

En la dimensión D, de retroalimentación:

• Se solicita el nivel de satisfacción en actividades específicas, con una escala ("Muy alto, alto, aceptable, pobre, no participé en esa actividad). Entre ellas, en el "desarrollo de ejercicios interactivos de entrenamiento de learning", un 43,75% de los alumnos que llenaron el cuestionario, indicó que la satisfacción alcanzada fue "Muy alta", un 31,23% que fue "Alta" y un 25% que fue "Aceptable". En "Tu aporte aparece en el video con actividad de investigación de alumnos, empleando EdPuzzle", un 25% de los alumnos que llenaron el cuestionario, indicó que la satisfacción alcanzada fue "Muy alta"; un 31,23% que fue "Alta"; un 6,25% que fue "Aceptable" y "Pobre"; y un 31,23%, no participó en ella. De todas las actividades consideradas, la actividad de learning fue la que obtuvo un mayor porcentaje en la calificación "Muy alto". Se solicitan luego, la razones por las que consideraron "Pobre" alguna de las actividades, y uno de los alumnos, respondió: "en su momento no pude realizar la actividad de edpuzzle y de "el ahorcado" por problemas de conexión". Se extendió el tiempo de desarrollo de la actividad, porque algunos alumnos lo solicitaron por mail, fue lenta la conexión para realizar la devolución del docente.

• En la pregunta abierta: ¿Cuál fue la actividad que consideraste más motivadora e interesante? ¿Podrías indicarnos qué aspectos valoras de ella? Dos de los comentarios se asocian al desarrollo de juegos: "La actividad learning ya que al hacer los ejercicios ponía en práctica lo aprendido y lo entendido de los materiales sobre los temas dados", y "Las actividades de los learning app eran entretenidas, y podías ir aprendiendo al hacerlas, también te servía de un buen resumen". Los demás comentarios están asociados a otras actividades.

• ¿Qué aspectos positivos valoras de la cursada?, todos los encuestados incluyeron aspectos positivos, entre ellos: "La cantidad de material disponible, el learning app, la posibilidad de promocionar, las conferencias explicativas", "Se adaptaron muy bien a

la enseñanza virtual producto de la pandemia, lo cual se reflejó en una cursada satisfactoria y útil", "Me gustó mucho la exigencia de trabajos constante, como para no perder el hilo y mantenerse activo. Super importante", y "Se puede valorar que se aprende mucho a lo largo de la cursada, es una materia que trata de mucha lectura y de tratar de agarrar los conceptos cuanto antes para hacer los distintos trabajos, hay un buen ambiente de trabajo y hay trabajos interesantes".

Los aportes de investigación realizados por los alumnos en la actividad anticipativa se brindaron como ejemplos de cambios a los que se enfrentan los gerentes, tanto al considerar el entorno externo, como al contemplar la cultura organizacional, en un video, indicando el nombre y apellido de cada alumno. Un trabajo no fue avalado, ya que no indicaba el enlace utilizado. Son catorce los aportes meritorios difundidos.

En el TI, los equipos elaboraron un trabajo escrito en formato científico y un poster, confeccionaron mapas conceptuales y realizaron debates sobre los procesos organizacionales típicos, las actividades que se desarrollan en las empresas, la información que se necesita, los distintos tipos de sistemas de información y tecnologías específicas de la industria 4.0. Integraron diversos temas de la asignatura. Cada alumno desarrolló una autoevaluación y evaluación de pares de posters de cada equipo. Uno de los equipos elaboró una actividad optativa propuesta utilizando colaborativamente, la herramienta coggle con mapa mental con algunas producciones de los alumnos, accesible para usuarios que se identifican en coggle, en: https://coggle.it/diagram/XzsCeMZTxnqfq7A7/t/-

5 Conclusiones

Con el uso de juegos interactivos, se intenta lograr motivación por aprender, fortaleciendo competencias del profesional de Informática y mejorando el compromiso de los alumnos. Estas actividades ayudaron a los alumnos a prepararse para el desarrollo de casos de aplicación y del TI. Los comentarios de alumnos en el diagnóstico y en el cuestionario de fin de cursada, muestran que han motivado a algunos alumnos a entrenarse en temas de la materia. La actividad anticipativa fortaleció habilidades de investigación de los estudiantes que participaron. Se analizarán aspectos de trabajo colaborativo, gamificación, y el uso y creación de recursos y herramientas educativas, que ayudan a los docentes a propiciar AS de temas de asignaturas.

References

1. Gary D. Fenstermacher. TRES ASPECTOS DE LA FILOSOFÍA DE LA INVESTIGACIÓN SOBRE LA ENSEÑANZA. En Merlin C. Wittrock. La investigación de la enseñanza I, Editorial Paidós, 1989. Universidad de Arizona Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195--197 (1981)
2. Hernández Rojas, Gerardo; Díaz Barriga, Frida. Una mirada psicoeducativa al aprendizaje: qué sabemos y hacia dónde vamos. Revista Electrónica Sinéctica, núm. 40, enero-junio, 2013, pp. 1-19 Instituto Tecnológico y de Estudios Superiores de Occidente Jalisco, México.

3. Díaz Barriga Arceo, Frida. Cognición situada y estrategias para el aprendizaje significativo. REDIE. Revista Electrónica de Investigación Educativa, vol. 5, núm. 2, 2003, pp. 105-117. Universidad Autónoma de Baja California. Ensenada, México.
4. Richard E. Mayer. Psicología de la Educación. Enseñar para un Aprendizaje Significativo. Volumen II. Pearson. Prentice Hall. 2004.
5. Mavilo Calero Pérez. Constructivismo pedagógico. Teorías y aplicaciones básicas. Alfaomega. 2008.
6. Alexander Ortiz Ocaña. MODELOS PEDAGÓGICOS Y TEORÍAS DEL APRENDIZAJE. ¿Cómo elaborar el modelo pedagógico de la institución educativa? Ediciones de la U. 2013.
7. lvarado Resendiz, J. L., García Munguía, M., & Castellanos López, L. Y. (2017). Aprendizaje Significativo En La Docencia De La Educación Superior. XIKUA Boletín Científico De La Escuela Superior De Tlahuelilpan, 5(9). https://doi.org/10.29057/xikua.v5i9.2239
8. Rodriguez, Facundo y otros. (2016). Flip-Flop. Aplicación de Buenas Prácticas a partir de la Gamificación. CACIC 2016. Disponible en: http://sedici.unlp.edu.ar/handle/10915/56649 Acceso: 13-12-19
9. Jéfferson Beltrán Morales. E-learning y gamificación como apoyo al aprendizaje de programación. 2017.
10. Coello Morán, L. J., & Gavilanes Aray, B. E. (2019). Tesis. Recuperado a partir de http://repositorio.ug.edu.ec/handle/redug/40728
11. Ruth S. Contreras Espinosa y Jose Luis Eguia (2016): Gamificación en aulas Universitarias. Bellaterra: Institut de la Comunicació, Universitat Autònoma de Barcelona.
12. Viktória Bieliková (2019). Aplicación de algunos elementos de la gamificación y del aprendizaje invertido en el aula de ELE. MASARYKOVA UNIVERZITA.
13. Víctor Hugo Perera Rodríguez. Carlos Hervás Gómez. Percepción de estudiantes universitarios sobre el uso de Socrative en experiencias de aprendizaje con tecnología móvil. Revista electrónica de investigación educativa versión On-line ISSN 1607-4041. https://doi.org/10.24320/redie.2019.21.e05.1850 Acceso: 24-8-2020.
14. Isabel Narbón-Perpiñá. Jesús Peiró-Palomino. La plataforma Socrative como herramienta de aprendizaje: Una aplicación a la asignatura Métodos Cuantitativos. Revista electrónica sobre la enseñanza de la Economía Pública. http://e-publica.unizar.es/wp-content/uploads/2018/02/2%C2%BA-17207-maquetado.pdf Acceso: 24-08-2020.
15. Juan Pablo Hernández-Ramos, Mª Victoria Martín-Cilleros, Mª Cruz Sánchez-Gómez. Valoración del empleo de Kahoot en la docencia universitaria en base a las consideraciones de los estudiantes. RISTI - Revista Ibérica de Sistemas e Tecnologías de Información. ISSN 1646-9895. Disponible en: http://dx.doi.org/10.17013/risti.37.16-30 Acceso: 25-8-2020.
16. Basilio Pueo, José Manuel Jiménez-Olmedo, Alfonso Penichet-Tomas y José Antonio Carbonell-Martinez. Aplicación de la herramienta EDpuzzle en entornos de aprendizaje individuales dentro del aula. Universidad de Alicante http://rua.ua.es/dspace/handle/10045/71190 Acceso: 28-08-2020
17. Belcastro, A. Bertone, R. "Experiencia de Acercamiento al Enfoque de Formación por Competencias que Intenta Propiciar Aprendizaje Significativo". Actas del 7mo Congreso Nacional de Ingeniería Informática- Sistemas de Información. 2019, Páginas: 389 a 398. ISSN 2347-0372 https://conaiisi2019.unlam.edu.ar/pdf/2019-CONAIISI-ACTAS-7MA-EDICION.pdf
18. Belcastro, A; Bertone, R. Tarea Auténtica Mediada por Tecnología. Experiencia desde una Asignatura Universitaria. Actas del 5to congreso Nacional de Ingeniería Informática /Sistemas de Información (CONAIISI 2017), pp. 1188-1200; UTN -Argentina (2017).
19. Belcastro, A; Bertone, R. "Apoyando el Ejercicio de Metacognición en el Ámbito Universitario", noviembre de 2018, "Actas del 6to Congreso Nacional de Ingeniería Informática /Sistemas de Información". ISSN 2347-0372. https://www.conaiisi2018mdp.org/memorias/memorias.html Acceso: 28-08-2020

ISBN 978 -987-633-574-4

Thematic Evolution of Scientific Publications in Spanish

Santiago Bianco[1], Laura Lanzarini[2], and Alejandra Zangara[2]

[1] *Information Systems Research Group, UNLa (GISI-UNLa)*

[2] *Computer Science Research Institute LIDI (III-LIDI), UNLP-CICPBA*

sabianco@unla.edu.ar, {laural,azangara}@lidiinfo.unlp.edu.ar

Abstract. Thematic evolution is a relevant technique when processing text documents about a specific topic but from different periods of time. Identifying changes in terminology and the evolution of research fields is of great interest to disciplines such as bibliometrics and scientometrics. In this article, strategies are proposed to improve the analysis of thematic evolution in scientific publications in Spanish, together with visualization techniques that allow highlighting the most relevant results. In principle, this methodology can be used in different contexts; however, here we apply it to the analysis of the scientific production related to technology in education and technology education, as an expansion of the work carried out in [1]. The tests implemented allow us stating that the inclusion index is an adequate metric to select the most relevant topic relationships, facilitating the understanding and visualization of the results obtained.

Keywords: Bibliometric Analysis, Text Mining, Thematic Evolution

1 Introduction

The analysis of text documents from specific contexts is a topic of interest for different areas such as information retrieval, document classification, bibliometric and scientometric analysis, and so forth.

When processing text documents written in different time periods, thematic evolution is an aspect that must be taken into account. Identifying the changes that took place over time in the nomenclature used on various topics within the same discipline or discourse area is an extremely useful tool when trying to apply text mining strategies.

As a specific case, any teacher, researcher or student who needs to write an article, thesis or research work, will have to carry out a review of the corresponding state of the art. In this sense, the possible topics of interest within a particular domain have to be identified. This bibliographic search process is generally time-consuming and, if not oriented correctly, can lead to blockages and frustration for the researcher. It would be interesting then to have methods and tools available to simplify the search and analysis of bibliography or publications of any kind for these processes.

ISBN 978 -987-633-574-4

In the first instance, bibliometric tools could be used to carry out the initial analysis of the texts of interest. Bibliometrics is known as a discipline capable of describing a set of publications applying statistical analysis techniques, identifying relevant thematic focuses, author collaboration networks, information on citations, and so forth. Scientometrics is a subdiscipline of bibliometrics that focuses specifically on scientific publications.

In any case, these approaches generally allow quantitative analyzes such as a list of the most cited authors, institutions with the highest number of publications, topics most written about, and so forth. When a more in-depth qualitative analysis is required, text mining and visualization techniques should be applied, such as thematic maps together with traditional bibliometric methods.

Thematic maps are a way of representing different topics covered in a field of a scientific discipline at a certain moment. Different types of bibliometric information can be used to build these graphs, one of them being the analysis and correlation between relevant terms.

An analysis technique called thematic evolution is derived from the thematic maps. It consists of showing the "evolution" of the relevance of a particular topic on a timeline. For example, it could be shown that in 2010 there was a thematic focus dedicated to research in neural networks and that the same group of people who worked on this topic gradually moved their research interests towards a different topic, such as black box model interpretability. The idea is to show that the first theme mutated or evolved into the second one. It should be noted that, in this context, the term "evolution" means "change and transformation", and does not imply there has been an improvement. Thus, saying that "Topic A evolved into Topic B" does not necessarily mean that Topic B is better in some ways than Topic A. In the following section, the methodology proposed for the analysis of thematic evolution in scientific publications is detailed, including all required metrics and techniques.

2 Proposed Methodology

2.1 Gathering the Documents

To carry out a thematic analysis, documents that are representative of the area of interest over a period of time long enough to be able to divide it into sub-periods (small periods of time into which a larger interval is divided) are required. Then, the thematic evolution will try to establish relationships between the central themes from different sub-periods.

When accessing the documents, we decided to work with scientific journals that would use the Open Journal System (OJS) for their digital issues, which allows consulting all available articles in a systematic and repeatable way, as they are supported on the same standard system. OJS [5] is an open source solution for managing and publishing academic journals online, thus reducing publication costs compared to print versions and other forms of dissemination. Unlike Scopus, Web of Science and others, no access codes or any other type of

ISBN 978 -987-633-574-4

authentication are required to extract information from the platform. This allows automating the journal articles extraction process through a script, which can easily be modified to extract the data from any journal that is implemented on OJS.

Raw data is downloaded as plain text. Key elements such as title, year of publication, abstract, and author's address are automatically pulled from the OJS system, without the need to directly access the full article. Authors and country affiliations are identified from their addresses and available metadata.

On the other hand, document publication language must be selected in advance. In this article, emphasis is placed on scientific documents in Spanish because this is still a little studied language from the point of view of thematic evolution. Inconsistent expressions, special characters, and ambiguities are processed after their download and collection, in a separate script. This script is used to give the final format to the publications so that they can be used as an input for the algorithms applied in the analysis.

2.2 Terms extraction

To analyze the documents collected, the title, abstract and keywords indicated by the authors are used. All these sections must be preprocessed and grouped so that they can be used properly by the algorithms. The process consists of the following steps:

1. The terms contained in the previously mentioned sections (title, abstract and keywords indicated by the authors) are extracted and standardized. This process consists of replacing special characters, unifying synonyms and acronyms, and writing all terms in lowercase.
2. The n-grams obtained, made up of two, three or four words, are added to the set of terms.
3. Those that exceed a certain threshold value of TD-IDF (Term Frequency — Inverse Document Frequency) are selected from the set of terms. This metric assigns high values to those terms that have a high frequency (in the given document) with a low frequency in the entire collection of documents, thus filtering common terms [7].
4. All selected terms are unified in a corpus to be analyzed by the algorithms.

2.3 Research Topics Identification

To detect the research topics and/or fields of interest for researchers, the joint occurrence or co-occurrence of previously identified terms is used [3]. This co-occurrence is calculated as indicated in equation 1 where c_{ij} is the number of documents in which both terms appear together, and c_i and c_j are the number of documents in which they appear individually.

$$e_{ij} = \frac{c_{ij}}{c_i c_j} \tag{1}$$

ISBN 978 -987-633-574-4

Using these co-occurrence values, the simple center algorithm [4]] is applied to build thematic networks made up of subgroups of strongly linked terms that correspond to interests or research issues of great importance in the academic field.

The detected networks can be represented using the density and centrality measures defined in [2].

By analyzing the relationship between the terms that make up the different thematic networks within a discipline in different sub-periods of time, the development of a given topic over the years can be analyzed, and the changes in relevant thematic focuses can be seen. This is known as thematic evolution, and is discussed in more detail in the next section.

2.4 Thematic Evolution

A thematic area is a set of topics that have evolved over different sub-periods. Each topic is made up of a set of terms. Let T_t be the set of topics detected in sub-period t and let $U \epsilon T_t$ be a topic detected in sub-period t. Let $V \epsilon T_{t+1}$ be a topic detected in the following sub-period $t+1$. A thematic evolution from topic U to topic V is considered to have happened if there are common terms in both sets. Each $k \epsilon U \cap V$ term is considered a *thematic link*. To weight the importance of a thematic link, the inclusion index defined in [8] calculated according to equation (2) is used. This index is a simple metric that in this context is used to measure how strong the relationship between two topics is. Its value is between 0 and 1; a higher value corresponds to a stronger relationship.

$$inclusion = \frac{\#(U \cap V)}{min(\#U, \#V)} \tag{2}$$

If a topic from a sub-period has no thematic link with another topic from a later sub-period, it is considered to be discontinuous, whereas if there is a topic unrelated to a previous sub-period, it is considered as a new or emerging topic.

Understanding that the inclusion index is important when it comes to identifying the degree of relationship between the topics, in this work it is used as a metric to simplify thematic evolution visualization.

3 Results

To measure the performance of the methodology proposed in this article, documents from the *EDUTEC* journal were used. This journal was selected because it addresses two specific topics; namely, technology applied to education and technology education. This latter aspect is relevant because all articles published use specific common vocabulary. In addition, it has publications in Spanish and has numbers published for more than 25 years.

During the article collection phase, only publications in Spanish with abstracts available for extraction and keywords or abstracts uploaded were taken into account. As a result, 392 documents were identified.

ISBN 978 -987-633-574-4

Subsequently, the set of documents obtained was divided into three sub-periods of similar duration, as follows:

- Sub-period 1: 1995-2005
- Sub-period 2: 2006-2013
- Sub-period 2: 2014-2020

For each sub-period, the most relevant topics were identified using two different strategies – first, using only the keywords and second, adding abstracts and titles to these keywords. The values of the parameters used in both cases are indicated in Table 1. For each case, the relationships identified were considered and the most relevant ones were selected according to their inclusion level value.

Parameter description	**Value**
Number of words to use in each topic (set of terms)	250
Minimum frequency for a term to be considered a member of a topic	20

Table 1. Parameters used in the analysis

As a result of the first strategy, that is, the evolution of topics from sets of terms selected taking into account only keywords, the graph in Figure 1 was obtained. In this figure, the thickness of the bands that join the topics in the different sub-periods is proportional to their inclusion index. In other words, the wider the band that joins two topics, the greater the value of the inclusion index between the two. As it can be seen, despite the fact that the number of terms involved is scarce, it is somewhat complex to identify the most relevant ones. This is where the use of the inclusion index can be very useful.

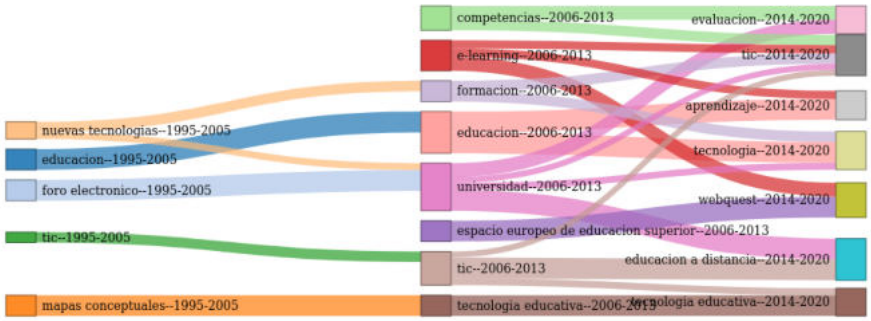


Fig. 1. Thematic evolution results using just keywords

ISBN 978 -987-633-574-4

For example, a simpler and more readable visualization applying a filter with a threshold of 0.5 for the inclusion index can be observed in Figure 2. By eliminating the topics with lower inclusion, charts become easier to read and the possibility of considering unreliable results in the analysis is reduced. Thus, it is easier to visualize those terms whose relationship between sub-periods is supported by a higher level of inclusion.

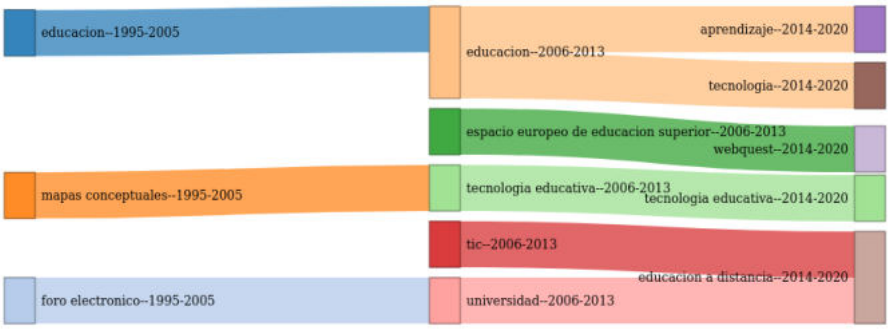


Fig. 2. Thematic evolution results using just keywords, filtered by inclusion index

Applying the second strategy, as expected, the topics identified were more closely related to the documents. This is reflected in how topics are linked, shown in Figure 3 This figure shows the relationships in a clearer way in the absence of an excessive number of crosses between connections, as it was the case in Figure 1. Regardless of this, the value of the inclusion index for each pair of terms is still directly proportional to the importance of their relationship.

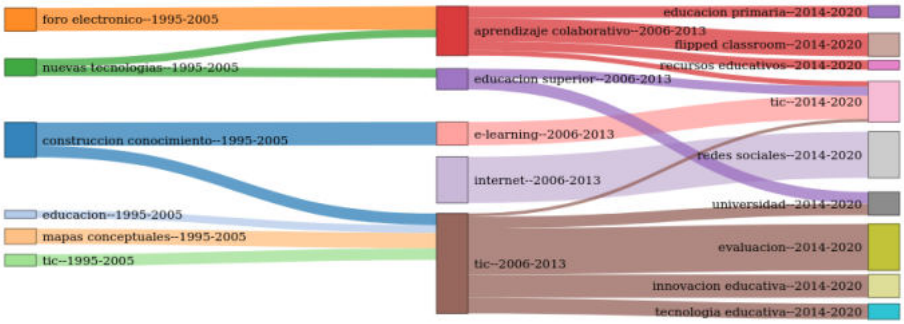


Fig. 3. Thematic evolution results using keyword terms, abstracts and titles

ISBN 978 -987-633-574-4

Table 2 shows the highest inclusion level values obtained as a result of the first procedure. These values correspond to the most interrelated topics, which were graphically joined with the thickest bands in Figure 1.

Topic A (Sub-period)	**Topic B (Sub-period)**	**Inclusion**
Education (1995-2005)	Education (2006-2013)	0.5
Electronic Forum (1995-2005)	University (2006-2013)	0.5
Conceptual Maps (1995-2005)	Educational Technology (2006-2013)	0.5
Education (2006-2013)	Learning (2014-2020)	0.5
Education (2006-2013)	Technology (2014-2020)	0.5
European Higher Education Area (2006-2013)	Webquest (2014-2020)	0.5
Educational Technology (2006-2013)	Educational Technology (2014-2020)	0.5
ICTs (2006-2013)	Distance Education (2014-2020)	0.5
University (2006-2013)	Distance Education (2014-2020)	0.5
Competencies (2006-2013)	Evaluation (2014-2020)	0.33

Table 2. Summary of results when using keywords.

Table 3 summarizes the results obtained with the second procedure; i.e., using n-grams of abstracts and titles in the analysis. As it can be seen, the first terms have high inclusion level values, indicating a consolidated relationship between both periods. Additionally, because new terms are added to represent document content, new topics appear, such as social media, flipped classroom and e-learning.

Topic A (Sub-period)	**Topic B (Sub-period)**	**Inclusion**
Internet(2006-2013)	Social Media (2014-2020)	1
ICTs (2006-2013)	Evaluation (2014-2020)	1
Knowledge-Building (1995-2005)	E-learning (2006-2013)	0.5
Electronic Forum (1995-2005)	Collaborative Learning (2006-2013)	0.5
Collaborative Learning (2006-2013)	Flipped Classroom(2014-2020)	0.5
E-Learning(2006-2013)	ICTs (2014-2020)	0.5
ICTs (2006-2013)	Educational Innovation (2014-2020)	0.5
Conceptual Maps (1995-2005)	ICTs (2006-2013)	0.33
ICTs (2006-2013)	Educational Technology (2014-2020)	0.33
Knowledge-Building (1995-2005)	ICTs (2006-2013)	0.25

Table 3. Summary of results when using keywords, abstracts and titles.

Through the expert consultation method, it was determined that the results obtained by this method are related to the development of ideas about teaching and ICTs in the field of research.

ISBN 978 -987-633-574-4

For example, the topics related to "Electronic Forums" may have become "Collaborative Work," since the ideas of collaborative work and learning went through an automation process when tools became available to carry out tasks. Thus, it would make sense for articles that covered tools to start working on conceptual models. This may also explain the transition between "Collaborative Learning" and "Flipped Classroom".

The transition between "ICTs" and "Educational Technology" could also be understood if we consider the discipline that builds conceptual models that include the use of technological tools.

4 Conclusions and future lines of work

This article describes a methodology capable of analyzing thematic evolution in scientific documents. Even though the results obtained come from the analysis of a journal in the domain of computer technology applied to education published in Spain, due to the nature of the text mining and bibliometric methods used, this can be replicated in other domains.

Two different procedures were carried out when creating the sets of terms on which co-occurrence would be measured; this is the metric used at the beginning of the topic identification process. We were able to corroborate that, by adding n-grams previously filtered by their TD-IDF value in the set of documents analyzed, an improvement is observed in the value of the metrics obtained and the identification of the terms that appear as most relevant in the results. Considering both keywords and the terms present in the title and the abstract of each document proved to yield relationships with a higher level of inclusion than when considering only keywords. This is because topic building is enriched and intersections with greater cardinality are obtained. On the other hand, both procedures proved that filtering the relationships by level of inclusion is effective in simplifying visualization. This is an extension of the work presented in [1] on that occasion, the articles (taken from the TEyET journal) were analyzed using only keywords. Term co-occurrence was represented in greater detail using the "density" and "centrality" metrics, but inclusion level was not used.

Once again, our conclusion is that the process used to analyze thematic evolution in Spanish is promising, and different methodologies for extending this work were identified. Firstly, the same analysis could be carried out using metrics other than TF-IDF to select the n-grams generated. This is useful because TF-IDF has certain limitations when applied to large sets of documents [6]. Other result visualization methods were also devised, so that the most relevant results can be highlighted automatically, modifying the threshold or the metric used.

Further extensions to this work are also planned - based on the results obtained, an analysis methodology will be built and validated by comparing its results with those produced by a group of experts. It would also be interesting to build an accessible and simple tool that would allow users who are not experts in computing or data analysis to take advantage of this methodology. These two future lines of work are complementary of each other and will also require the

ISBN 978 -987-633-574-4

design of evaluation devices, both for the experts who validate the methodology and for the potential users of the tool, so that process efficiency and usability can be verified.

References

1. S. Bianco, L. L., and Z. A. Evolución temática de publicaciones en español. una estrategia posible para el diseño de situaciones didácticas. In *XVI Congreso de Tecnología en Educación y Educación en Tecnología (TEyET 2021)*. RedUNCI, 2021.
2. M. Callon, J.-P. Courtial, and F. Laville. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, 22:155–205, 1991.
3. M. Callon, J.-P. Courtial, W. A. Turner, and S. Bauin. From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2):191–235, 1983.
4. N. Coulter, I. Monarch, and S. Konda. Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science*, 49(13):1206–1223, 1998.
5. P. K. Project. Open journal system, 2001. http://pkp.sfu.ca/ojs.
6. J. Ramos et al. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, volume 242, pages 29–48. Citeseer, 2003.
7. S. Robertson. Understanding inverse document frequency: on theoretical arguments for idf. *Journal of documentation*, 2004.
8. C. Sternitzke and I. Bergmann. Similarity measures for document mapping: A comparative study on the level of an individual scientist. *Scientometrics*, 78:113 – 130, 2009.

ISBN 978 -987-633-574-4

Aportes para pensar la educación en pandemia desde la accesibilidad

Javier Díaz, Ivana Harari, Alejandra Schiavoni, Paola Amadeo, Soledad Gómez, Alejandra Osorio

LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas
Facultad de Informática - Universidad Nacional de La Plata
La Plata, Argentina
jdiaz@unlp.edu.ar, iharari@info.unlp.edu.ar, ales@info.unlp.edu.ar,
pamadeo@linti.unlp.edu.ar, sgomez@cespi.unlp.edu.ar, aosorio@unlp.edu.ar

Resumen. La pandemia que comenzó a principios del año 2020 y la situación de aislamiento virtualizaron en forma muy rápida las propuestas educativas presenciales. La modalidad a distancia nos enfrentó a la revisión de las prácticas docentes incluyendo la planificación, evaluación y seguimiento de los estudiantes. Las herramientas de videoconferencia emergieron como el único recurso que permitió los encuentros sincrónicos y la interacción en forma colaborativa. En este artículo, se presenta el análisis de accesibilidad de los sistemas de videoconferencia que utilizan los estudiantes con discapacidad en las asignaturas de la Facultad de Informática de la Universidad Nacional de La Plata. Se analizaron las funcionalidades provistas para discapacidad, su cumplimiento y conformidad de estándares internacionales de accesibilidad y además fueron testeadas manualmente en distintos escenarios de interacción. Se proponen también un conjunto de recomendaciones como aporte para mejorar el abordaje de estrategias docentes con tecnologías, teniendo en cuenta las distintas situaciones de los estudiantes con discapacidad.

Palabras claves: accesibilidad, sistemas de videoconferencia, educación en pandemia.

1 Introducción

La situación de aislamiento provocada por la pandemia, hizo que en el inicio del ciclo lectivo 2020, fuera no sólo indispensable sino urgente adecuar las estrategias de enseñanza en todos los niveles educativos. La modalidad a distancia, que experimentamos por primera vez y sin suficiente preparación previa, nos enfrentó a la revisión de nuestras prácticas docentes, y a la revisión desde la planificación pasando por la cursada, evaluación, seguimiento y otros aspectos.
Al mismo tiempo, herramientas como la conferencia web, una herramienta bidireccional y sincrónica, emergió como el único recurso que nos permitió tener un encuentro simultáneo e interactuar en forma colaborativa. La envergadura, que estas herramientas cobraron durante los periodos de aislamiento, estuvo en relación a: un aspecto pedagógico, a la consolidación de vínculos, una potencialidad educativa, la comunidad virtual y un aspecto social, en tanto permitió la inclusión de participantes que se encuentran dispersos geográficamente. Hace algunos años los sistemas de conferencias web eran bastante limitados pero actualmente se puede contar con sistemas interesantes de videoconferencias que ofrecen una diversidad de funcionalidades, constituyendo una herramienta con un potencial pedagógico significativo.

Si bien las herramientas de videoconferencia, en muchos casos potentes, nos permiten un determinado acercamiento, éste no es comparable con el presencial. El medio virtual, el software intermediario, los problemas de comunicación, la interferencia, los ruidos, hacen que dicha virtualidad se haga presente y la distancia se sienta. En el caso de las personas con discapacidad, las limitaciones que presentan las herramientas de videoconferencia (problemas de comunicación, la interferencia, los ruidos) se profundizan aún más, dificultando el desarrollo de prácticas educativas que promuevan procesos de formación y aprendizaje. El no poder conectarse, o conectarse pero no poder acceder al material completo, o no poder escuchar lo que el docente está explicando, o no poder ver las diapositivas que el docente está compartiendo a través de la videoconferencia, agudiza aún más las distancias. Esta brecha genera un distanciamiento difícil de sortear si no es a través de la implementación de estrategias específicas, que promuevan alternativas consensuadas con los 1500 estudiantes con discapacidad. La inclusión de estas dinámicas, se configuran como puentes que la cátedra debe posibilitar, que deben ser construidos en conjunto con la persona con discapacidad atendiendo sus necesidades de comunicación, adaptando el material que provee y su forma de abordarlo, según la especificidad de su caso y de sus limitaciones que pueden ser mitigadas si las condiciones requeridas son otorgadas.

El tema de accesibilidad web, representa una línea de investigación de mucha relevancia, que se trabaja en la Facultad de Informática desde el año 2002, y que se incorporó al plan de estudios de las carreras de Licenciatura en Informática y Licenciatura en Sistemas. Es importante destacar que se institucionalizó su abordaje mediante la creación de una Dirección de Accesibilidad desde el año 2010 que tiene un contacto estrecho con los

estudiantes con alguna discapacidad. También, se fortaleció el trabajo con la temática de discapacidad, a partir del desarrollo de tesinas, trabajos de cátedra, proyectos de innovación y desarrollo con estudiantes de Informática y diferentes proyectos de extensión acreditados por la Universidad Nacional de La Plata.

En este artículo, se presenta el análisis de accesibilidad de los sistemas de videoconferencia que utilizan los estudiantes con discapacidad en las asignaturas de la Facultad de Informática de la Universidad Nacional de La Plata. Los sistemas analizados fueron Zoom, Big Blue Button y Webex, y para ello, se realizaron una serie de testeos y comprobaciones de tales aplicaciones. Se analizaron las funcionalidades provistas para discapacidad, su cumplimiento y conformidad de estándares internacionales de accesibilidad y además fueron testeadas manualmente en distintos escenarios de interacción tales como el uso de lectores de pantalla, la interacción sólo con teclado, o únicamente con el uso del mouse y un teclado virtual, el uso de magnificadores, y otras condiciones como la navegación sin hojas de estilo, sin imágenes y sin JavaScript. Se describe, también el sistema de gestión de aprendizaje Moodle que cumple con varias pautas de accesibilidad, según las normas internacionales. Además, se plasman recomendaciones importantes a tener en cuenta sobre la accesibilidad principalmente desde el punto de vista de estos sistemas de videoconferencia que se están utilizando para suplantar las clases presenciales.

2 Educación virtual y accesibilidad

Pensar las problemáticas de accesibilidad en medio del contexto de la educación remota o educación de emergencia, propone todo un desafío. La rapidez con que las propuestas educativas presenciales se virtualizaron en contexto de ASPO, homogeneizó las estrategias pedagógicas, que en cuestiones sobre temas de accesibilidad merecen un tratamiento desde la diversidad. Se conjugan muchos factores cuando se trasladan las actividades académicas a un plano totalmente virtual que afecta al proceso de enseñanza y aprendizaje, alternando todo su contexto. Lo que antes se manifestaba dentro de un aula, ahora se conjugan: las condiciones particulares de cada sujeto sea docente, estudiante o auxiliar terapeuta o intérprete en el caso de los estudiantes con discapacidad, que interviene en el proceso educativo. Cada entorno particular que antes quedaba "rezagado o invisibilizado" ante el contexto del aula junto a compañeros en un lugar externo al hogar conformando un ambiente aparte, social e integrador, ahora se hace "evidente" y "afecta" e incide en el proceso educativo.
El medio tecnológico, el entorno particular, las características específicas de cada participante que interviene en el proceso formativo, el software elegido para las actividades académicas ya sea para repositorio de materiales y recursos educativos, para la comunicación, para las consultas, para las clases en videoconferencias, para la evaluación, va configurando "un aula virtual" única, específica y particular para cada uno. Esto afecta de distinta manera, su nivel de percepción, comprensión, comunicación, adquisición de conocimientos y posterior aplicación.

Entendiendo que el colectivo de personas con discapacidades de ningún modo es homogéneo, se avanza hacia la idea de efectivizar una verdadera inclusión educativa para todos los estudiantes desde la perspectiva de derechos y basada en la particularidad y las diferencias individuales. Asumimos el desafío de trabajar bajo el enfoque del paradigma social que se convierte en la base de algunas de las intervenciones pedagógicas que llevamos adelante y sobre las cuales investigamos. Nos aporta penser, el trabajo con las discapacidades "sobre la integración educativa, y conceptos tales como necesidades educativas derivadas de las discapacidades toman relevancia, en la intención de marcar que la discapacidad de la persona está más ligada a "barreras para el aprendizaje y la participación" producidas en el entorno (Booth & Ainscow, 2002) que a la deficiencia misma" [1].

En términos normativos, la Ley 24.521 sobre la Educación Superior de la Nación Argentina, que las universidades, los institutos universitarios y los institutos de educación superior deben respetar, estipula que se deben tener políticas de inclusión educativa tomando medidas para dar iguales oportunidades y posibilidades a las personas con discapacidad, y que puedan acceder al sistema sin discriminaciones de ningún tipo [2]. A su vez, el espíritu del Estatuto de la propia Universidad Nacional de La Plata, que desde su preámbulo la define como una institución pública y gratuita de educación superior, que se debe ofrecer abierta e inclusiva para toda la sociedad estableciendo políticas que tiendan a facilitar el ingreso, permanencia y egreso de los sectores más vulnerables de la sociedad. También indica, la ejecución de políticas con el objeto principal de propender al mejoramiento constante de la calidad de vida de los integrantes de la comunidad universitaria, a la vez que garantizar la efectiva igualdad de oportunidades para el acceso a la educación superior.

A nivel de accesibilidad digital, la Argentina cuenta con un marco legislativo desde el año 2010 cuando se aprobó por unanimidad la Ley 26.653 sobre accesibilidad de la información en las páginas web. La misma estipula que los sitios nacionales gubernamentales deben respetar las normas de accesibilidad de la W3C, específicamente los criterios A y AA de las WCag 2.0 [3]. Como la UNLP es una institución nacional, los sitios pertenecientes al dominio unlp.edu.ar están incluidos en el alcance de esta normativa [4].

ISBN 978 -987-633-574-4

El desafío de integrar, asociados la responsabilidad de promover espacios educativos basados en los principios de igualdad e inclusión, trasciende las barreras de las modalidades presencial, virtual, sincrónico y asincrónico en tanto forman parte de las responsabilidades éticas a las que como docentes nos hemos comprometido. La recuperación de un enfoque de integración que nos permite pensar más allá de las discapacidades, no busca invisibilizar sino pensar su complejidad, asumiendo los condicionamientos sociales, culturales y educativos en los que están inmersos los sujetos. Entender y repensar las prácticas docentes desde la complejidad y en pos de la igualdad e inclusión nos permitirá pensar la Universidad como "La escuela de diseño inclusivo pasa a ser así parte de una filosofía de vida que reconoce la diversidad como valor y como fuente de enriquecimiento, cuyo éxito se plasma en los logros de todos y cada uno/a de los/as alumnos/as y en el desarrollo y bienestar de la comunidad. Este tipo de educación para la inclusión -concebida como educación para toda la vida- debería ser accesible especialmente para los grupos más vulnerables y marginados, destacándose la importancia de que este proceso continuo y gradual se inicie tempranamente y pueda ser articulado a lo largo de los diferentes niveles de enseñanza" [1].

3 Características de accesibilidad del Sistema de Gestión de Aprendizaje Moodle

Las plataformas virtuales o LMS constituyen hoy herramientas fundamentales para el aprendizaje y enseñanza a través de Internet. En el mercado actual cada vez más empresas suman nuevas plataformas entre sus productos ofrecidos siendo Moodle [5] como producto open source y Blackboard las más antiguas y mejor posicionadas en cuanto a su uso y versatilidad de la solución ofrecida. Si bien es una realidad que cada solución cuenta con sus pros y sus contras, al momento de elegir dependerá de los recursos identificados como prioritarios, costos de las licencias, servicios de infraestructura ofrecida, entre otros.

A continuación se describe Moodle, siendo un estándar de facto en los LMS de código abierto, además de ser la herramienta que se utiliza desde el año 2003 en la Facultad de Informática, en gran variedad de cursos y destinatarios, además de seguir entre las primeras en las tendencias globales como plataforma de e-learning y ser propuesta como plataforma virtual de cursos en línea por el consorcio SIU de las Universidades Nacionales de Argentina [6].

Moodle es un estándar de facto para entre las plataformas de aprendizaje virtual, siendo una de las utilizadas en el dictado de los cursos de la Facultad de Informática de distintos años. Desde sus inicios en el año 2001, buscó generar una comunidad que llegue a la mayor cantidad de personas en todo el mundo, adoptando una filosofía de software libre. Actualmente la comunidad está formada por administradores, docentes y usuarios que colaboran de distinta manera generando espacios de intercambio y colaboración entre todas las partes interesadas.

El objetivo de Moodle es que sea usable y accesible para todos los usuarios, más allá de sus capacidades, dispositivos de acceso, contexto, entre otros. Los desarrolladores del núcleo de Moodle se encuentran comprometidos en el uso de estándares de la Web, la Web Content Accessibility Guidelines WCAG 2.1 nivel AA [7], Authoring Tool Accessibility Guidelines ATAG 2.0 [8] para la generación de contenido, Accessible Rich Internet Applications ARIA 1.1 [9] para el contenido enriquecido para crear presentaciones interactivas y atractivas así como la sección 508 de la normativa de Estados Unidos. Es importante destacar también que la plataforma cuenta con muchísimos módulos, que los usuarios habilitan y deshabilitan, utilizan temas personalizados y es posible realizar diferentes configuraciones de acuerdo a las necesidades de la institución y el conocimiento del administrador. Esto implica que es imposible asegurar el 100% de accesibilidad en la plataforma Sin embargo, la siguiente frase ilustra la filosofía de Moodle "La accesibilidad no es un estado, es un proceso de mejora continua en respuesta a nuestros usuarios y el mayor ambiente técnico" Accesibilidad en Moodle. [10]

Desde la versión 2.7 tiene soporte para lectores de pantalla y se incluyen en sus notas de versión. A continuación se presentan las principales consideraciones respecto a la generación de contenido a través de los recursos y actividades más populares y temas personalizados.

Dado que el entorno Moodle brinda herramientas para construir contenido, así como también para consumirlo, es importante analizar el cumplimiento de las normas ATAG 2.0. Tanto para que las herramientas sean accesibles como también el contenido generado sea accesible. Es interesante la herramienta ATAG Report Tool [11] que provee la W3C para asistir en la construcción de reportes que permiten evaluar la accesibilidad de una herramienta de generación de contenidos.

El editor de texto Atto de Moodle permite generar contenido accesible. Se encuentra disponible desde la versión 2.7, es una alternativa al tradicional TinyMCE HTML y texto plano. Atto [12] pone el foco en la usabilidad y la accesibilidad, siendo la mejor alternativa para cualquier tipo de usuario. Permite generar contenido accesible. Desarrollado en Javascript, específicamente para Moodle. El editor Atto ofrece opciones de configuración a través del panel Administrador, como que aparezca o no el validador de accesibilidad, el

ayudante de lector en línea, el color de fondo, entre otros. Desde el perfil docente o editor del curso, este editor ofrece validaciones de accesibilidad de acuerdo al estándar HTML5 y la WCAG desde la barra de herramientas, como es el botón de comprobación de accesibilidad y el lector de pantalla.

Por ejemplo, el recurso Página, despliega una serie de campos a desarrollar, como la Descripción y el Contenido. Presionando el botón comprobación de accesibilidad se validan los siguientes puntos:

- Contraste adecuado de fuente y fondo de acuerdo a las recomendaciones de la WCAG2.0
- Organización jerárquica de contenido. Uso adecuado de títulos, subtítulos, párrafos, et.
- El uso de ìconos con descripciones y usos adecuados y las recomendaciones de ARIA para no resultar redundantes en el contenido.
- Soporte de teclado. Todas las componentes deben ser accesibles desde el teclado y recibir el foco adecuado.
- Soporte para contenido enriquecido WAI ARIA. Las imágenes o elementos decorativos incluyen el atributo ARIA “presentation” o aria-hidden=”true”.
- Páginas con título adecuado en todas ellas. Descriptivo, completo y jerárquico.
- El contenido matemático desarrollado con Mathjax, validación de las imágenes con texto alternativo, entre otros.

El selector de archivos, que permite seleccionar imágenes y fotos en los distintos recursos, brinda un espacio donde completar un texto descriptivo de la imágen. Los temas son una herramienta muy útil, que permite configurar el look and feel de la instalación de Moodle en forma rápida y completa. En el menú administración de Moodle, la sección Plugins - Temas de la plataforma, se presentan los temas por orden de cantidad de descargas. También en el artículo de enero de este año presenta una lista de los 10 mejores temas [13], estando Adaptable y Moove entre los destacados. Tanto Moove como Adaptable son accesibles.

Como se puede observar, no sólo la plataforma de gestión de aprendizaje debe ser accesible sino también el contenido que se genera y el análisis de las herramientas complementarias a las clases virtuales, como son las sesiones sincrónicas utilizando herramientas de videoconferencias. En la siguiente sección se presenta un análisis de accesibilidad de las herramientas de videoconferencia utilizadas en la Facultad de Informática, durante este período.

4 Análisis de las herramientas de videoconferencia según la discapacidad

Como se mencionó anteriormente, debido al contexto de salud, dio lugar a restricciones de movilidad y de confinamiento social preventivo, que se tuvo que implementar de manera obligatoria y rigurosa en la mayoría de los países, y Argentina no fue la excepción. El uso de dispositivos y medios tecnológicos digitales, específicamente de la Web, se convirtió en un medio demandante y obligatorio, principalmente para continuar en el proceso de enseñanza y aprendizaje.

Las clases presenciales se convirtieron en encuentros sincrónicos a través de sistemas de videoconferencias, las consultas, en foros, las instancias evaluativas, en sistemas de examen remoto, con cámaras y micrófonos encendidos, intentando que el proceso educativo no se interrumpa y pueda establecerse.

En el caso de las personas en situación de discapacidad, no sólo basta con que las clases, los materiales y demás recursos educativos estén disponibles en Internet, sino que además dichos medios y herramientas digitales provistas, sean accesibles.

Por esto, en este artículo se trata de analizar el nivel de accesibilidad de las herramientas digitales que se utilizan para realizar las clases sincrónicas remotas, como lo son los sistemas de videoconferencias, más allá del cumplimiento de estándares internacionales de accesibilidad mencionados por las propias herramientas desde la perspectiva de estudiantes de educación superior con distintas discapacidades.

Puntualmente, se analizaron las herramientas Zoom v5.4.6 [14], Big Blue Button 2.2 [15] y Webex 39.3.0 [16], las cuales fueron utilizadas en las asignaturas donde se contaron con estudiantes con discapacidad de la Facultad. En esta sección se detalla el análisis realizado con herramientas automáticas tales como Wave [17] y las validaciones manuales se llevaron a cabo utilizando el lector de pantalla NVDA [18], la extensión Web Developer [19], Colour Contrast Checker [20] el teclado virtual y la lupa de Windows, Insights for Web [21].

Más adelante se presenta un cuadro comparativo entre las distintas herramientas, respecto al grado de cumplimiento en cuanto a:

- Validadores automáticos de accesibilidad: se validaron las herramientas con los validadores de accesibilidad. Es posible que se cumpla con un nivel adecuado en todas las pantallas o presente dificultades en secciones específicas. Por ejemplo en WebEx, el validador detecta algunos problemas de accesibilidad como problemas de contrastes y algunos íconos sin textos alternativos, también presenta algunos problemas en la estructura de encabezados y algunos sectores carecen de ellos. En la Fig. 1 se presenta la evaluación automática de accesibilidad de WebEx utilizando Wave.

ISBN 978 -987-633-574-4

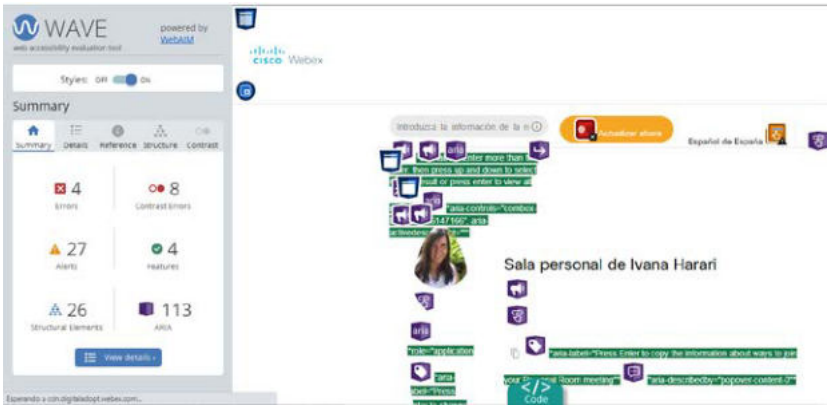


Fig. 1. Evaluación de la accesibilidad de Webex utilizando Wave

- Lectores de pantalla: se utilizó la herramienta con el lector de pantalla activado. Zoom no presentó mayores dificultades pero en BBB, el lector no lee la ventana de "entrar como oyente" o que se está activando el dispositivo de audio. Por su parte WebEx no lee los íconos de copiar invitación así como tampoco el actualizar calendario y lee el nombre interno de la componente como ser botón501139, haciendo muy confusa la interacción. Tampoco, el lector lee el ícono de Ayuda. Y algunas frases aparecen en inglés. Está característica es particularmente relevante para las personas ciegas o con visión reducida.
- Contraste: se evaluó el nivel de contraste adecuado en distintos escenarios como al utilizar fuentes pequeñas o la lupa magnificadora. Las 3 herramientas presentaron problemas en determinados sectores donde se utilizan letras muy pequeñas, íconos y botones cuya área de cliqueo es inferior a la recomendable. Esto es particularmente significativo para las personas con visión reducida.
- Diseño responsivo: la adecuación automática de la herramienta a los distintos dispositivos es relevante para operar el software en forma adecuada, más allá del dispositivo que se utilice. En este sentido, BBB permite acceder a las funciones que no se visualizan en el marco de la ventana. Mientras que WebEx también presenta algunos problemas en este sentido.
- Subtitulado: se analizaron las posibilidades de las herramientas para la gestión adecuada de subtítulos, como la posibilidad de contar con subtítulos en línea o un rol especial durante las sesiones o la posibilidad de grabar las clases para su posterior subtitulado. Está característica es particularmente relevante para las personas sordas. En la Fig. 2 se puede observar las facilidades de subtitulado de Zoom.

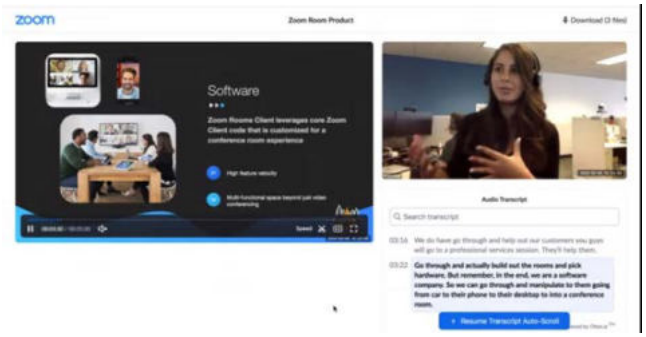


Fig. 2. Subtitulado utilizando Zoom

- Navegación con teclado: se interactúo con la herramienta utilizando sólo el teclado como mecanismo de interacción. Se verificó el orden adecuado de tabulación y los accesos directos a las funciones principales. Si bien las 3 herramientas no presentaron grandes dificultades, WebEx presentó algunas dificultades en la configuración del audio que no presenta una opción de salida y en el manejo de las barras de aumento o disminución del nivel. También presenta algunos menús pop-up, que se despliegan al presionar el ícono de puntitos que es cíclico si se lo opera desde el teclado con la tecla TAB. Esto es correcto, pero debería tener una cruz para poder salir de allí. Está característica es particularmente relevante para las personas con discapacidad motriz. En la Fig. 3 se presenta un ejemplo de esquema de navegación utilizando teclado en BBB.

ISBN 978 -987-633-574-4

Fig. 3. Esquema de navegación utilizando teclado en BBB

- Navegador con JavaScript y hojas de estilos deshabilitado: el único caso que aplicó está validación fue BBB, donde deshabilitando las hojas de estilo y JavaScript se puede interactuar adecuadamente. Particularmente relevante para la navegación utilizando distintos dispositivos y versiones antiguas de navegadores.

En la Tabla 1 se presenta un resumen del análisis realizado para las distintas herramientas de videoconferencia mencionadas.

Tabla 1. Resumen de las características de accesibilidad de las herramientas de VC analizadas

	Zoom	**WebEx**	**BBB**
Lector de pantalla	Cumple	Cumple parcialmente	Cumple parcialmente
Diseño responsivo	Cumple	No cumple	No cumple
Validador de accesibilidad	Cumple	Cumple parcialmente	Cumple parcialmente
Deshabilitar estilos y Javascript	No aplica	Cumple	No aplica
Subtitulado en línea	No cumple	No cumple	No cumple
Rol para subtitulado manual	Cumple	Cumple	Cumple
Transcripción sobre reuniones grabadas	Cumple	Cumple	Cumple
Grabado de las sesiones	Cumple	Cumple	Cumple
Navegación con teclado, orden de tabulación	Cumple	Cumple	Cumple
Navegación con teclado entre las opciones más importantes	Cumple	Cumple	Cumple parcialmente
Contraste	Cumple parcialmente	Cumple parcialmente	Cumple parcialmente
Lupa	Cumple	Cumple	Cumple

Tomando sólo los puntos que aplican a las 3 herramientas, se puede observar que Zoom cumple con el 81,81% de los aspectos evaluados, mientras que BBB cumple con el 63,63% y WebEx con el 45,45%. Por su parte Zoom cumple parcialmente con el 9,09% de los aspectos evaluados mientras que BBB cumple parcialmente con el 27,27% y WebEx con el 36,36% de las características analizadas. Finalmente, Zoom no cumple con el 9,09%, BBB y WebEx no cumplen con el 18,18% de las características mencionadas.

ISBN 978 -987-633-574-4

5 Recomendaciones generales

A raíz del análisis realizado y recuperando un enfoque educativo que sobre la inclusión, propone la integración de estrategias para el abordaje de las problemáticas que afrontan las personas con discapacidad, se puede realizar una serie de recomendaciones para docentes que utilicen sistemas de videoconferencias como Zoom, Webex y Big Blue Button (BBB).

Se considera una recomendación fundamental, relevar las condiciones de los estudiantes con discapacidad para lo cual resulta pertinente: preguntar a los estudiantes y en especial a los estudiantes con discapacidad si cuentan con los recursos tecnológicos que se le exige al utilizar estrategias virtuales a distancia. En este punto es importante relevar que nivel de alfabetización digital poseen para ofrecer herramientas que faciliten la interacción. En este sentido, consensuar formas y estrategias de comunicación constituye un factor importante a la hora de pensar las clases por videoconferencia. En el caso de los estudiantes con discapacidad, se hace imprescindible poder contar con reuniones personalizadas mediante una videoconferencia donde haya pocas personas. Así por ejemplo, los estudiantes sordos pueden leer los labios, o se puede permitir la presencia de la intérprete. Lo mismo ocurre con las personas con dislexia y displasia, y con aquellas personas con problemas motrices que le dificultan el tipeo.

En términos de las estrategias docentes, adecuar los tiempos, las estrategias y los materiales permite planificar el tiempo, es importante tener en cuenta los tiempos que puede llevar adaptar los materiales para que sean aptos para los estudiantes con discapacidad. Por lo que las exigencias en los tiempos de entrega deben dilatarse. En esta línea es recomendable que cuando se comparte documentos PDFs, videos, diapositivas u otro material educativo, el mismo sea accesible y en lo posible se entregue previamente al estudiante.

Entender el contexto de interacción de los estudiantes con discapacidad, nos posibilita diseñar estrategias y seleccionar las herramientas desde una perspectiva accesible. Así por ejemplo, los estudiantes ciegos requieren manejar todo mediante el teclado y que la aplicación que se utilice admita lectores de pantallas como el Jaws, NVDA o los lectores de los celulares. En el caso de los estudiantes sordos requieren subtitulados o interpretación de lengua de señas, por lo que la elección de aplicaciones de videoconferencias o videollamadas que subtitulen automáticamente o al menos mediante un rol que subtitule es recomendable.

Respecto de las decisiones pedagógicas sobre la técnica, es recomendable, verificar que el acceso a la videoconferencia sea accesible, la invitación a la reunión debe otorgarse por varias alternativas. Whatsapp, correo electrónico, la plataforma Moodle presentan características accesibles. En esta línea, es importante que la aplicación de videoconferencia en sí sea accesible.

Pensar y analizar las funcionalidades y servicios de accesibilidad que ofrecen las aplicaciones de videoconferencia, son acciones que nos permitirán prever que, por ejemplo para las personas sordas, se requiere que se subtitule lo que los docentes están hablando. O nos permitirán adaptar las grabaciones que se realicen de las reuniones. Es importante que todas las reuniones que realice la cátedra sean grabadas para que el estudiante con discapacidad tengan acceso y revisión posterior al encuentro, las mismas deberán incorporar subtitulado a la vez que resulta necesario analizar el contenido visual que no fue oralizado por el docente, y comunicárselo a la persona ciega.

En relación también a los materiales resulta imprescindible, hacer accesible lo que se comparte por pantalla u oralizarlo. Los docentes suelen compartir por pantalla diapositivas o documentos. Estos deben ser accesibles y en lo posible haberlos entregado a los estudiantes previamente. Existen algunas limitaciones que condicionan las posibilidades de trabajo, por ejemplo los lectores de pantalla no leen el material compartido por los sistemas de videoconferencias, ya que el audio lo tiene el docente que está dictando la clase. Se recomienda que el docente evite palabras de ubicación como ser: “Acá se encuentra…”, “deben fijarse esto…”, “como ven acá…”, “esto y lo que hicimos allá..”

Por último, es importante verificar que las notificaciones de la aplicación de videoconferencia funcionen en lectores de pantalla. Estas recomendaciones, son aportes para profundizar el abordaje de estrategias docentes con tecnologías, en el reconocimiento de las múltiples realidades que presentan los estudiantes con discapacidad.

6 Conclusiones

En este artículo se analizó la accesibilidad de la plataforma virtual Moodle, estándar de facto de entornos de aprendizaje virtual en Argentina y en el mundo, y los sistemas de videoconferencia que se están utilizando para suplantar las clases presenciales en la Facultad de Informática. Es importante destacar que no es lo único determinante en el proceso de enseñanza. Tampoco lo es la particularidad de cada estudiante desde su perspectiva en su entorno virtual, ni la formación del docente respecto a cuestiones técnicas y de uso de dispositivos tecnológicos, o a la selección que realice respecto a las aplicaciones de software que utilice en su

proceso de enseñanza, también la especificidad del contenido a enseñar, afecta en todo este entramado. Estos y otros aspectos más, hacen que se requieran canales de adaptación particulares, niveles de estimulación y de contención, consensuados y bien diseñados, considerando la especificidad del estudiante, del medio, del contexto, de los recursos empleados, como también, de lo disciplinar.

Los sistemas de videoconferencia constituyen una herramienta con un potencial pedagógico significativo, pero es necesario analizarlas respecto a sus servicios de accesibilidad para entender el impacto que puede generar cuando son utilizadas por parte de personas con discapacidad. Deben ser seleccionadas y utilizadas de manera tal que todos los estudiantes puedan gozar de este recurso de la manera más equitativa posible. Las características de accesibilidad deben ser consideradas si se pretende garantizar efectividad y eficacia en los encuentros sincrónicos a todo el alumnado. Por tal motivo, en este artículo se realiza un aporte sobre el nivel de accesibilidad de las mismas, teniendo en cuenta sus servicios, si cumplen con la conformidad de las WCAG, si soportan distintos escenarios de interacción y lo más importante, cómo es el impacto frente a los estudiantes con discapacidad que en forma remota, desde la distancia y sin asistencia tuvieron y tienen que acceder a las clases y comprenderlas.

Referencias

1. Sonia L. Borzi ... [et al.] Infancia, discapacidad y educación inclusiva : investigaciones sobre perspectivas y experiencias 1a ed. - La Plata: Universidad Nacional de La Plata. Facultad de Psicología, 2019. http://sedici.unlp.edu.ar/handle/10915/83689
2. Ministerio de Justicia y Derechos Humanos. Ley de Educación Superior Ley 24. 521. http://servicios.infoleg.gob.ar/infolegInternet/anexos/25000-29999/25394/texact.htm
3. Congreso Nacional de la República Argentina; Ley 26653 Accesibilidad de la información en las páginas web. http://servicios.infoleg.gob.ar/infolegInternet/anexos/175000-179999/175694/norma.htm
4. Honorable Asamblea Universitaria: Estatuto de la Universidad Nacional de La Plata (2009) https://unlp.edu.ar/frontend/media/20/120/722e7f1b616ac158e02d148aaeb762aa.pdf
5. Moodle. https://moodle.org/
6. Interfaz SIU-Guaraní - Moodle. https://documentacion.siu.edu.ar/wiki/SIU-Guarani/version3.18.0/interfaces/moodle
7. Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/
8. Authoring Tool Accessibility Guidelines (ATAG) 2.0. https://www.w3.org/TR/ATAG20/
9. Accessible Rich Internet Applications (WAI-ARIA) 1.1. https://www.w3.org/TR/wai-aria-1.1/
10. Accesibilidad en Moodle. https://docs.moodle.org/all/es/Accesibilidad
11. ATAG Report Tool. https://www.w3.org/WAI/atag/report-tool/
12. Editor Atto. https://docs.moodle.org/all/es/Editor_Atto
13. The Best Free Themes for your Moodle-based Learning Environment EVER (New Decade Update) https://www.lmspulse.com/2020/top-10-free-moodle-themes-showcase-learning-environment-moodlethemes/
14. Zoom: Accessibility Features: zoom.us/accessibility
15. Big Blue Button Accessibility: https://docs.bigbluebutton. org/2.2/ html5-accessibility.html#overview
16. Webex Meetings Accesibilidad. https://help.webex.com/es-co/84har3/Webex-Meetings-and-Webex-Events-Accessibility-Features.
17. WAVE Web AccessibilityEvaluation Tool: https://wave.webaim.org/
18. NVDA en Español: https://nvda.es/
19. Web Developer. https://chrome.google.com/webstore/detail/web-developer/bfbameneiokkgbdmiekhjnmfkcnldhhm
20. Colour Contrast Checker. https://chrome.google.com/webstore/detail/colour-contrast-checker/nmmjeclfkgjdomacpcflgdkgpphpmnfe/related?hl=en-GB
21. Accessibility Insights for Web. https://accessibilityinsights.io/docs/en/web/overview/

ISBN 978 -987-633-574-4

Metodologías para el Diseño de Juegos Serios. Análisis Comparativo

Edith Lovos[1], Mónica Ricca[1], Cecilia Sanz[2],

[1] Universidad Nacional de Río Negro, Sede Atlántica, CIEDIS, Viedma, Río Negro
[2] III-LIDI Facultad de Informática Universidad Nacional de La Plata
[2] Investigador Asociado de la Comisión de Investigaciones Científicas de la Pcia. de Bs. As.
{*elovos,mricca*}*@unrn.edu.ar, csanz@lidi.unlp.edu.ar*

Abstract. La valoración de los juegos serios como herramienta educativa se puede reconocer en la literatura sobre el tema, y alcanza a los diferentes espacios formativos. Sin embargo, un tema aún en discusión se vincula a las metodologías o modelos qué permiten guiar el proceso de diseño y producción de un juego serio. En este trabajo se presenta un análisis comparativo entre diferentes metodologías de diseño, recuperadas en el marco de una revisión bibliográfica. Respecto a los enfoques teóricos, los materiales analizados se categorizan en aquellos vinculados al estudio del comportamiento humano, y aquellos enfocados específicamente en aspectos del aprendizaje.

Keywords: Serious Games, Design, Methodology

1 Introduction

En los últimos años los juegos serios comienzan a emplearse como recurso para fomentar y propiciar el aprendizaje en diferentes niveles educativos. Algunas propuestas parten del hecho de que los juegos tienen un papel fundacional en el desarrollo de los seres humanos desde sus primeros años de vida y también en la adolescencia. Señalan las investigaciones que el juego cumple un rol fundamental en la construcción del conocimiento, en la plasticidad cerebral, en su rendimiento académico y también en el desarrollo socioemocional, cognitivo y físico. Según los estudios consultados, los juegos provocan experiencias complejas que activan una serie de funciones ejecutivas como la toma de decisiones [1,2,3] y mejorar el rendimiento en una amplia variedad de tareas y habilidades cognitivas [4]. Dentro de estas últimas, se plantea la mejora de la asignación de recursos atencionales, el cambio de tareas, la promoción y desarrollo de la creatividad, la resolución de problemas, el desarrollo de habilidades espaciales y también otras habilidades no cognitivas como la persistencia y la apertura a nuevas experiencias [5,6,7,8]. Ha sido tal el interés que han generado los juegos con otros propósitos que no sea solo el entretenimiento, que comenzaron a ser llamados juegos serios (JS), es decir aquellos donde se agrega un objetivo educativo en sus diferentes formas. Sin embargo, aunque los aportes de estos juegos son motivadores para su inclusión en una propuesta didáctica, no siempre es posible encontrar aquel que se ajuste a la misma, generando así la necesidad para docentes e investigadores de

involucrarse en el diseño de sus propios juegos a la vez que trabajar con otros perfiles como diseñadores y/o programadores.
En este trabajo presentamos los resultados alcanzados a partir de un proceso de comparación entre metodologías, métodos y framework que pueden ser utilizados para el diseño de juegos educativos. Es importante señalar que el trabajo se realizó en el marco de un proyecto de investigación (PI-UNRN-40C-750) acreditado y financiado por la Universidad Nacional de Río Negro. A través de éste, y siguiendo la metodología de investigación acción participativa, se busca generar conocimiento sobre el diseño, producción e integración de juegos serios en particular aquellos que incluyan tecnologías emergentes como la realidad aumentada y puedan jugarse usando dispositivos móviles.

2 Juegos Serios

En la publicación del libro "Serious Games", Clark Abt [9], define al juego serio como aquel que es diseñado con un propósito primario distinto al puro entretenimiento y explora las diversas maneras en que se puede incluirlo en los procesos de enseñanza y aprendizaje, manteniendo la diversión y el placer para los jugadores.
Los llamados juegos serios (JS) incluyen diferentes propósitos como educar o mejorar las capacidades de los usuarios de una manera entretenida. La mayoría de estos juegos desarrollan escenarios simulados y generan la posibilidad de que los jugadores experimenten situaciones específicas en la virtualidad. Michael y Chen [10] afirman que el hecho de que sean denominados juegos serios no implica que no ofrezcan diversión y entretenimiento, sino que se focalizan aspectos formativos e instructivos, dando lugar en forma entretenida a que el jugador pueda aprender, aplicar y demostrar lo aprendido. Los JS se aplican en diferentes ámbitos, desde formación a empleados, en profesiones que requieren la simulación de escenarios de peligro, y como herramienta para ayudar al cambio de conductas. Con independencia del ámbito de aplicación, un JS contiene cuatro componentes principales: objetivos (educativos), reglas, retos e interacción. A través de la mecánica del juego se generan las acciones que posibilitan construir las reglas y métodos diseñados para la interacción con el juego. Así, la mecánica del juego, da lugar a la comunicación, la puntuación, las recompensas, castigos y el flujo del juego [11].

3 Diseño de Juegos Serios

Llevar adelante el proceso de construcción de un JS, requiere combinar el diseño instruccional con el diseño del videojuego (características del juego, mecánicas, y jugabilidad). De esta forma, la producción implica poner en práctica conocimientos sobre diseño de juegos, teorías de aprendizaje y dominio del contenido a abordar con el mismo [12]. El diseño instruccional permitirá definir los contenidos, las habilidades a desarrollar, las estrategias que se usarán para ofrecer los contenidos y los mecanismos de evaluación, todo ello en base a las necesidades de los estudiantes, sus características y el contexto de aplicación [13]. Para Silva [14], teniendo en cuenta que un JS es un

producto de software, los modelos que permiten guiar el proceso de diseño y producción, se componen en general al menos 4 etapas iterativas: análisis, diseño, implementación y evaluación, y requieren de un trabajo multidisciplinar que involucra: diseñadores de juegos, desarrolladores y expertos en contenidos. El mismo autor, señala que un equipo interdisciplinario como el que demanda la construcción de un JS, exige además avanzar hacia la unificación de un vocabulario que permita mejorar la comunicación.

4 Análisis Comparativo

En el proyecto de investigación mencionado en el apartado Introducción, se llevó adelante una revisión de bibliografía con la intención de recuperar metodologías, modelos y/o frameworks que permitan orientar el proceso de diseño de un juego serio. La revisión se limitó al periodo 2015-2019, y buscó responder a las siguientes preguntas:

- ¿Qué enfoques teóricos predominan en las metodologías?
- ¿Qué tipos de juegos serios es posible diseñar con las metodologías?
- ¿Qué etapas incluyen las metodologías? y qué conocimientos demandan?

Es importante resaltar que se definieron criterios de exclusión tomando como base los aportes de Ávila-Pesántez [26] así se descartaron: artículos en los que se aborda el tema de diseño de JS, pero no se definen explícitamente las etapas o pasos que demanda, y aquellos enfocados en grupos de destinatarios específicos.

En la Tabla 1, se presenta un resumen de las metodologías recuperadas categorizadas por ID (autor/es), tipo(metodología, framework, modelo), enfoque teórico, tipos de JS que permite generar, cantidad de etapas/fases que componen el modelo, aplicación práctica del modelo, y por último país de origen.

Tabla 1. Herramientas de Autor y Frameworks recolectados

ID	Tipo	Enfoque Teórico	Tipo de Juegos	Cantidad de fases/ etapas	Aplicación Práctica	País de origen
[15]	Metodología	Aprendizaje Basado en Problemas	Aventura / Rompecabezas	3	Aplicación en el JS “Clean World”	Portugal
[16]	Framework CBDG	Teoría cognitiva social e Inteligencias Múltiples	Aventura/ Rompecabezas	3	Aplicado en la evaluación del juego “The Journey Project 3”	Australia

ISBN 978 -987-633-574-4

[19]	Metodología SADDIE	Extensión del modelo ADDIE	No específica	6	Producción de JS en Formación docente	Slovenia
[21]	Metodología	Investigación - Acción	Aventura	5	Juego de historia "Ferran Alsina"	España
[22]	Método Co-CreARGBL	SADDIE	Aventura	8	Diseño de JS y formación docente	España
[23]	Modelo ATMSG	Teoría de la actividad	Aventura	2 Fases (4 pasos en total)		Países Bajos
[24]	Extensión ATMSG	Teoría de la actividad + Analíticas de Aprendizaje	Aventura	2 Fases (6 pasos en total)	Juego "Circuit Warz"	Inglaterra
[25]	Metodología	No se indica	Aventura	7	Urano: Invasion of the thieves of planets	España
[26]	Metodología DIJS	Combina diseño instruccional + diseño de videojuegos	JS en general	4 etapas con ingredientes y utensilios	Aplicada al análisis y extensión de JS Desafiate	Argentina

Barbosa et al., [15] proponen una metodología para el diseño y la producción JSs, que busca incluir tareas educativas en paralelo al ambiente principal del juego y para ello emplean diferentes mecánicas (mini-juegos, rompecabezas y cuestionarios), que permiten reforzar objetivos de aprendizaje mediante experiencias que promuevan enlaces, motiven a los jugadores y los "*inviten*" a continuar jugando. Desde la perspectiva pedagógica, esta metodología busca avanzar en línea con el aprendizaje basado en problemas (ABP), así sus componentes principales son: un juego principal con preguntas y un conjunto de mecanismos de aprendizaje que se vinculan con el juego

ISBN 978 -987-633-574-4

principal, que tienen independencia y pueden ser jugados en paralelo al juego principal. El primer paso en el diseño de un JS usando esta metodología, consiste en la creación de la historia que dará soporte al juego y sus diferentes niveles. Luego, cada nivel del JS contará con sus propios mecanismos de aprendizaje, de esta forma el juego principal concentra la diversión y los niveles presentan los retos que materializan el aprendizaje. En resumen, esta metodología permite visualizar la construcción de un JS estructurada por niveles, donde cada nivel se compone de diferentes mecanismos de aprendizaje. Los autores presentan un caso de aplicación de la misma, a través del juego "Clean World", diseñado y desarrollado para jugar con la consola de juegos Xbox. En cada nivel del juego, el jugador tiene que completar misiones para acceder a los retos de aprendizaje y de esta forma poder avanzar en el mismo.

Otra propuesta es la de Starks [16], un framework denominado CBDG por sus siglas en inglés (Diseño de Juegos Cognitivos Conductuales), el cual busca durante el proceso de creación del juego, alinear las metas de éste y con las del aprendizaje. Para ello el framework combina el marco teórico de la teoría cognitiva social de Bandura [17] y la propuesta de inteligencias múltiples de Gardner [18] en un modelo unificado que permita crear juegos para promover el aprendizaje y el cambio de comportamiento, y dónde como remarca la autora[16], la pregunta que guía el proceso se resume en : *"cómo es posible expresar uno o más elementos cognitivos sociales usando mecanismos de inteligencias de una manera que facilite el disfrute proceso"* (pp. 14). Por elementos cognitivos sociales, la autora hace referencia a: conocimiento, autoeficacia, objetivos, planes, resultados esperados, soporte social y obstáculos. Estos pueden ser incorporados en el diseño a través de: librerías de juegos, moderando los niveles de dificultad del juego, así como guardando y restaurando puntos y haciendo uso de personajes no jugadores (NPC) que actúan durante el juego como tutores o coaches. Aunque no se presenta ningún caso de aplicación del framework para el diseño de un JS, en cambio se presentan los resultados de su aplicación para evaluar si los componentes del framework se encuentran presentes en el juego "The Journey Project 3" (1998).

Rugelj[19], con la intención de estimular y mejorar diversas competencias en cursos del profesorado (formación docente), propone para el diseño de JS un modelo al que denomina SADDIE. Este es una extensión del modelo secuencial ADDIE (Análisis, Diseño, Desarrollo, Implementación y Evaluación), utilizado en el diseño de materiales instruccionales. Así, SADDIE se basa en la teoría del aprendizaje activo y consta de seis fases principales: Especificación, Análisis, Diseño, Desarrollo, Implementación y Evaluación. Se espera que al diseñar un JS siguiendo esta metodología, los estudiantes adquieran competencias didácticas, técnicas y habilidades esenciales para el trabajo en equipo. La puesta en práctica de SADDIE, como señala Rugelj [20], permite a sus usuarios desarrollar habilidades para determinar los objetivos de aprendizaje que resultan consistentes con el currículum, seleccionar el/los enfoques de enseñanza que resultan más apropiados para alcanzar los objetivos, implementarlos en el proceso de aprendizaje, así como también preparar las devoluciones y evaluar de los conocimientos adquiridos, entre otros.

Siguiendo esta idea de involucrar a los docentes en el proceso de diseño y construcción de un JS, Contreras-Espinosa y Eguía Goméz [21], proponen la co-creación de JSs siguiendo la metodología de investigación-acción (IA). La IA, permite trabajar de forma espiralada en ciclos, donde se llevan adelante las siguientes fases: diagnóstico,

planificación de la acción, acción, evaluación y especificaciones. Por otra parte, la co-creación implica la participación activa de todos los actores (diseñadores, usuarios, desarrolladores), desde los inicios del proyecto, así cada uno plantea necesidades y se definen los objetivos en forma conjunta. Los autores han aplicado la IA en una experiencia de trabajo multidisciplinario entre investigadores y docentes de nivel primario para el co-diseño de un juego serio. En la misma, se llevó adelante el diseño de un JS sobre historia, que buscó promover competencias del tipo comunicativa lingüística y audiovisual, a la vez que el desarrollo de competencias culturales. Sobre la experiencia, los autores destacan que posibilitó asignarles a los docentes el rol de usuarios activos (incluyéndolos en el proceso de innovación desde la idea a la etapa de prueba). Además, posibilitó a los desarrolladores crear contenidos con un alto nivel de apropiación por parte de los usuarios. Los autores indican que esta forma de trabajo resulta un aporte a la inclusión de juegos en propuestas didácticas, ya que son los docentes quienes finalmente decidirán la inclusión del juego en el espacio de la práctica docente.

En el caso especial de JS que incluyen tecnologías emergentes, Tobar-Muñoz et al., [22] proponen un método de co-diseño, al que han denominado Co-CreARGBL (Co-creación de juegos basados en realidad aumentada para el aprendizaje). El método se compone de tres fases: entrenamiento, diseño interactivo y evaluación en el aula, en las que se incluyen diferentes actividades y roles (Líder de proyecto, Diseñadores, Desarrolladores, Investigadores, docentes y estudiantes). En relación a las etapas y actividades que se llevan adelante en la fase de diseño, las mismas siguen el modelo SADDIE [19] descripto anteriormente. Y en particular, la fase de entrenamiento consiste en introducir a los docentes en los conceptos de RA y aprendizaje basado en juego (GBL, por sus siglas en inglés), de manera que ellos puedan comprender sus potencialidades y beneficios desde una perspectiva práctica. En el caso de Carvalho y otros [23], proponen un modelo conceptual con base en una línea de investigación de las Ciencias Sociales, que estudia las diferentes formas del comportamiento humano y las interacciones que estos generan, conocida como teoría de la actividad (AT). Así, a partir de este modelo se busca describir durante el diseño/análisis de un JS, las formas en que los elementos que lo componen se conectan entre sí durante el juego, y cómo estos contribuyen a alcanzar los objetivos pedagógicos establecidos. El modelo propuesto lo denominaron ATMSG, se compone de tres actividades principales: las vinculadas a la jugabilidad del JS, las vinculadas al aprendizaje (desde la perspectiva del estudiante), y por último las actividades instruccionales (desde la perspectiva del docente). Esta diferencia, remarcan los autores, permite diferenciar posibles conflictos en los motivos que conducen las actividades que podrían afectar los objetivos de aprendizaje del JS. Asimismo, proponen una diferenciación entre las actividades instruccionales: intrínsecas y extrínsecas. Las primeras tienen lugar en forma solitaria dentro del juego, aquí se incluyen aquellas que soportan el aprendizaje (tips, mensajes de ayuda, evaluaciones automáticas entre otras), y las extrínsecas, por el contrario las lleva adelante el instructor/docente, por fuera del juego y en un espacio temporal establecido por el mismo (antes, durante, después) de una sesión del juego, por ejemplo una clase, un curso. La estructura jerárquica del modelo ATMSG, incluye 2 etapas y en cada etapa pasos a realizar. La primera etapa, considerada de alto nivel, corresponde al Análisis de Actividades (se identifican y describen las actividades del JS como red de actividades), y la segunda de un nivel intermedio corresponde al Análisis de

ISBN 978 -987-633-574-4

Acciones (se representa la secuencia del JS, se identifican acciones, herramientas y objetivos, se provee una descripción de las implementaciones). En relación a la usabilidad y utilidad del modelo, los autores indican, que el nivel de detalle demandado, puede dificultar su uso por parte de usuarios no expertos o que están menos familiarizados con los juegos digitales.

Por su parte, Callaghan et al., (2018) proponen una extensión del modelo ATMSG que facilita la identificación, selección e integración de analíticas de aprendizaje en JS. Esta integración forma parte del proceso de evaluación, con la intención de determinar si se alcanzan los objetivos de aprendizaje, y por otra parte proporcionan información en tiempo real sobre las diferentes interacciones que tienen lugar en éste. Así a la estructura original de ATMSG, se agrega un paso más en la fase de análisis de las acciones, que permite mapear acciones con trazas específicas del juego. Los autores presentan un caso de aplicación de la metodología, para el diseño y evaluación del JS "Circuit Warz", destinado a estudiantes de Ingeniería.

Otra propuesta es la De Lope et al., (2017), quienes presentan una metodología de diseño de JS basada en una narrativa interactiva que permite integrar todos los aspectos transversales del juego, sumado a ello, gestiona un conjunto de representaciones visuales para facilitar la comunicación entre los integrantes del equipo de trabajo. La metodología se compone de pasos ordenados e iterativos comprendidos en 6 fases de diseño, teniendo en cuenta que el juego se organiza en capítulos y escenas, donde se aplican técnicas de evaluación a lo largo del proceso, en particular recopilando datos sobre el impacto emocional de la historia del juego. Un aspecto distintivo de la metodología, es el uso de la notación gráfica, como medio de comunicación entre los miembros del equipo, sin importar su perfil. De esta forma, es posible obtener una vista abstracta del juego, que facilita la implementación del mismo y que también puede ser usada por los desarrolladores que no forman parte del diseño. Los autores presentan un caso de aplicación de la metodología en el diseño del JS "Urano" cuyo objetivo es la comprensión lectora.

Siguiendo esta idea de facilitar la comunicación entre los integrantes del equipo de diseño y desarrollo del JS, Silva (2019), propone una metodología práctica que toma como base la propuesta por Barbosa[15], y separa los contenidos de aprendizaje del juego de las mecánicas que se usan para mantener la diversión. Esto se logra a través del uso de un vocabulario simple acompañado de diagramas, que permite identificar los pasos principales involucrados en el diseño: desde la elección del tema hasta la experiencia del usuario. De esta forma, es posible identificar los componentes del juego que darán soporte al proceso de aprendizaje, y a aquellos que se utilizarán sólo con el fin de mantener la diversión y generar en el usuario ganas de continuar jugando. El autor señala, que si bien la metodología no se asocia al diseño de juegos de un género específico, no puede aplicarse al diseño de juegos de simulación, ya que aquí las mecánicas dependen del sistema a simular y el aprendizaje se construye en base a las mismas.

Finalmente, en [26] se presenta una propuesta de metodología denominada DIJS que integra algunas de las ideas encontradas en una revisión de la literatura realizada por los autores. Propone atender tanto al diseño, como al desarrollo y la evaluación del juego serio que se desea crear. Incluye una metáfora encontrada en una de las metodologías estudiadas, en donde se plantea que este proceso es como un menú de platos (etapas), y para cada plato se utilizan ingredientes y utensilios para realizarlo.

Así las etapas de DIJS recomiendan ingredientes y utensilios en cada una, para guiar y viabilizar su realización. Las etapas incluyen desde la definición de los objetivos pedagógicos del juego, la definición del perfil del jugador (atendiendo a cuestiones de interés para su aprendizaje), hasta la evaluación del juego, con estudiantes y docentes, y con foco en la jugabilidad.

4 Discusión y Conclusiones

La inclusión JS en propuestas educativas, se basa no solo en la experiencia lúdica placentera, sino en su aporte como facilitador para construir conocimientos, trabajar la creatividad y la construcción social. En tal sentido, los JS se proponen como un dispositivo de socialización y de promoción del aprendizaje en sí mismos, donde al jugar también se aprende a negociar sentidos, a competir con otros, a justificar, a crear y recrear historias, a apropiarse de roles y valores, a elegir entre diferentes opciones, entre otras acciones. En resumen, los JS como recurso educativo, requieren combinar el diseño instruccional con el diseño específico del videojuego y así el diseño y producción de un juego educativo se convierte en un trabajo multidisciplinario. Las metodologías, frameworks y modelos analizados en el apartado anterior, destacan la importancia de incluir a los docentes en estos equipos de trabajo no solo en las etapas de alto nivel sino a lo largo del proceso. Esto como señalan Rugelj [19] y Tobar- Muñoz [22], su participación no es solo un aporte al diseño y producción del JS, sino que permite que los docentes puedan acercarse, comprender y utilizar diferentes tecnologías entre ellas las emergentes (realidad aumentada, por ejemplo). Sin embargo, como señala Silva [14], esto exige un lenguaje de comunicación común, en algunos casos se avanza adaptando recursos del diseño de software como el Lenguaje de Modelado Unificado (UML) [23, 24], y metodologías usadas en el campo de la educación, ya sea para el diseño de recursos didácticos (ADDIE, por ejemplo), como para la investigación (caso de la investigación -acción). En relación a los enfoques teóricos, se categorizan en aquellos vinculados al estudio del comportamiento humano como: teoría de la actividad y teoría cognitiva social, y aquellos enfocados en aspectos del aprendizaje. Este análisis es el inicio para avanzar en la selección de una metodología específica para la creación de juegos serios educativos, que se realizarán en el marco del proyecto mencionado.

5 References

1. Moncada Jiménez J. & Chacón Araya, Y. (2012). “El efecto de los videojuegos en variables sociales, psicológicas y fisiológicas en niños y adolescentes”, Retos: nuevas tendencias en educación física, deporte y recreación, nº 21, pp. 43-49, 2012
2. Greitemeyer T.& Mügge D. O. (2014). Video games do affect social outcomes: A meta-analytic review of the effects of violent and prosocial video game play, Personality and social psychology bulletin, vol. 40, nº 5, pp. 578-589, 2014
3. Bosch, H. E., Bergero, M. S., Nasso, C., Pérez, M. M., & Rampazzi, M. C. (2017). Innovaciones didácticas para ciencias y matemática asistida por TIC. TE & ET.

4. Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. Acta psychologica, 129(3), 387-398.
5. Green, C. S., Li, R., & Bavelier, D. (2010). Perceptual learning during action video game playing. Topics in cognitive science, 2(2), 202-216.
6. Oei, A. C., & Patterson, M. D. (2014). Playing a puzzle video game with changing requirements improves executive functions. Computers in Human Behavior, 37, 216-228.
7. Jackson, L. A., Witt, E. A., Games, A. I., Fitzgerald, H. E., Von Eye, A., & Zhao, Y. (2012). Information technology use and creativity: Findings from the Children and Technology Project. Computers in human behavior, 28(2), 370-376.
8. Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. Computers & education, 80, 58-67.
9. Abt, C. C. (1987). Serious games. University press of America.
10. Michael D. & Chen, S. (2006). Serious Games. Games that educate, train and informs. Canadá: Thomson, 2006.
11. Taipe, M. S. A., Pesántez, D. Á., Rivera, L., & Vizueta, D. O. (2017). Juegos Serios en el Proceso de Aprendizaje. UTCiencia" Ciencia y Tecnología al servicio del pueblo", 4(2), 111-122.
12. Mestadi, W., Nafil, K., Touahni, R., & Messoussi, R. (2018). An Assessment of Serious Games Technology: Toward an Architecture for Serious Games Design. International Journal of Computer Games Technology, 2018.
13. Fernández-Robles, J. L., & Hernández-Gallardo, S. C. (2019). Diseño instruccional de un juego serio que facilite a niños de tercer grado de primaria el ejercicio de operaciones matemáticas básicas.
14. Silva, F. G. (2019). Practical methodology for the design of educational serious games. Information, 11(1), 14.
15. Barbosa, A. F., Pereira, P. N., Dias, J. A., & Silva, F. G. (2014). A new methodology of design and development of serious games. International Journal of Computer Games Technology, 2014.
16. Starks, K. (2014). Cognitive behavioral game design: a unified model for designing serious games. Frontiers in psychology, 5, 28.
17.Bandura, A., and Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. J. Pers. Soc. Psychol. 41, 586–598. doi: 10.1037/0022-3514.41.3.586
18. Gardner, H. (1983). Frames of Mind: The Theory of Multiple Intelligences. New York: Basic Books
19. Rugelj, J., Zapušek, M.: Achieving teacher's competencies in the serious games design process. In: Busch, C. (Ed.) Proceedings of the 8th European Conference on Games Based Learning ECGBL 2014. Academic Conferences and Publishing International Limited, Sonning Common (2014)
20.Rugelj, J. (2015). Serious games design as collaborative learning activity in teacher education. IN: Busch, C. (ed.).Proc. of the 9th European Conference on Games Based Learning: Steinkjer, Norway 8-19 October 2015. Reading: Academic Conferences and Publishing International Limited, 456-460.
21.Contreras-Espinosa, R. S., Eguia-Gómez, J. L., & Solano Albajes, L. (2016). Investigación-acción como metodología para el diseño de un serious game. RIED: revista iberoamericana de educación a distancia, 19(2), 71-90.
22.Tobar-Muñoz, H., Baldiris, S., & Fabregat, R. (2016). Method for the Co Design of Augmented Reality Game-Based Learning Games with Teachers. In Proceedings of the VIII International Conference of Adaptive and Accessible Virtual Learning Environment (pp. 103-115).

ISBN 978 -987-633-574-4

23.Carvalho, M. B., Bellotti, F., Berta, R., De Gloria, A., Sedano, C. I., Hauge, J. B., ... & Rauterberg, M. (2015). An activity theory-based model for serious games analysis and conceptual design. Computers & education, 87, 166-181.
24.Callaghan, M., McShane, N., Eguíluz, A., & Savin-Baden, M. (2018). Extending the activity theory based model for serious games design in engineering to integrate analytics.
25. De Lope, R. P., Arcos, J. R. L., Medina-Medina, N., Paderewski, P., & Gutiérrez-Vela, F. L. (2017). Design methodology for educational games based on graphical notations: Designing Urano. Entertainment Computing, 18, 1-14.
26. Archuby, F. H. (2020). Metodologías de diseño y desarrollo para la creación de juegos serios digitales (Doctoral dissertation, Universidad Nacional de La Plata).

ISBN 978-987-688-574-8

CACIC 2021

WORKSHOP COMPUTACION GRAFICA, IMAGENES Y VISUALIZACION

COORDINADORES

Silvia Castro (UNS)
Roberto Guerrero (UNSL)
Oscar Bría (INVAP)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978 -987-633-574-4

Virtual Reality Volumetric Rendering Using Ray Marching with WebGL

Federico Marino[1], Horacio Abbate[1] and Ricardo A.Veiga[1],

[1] Universidad de Buenos Aires, Facultad de Ingeniería, Buenos Aires, Argentina
{fmarino, habbate, rveiga}@fi.uba.ar

Abstract.
This work is concerned with virtual reality web applications that implement volumetric visualizations based on Ray Marching, combined with traditional triangle-based rendering. The use case selected was an interactive tool to navigate through a 3D model of sedimentary basins and natural reservoirs of oil and gas. The surfaces are described by signed distance functions computed in the GPU within a pixel shader. A procedural 3D texture defines the distribution of sedimentary strata inside the volume and provides visual clues of the internal structures. In VR mode, the user can navigate the scene by moving his body. He can subtract portions of the 3D volume in real time through a set of Boolean operators (plane, cylinder or sphere) using VR hand controllers. A virtual control panel, dynamically generated from a JSON file, allows parameters to be adjusted within VR. The application can also be used in desktop mode.

Keywords: Ray Marching, Virtual Reality, Sedimentary Basin Visualization, Oil and Gas Industry, WebXR

1 Introduction

There are many uses for Virtual Reality (VR) systems in the gas and oil industry with the purpose of improving the efficiency and reducing the risks of exploration and operation [1]. Even though these techniques were widely applied by the industry for training, their uses were extended to other areas like exploring, production, and operation and maintenance of surface installations.

As Jampeisov [1] pointed out, a person receives up to 80% of the information of the world through sight and the use of 3D visualization helps to enhance the efficiency of analyzing large amounts of information. These ideas led to an increasing interest in using different computer-based technologies to train engineers [2].

The inclusion of computer and information technologies (TICs) in teaching methodologies is considered an alternative to improve the comprehension and use of scientific and technology models [3], [4], [5].

The present paper shows the implementation of a volumetric rendering solution based on ray marching that is compatible with WebGL 1.0 and WebXR and can run on a web browser. This project was financed by a PIDAE 2018 (1-116) grant.

ISBN 978 -987-633-574-4

1.1 The Problem to Solve

The structure of a seismic image of an oil basin is shown in Figure 1. This kind of schematic representation is commonly used and it consists in a sequence of bidimensional cuts. Each cut has a label that indicates the type of geological component (e.g.: the basin and its geometry, the filling represented by strata affected by geological faults).

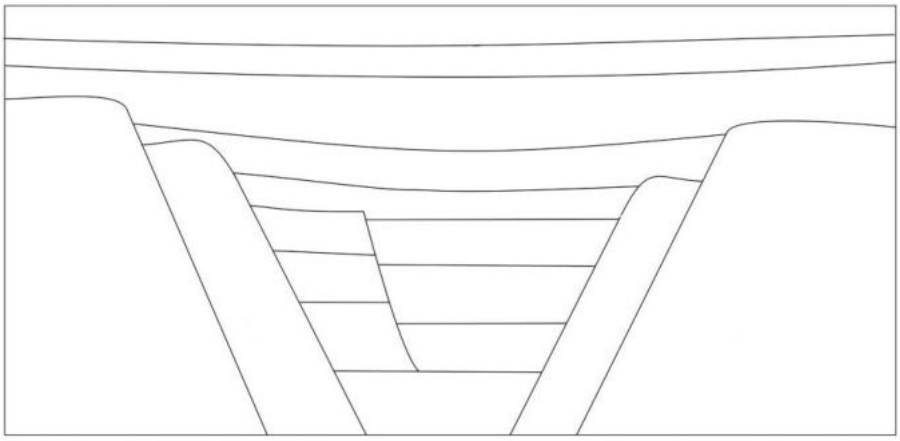


Fig. 1. Initial sketch of a simulated geological section for the RV model.

So far, the user had to analyze a two-dimensional scheme like the one shown in figure 1, in order to perceive the three-dimensional shape of each layer like porous and possibly permeable media. The VR application allows the user not only to see the bidimensional images, but also to get a tridimensional insight of the basin and to explore inside in a dynamic and intuitive way. This way the lithology, the textures and the sedimentary structures, and also the physical properties related to the context could be analyzed.

1.2 VR Devices

VR devices are more accessible to the general public nowadays. A high-resolution screen and an optical system provide an independent image to each eye for stereoscopic views. These systems also have a set of sensors to determine the attitude and the position of the user's head into the three-dimensional space. Many of them have wireless joysticks (hand controllers) whose orientation and position are also sensed.

In this project, an Oculus Quest device connected to a PC computer (which does the graphics processing) was used. The GPU used was a Nvidia RTX 2070.

ISBN 978 -987-633-574-4

1.3 WebGL, WebVR and WebXR Standards

The WebGL (Web Graphics Library) is a standard specification for an API implemented in Javascript for 3D graphics rendering by any web browser. It runs on any platform that supports OpenGL 2.0 and OpenGL 2.0 ES.

The WebVR standard was created by Mozilla in 2014, and it offers access to VR devices like HTC Vive, Oculus Rift, Oculus Quest or Google Cardboard from a web browser. The last version, launched in 2017, depends on special web browsers to run, like Firefox Nightly. The standard was never completely implemented and was finally substituted by the WebXR [6].

The WebXR standard brings together support for Virtual Reality and Augmented Reality applications on the same API (XR = AR + VR); as the devices for both applications have some common needs, like sensing the position and orientation, and also rendering the images from the corresponding point of view.

As of August 2021, WebXR is still a working draft and it's not available for all web browsers.

1.4 The Platform Chosen and the Solution Overview

A web application based on Javascript, HTML, CSS, WebGL and WebXR was implemented. Some open-source libraries such as React.js and Bootstrap were used to build the graphical user interface (2D menu).

Additionally, Three.js library was used for the management of 3D meshes, object hierarchies, cameras, light sources, materials, shaders and rendering.

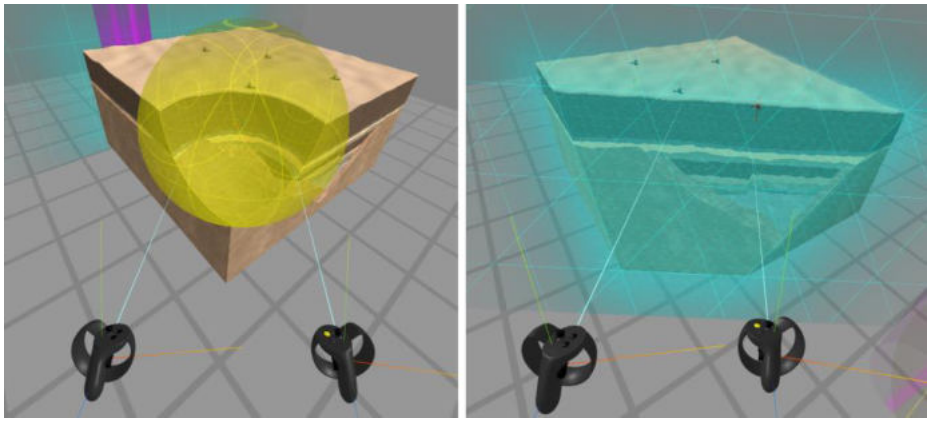


Fig. 2. Boolean operators being manipulated: sphere operator (left), flat operator (right).

The software lets the user navigate around a 3D volume that represents the basin. Using the hand controllers (as seen in figure 2), it is possible to position certain 3D objects called “Boolean operators” (sphere, plane or cylinder) that can interactively cut the volume, thus subtracting portions and allowing the user to visualize the interior of the basin.

ISBN 978 -987-633-574-4

2 The Rendering Engine

2.1 Representation of Surfaces

The problem of constructing a graphical representation of the basin can be divided into two parts, as shown in Figure 3: a) compute the coordinates of all the points that make up the surfaces resulting from subtracting the Boolean operators from the cube that represents the portion of terrain; b) compute the color of each pixel of the surfaces projected on the screen. To solve the first part, some way of modeling such surfaces dynamically must be considered, since the user can manipulate in real time the position, orientation and scale of Boolean operators.

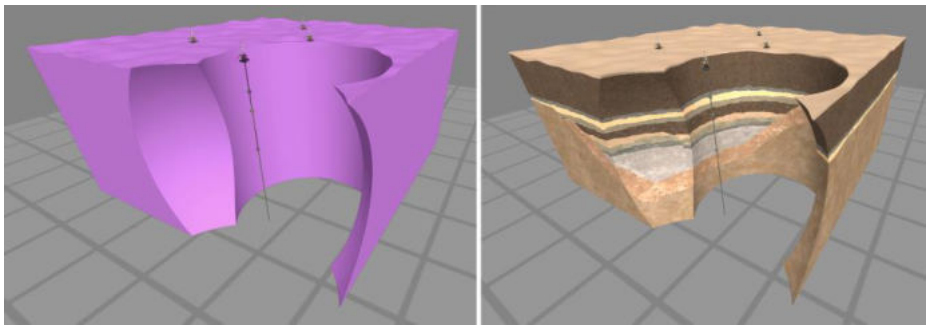


Fig. 3. Surface resulting from subtracting the spherical and cylindrical Boolean operators (left). Surface colored by a 3D procedural texture that exposes the internal structure of the oil reservoir. (right).

In computer graphics there are at least two ways of representing surfaces: the parametric form [7] and the implicit form [8]. The WebGL API allows the rendering of 3D scenes through a graphics pipeline that receives geometric information as input (vertices, triangles, normals, etc.) and generates an image on the output device.

In order to be able to visualize the surfaces, using the traditional graphical pipeline, it is necessary to obtain a mesh of triangles of the modeled surfaces.

The Marching Cubes technique [9] is an algorithm whose objective is to obtain a polygonal mesh of an iso-surface in a three-dimensional discrete scalar field (voxels). This technique is usually used for visualizing medical images taken by MRI scanners, where a density function is represented by values between 0 and 1.

In this case the density function could simply be computed from the Boolean operators and their position and orientation relative to the volume, assigning a density of 0 for an empty voxel and 1 for an occupied voxel.

Then, using Marching Cubes, it would be possible to get the triangle mesh that represents the surface shown in Figure 3.

One of the disadvantages of this approach is the processing time required to get the triangles, because this process should run every time any of the Boolean operators are modified. Although there are ways to implement it directly using a GPU, as

described in GPU GEMS 3[10], it requires the use of geometry shaders, and this functionality is not available in the WebGL 1.0 platform.

2.2 Rendering of Implicit Surfaces Using Ray Marching

In this project, the ray marching technique was used for representing the surfaces of the basin. It is not based on triangles, but on evaluating each pixel in an image, based on implicit representations of 3D surfaces [7] described by Signed Distance Functions (SDF). These functions are analytic expressions representing distance fields, which measure how close a point x is to a set S, and its sign changes as it is on one side or the other of the set [11].

The 3D surfaces to be represented can be defined using multiple distance functions of primitive geometric shapes [12] (spheres, planes, cubes, cylinders, etc.), combined with operations like union, intersection, subtraction, etc.

This poses a challenge when it is needed to represent some specific arbitrary surface. In the particular case of this application, the primitives involved are few and simple (cubes, spheres, etc.).

This technique offers great advantages in terms of performance, since all the logic and data necessary to solve the scene can be embedded completely within a pixel shader program written in GLSL language; minimizing memory reads.

This program is supplied to the GPU and executed very efficiently by multiple processing cores working in parallel; solving an entire view in less than 10 milliseconds, with the appropriate hardware.

2.3 The Ray Marching Algorithm

The essence of the algorithm can be summarized as follows: for each pixel on the screen, a ray is defined by an origin (position of the camera or the observer's eye) and a direction vector towards the pixel in the near plane of the camera. Then, through an iterative process, the algorithm advances in that direction, evaluating at each step if the ray impacted any surface.

There are different variants of this algorithm; some of them, like the one proposed in [11], advance at a constant pace and therefore require a very high sampling rate on the path of the ray, which impacts on performance.

On the other hand, Hart [13] introduces a more robust method known as Sphere Tracing, where the advanced distance at each step is not constant, but depends on the evaluation of an SDF. If the resulting value d (distance to the surface) is positive, the position over the ray should be increased by d. If d is negative, the resulting position is within the implicit surface defined by the SDF.

The following code shows the essence of the algorithm, where *distanceToScene* is the actual SDF.

ISBN 978 -987-633-574-4

Ray marching function in GLSL

```
float raymarching(vec3 rayOrigin, vec3 rayDir) {
  float t=0.0;
  for(int i=0; i<MAX_STEPS ; i++) {
     float delta = distanceToScene(rayOrigin+rayDir*t);
     if (delta<epsilon)  return t;
     t+= delta;
  }
  return INFINITY; // max steps where reached
}
```

When the condition $d < epsilon$ is met, the ray is considered to have hit the surface. Then the color of the pixel is evaluated at the point of impact using Phong's model [14]. The normal vector at the pixel can be obtained by calculating the gradient vector of the distance field, using 4 sampled points close to the impact point.

2.4 Procedural 3D Texture

The Phong model calculates the ambient (*ka*) and diffuse (*kd*) coefficients, based on the ambient light and a directional light representing the sun. Then, the diffuse color of the surface is the most relevant piece of information for the user, as it will help him recognize the different basin strata. The *getTerrainColor* function returns an RGB color based on two input parameters: *position* (vec3) and *normal* (vec3). A *colorMix* uniform variable controls whether the output should be a flat color or a pixel sampled from texture maps, as shown in figure 4.

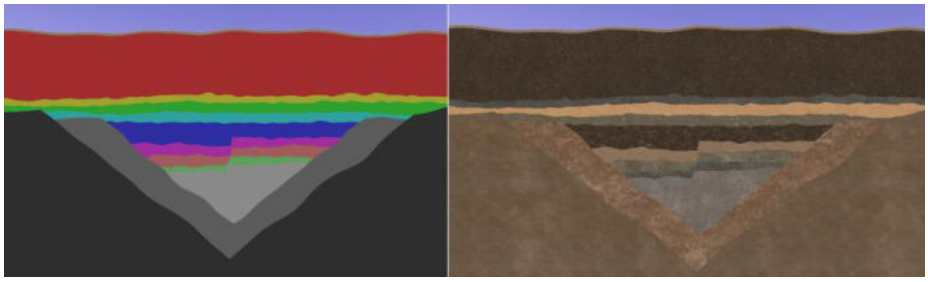


Fig. 4. Flat color output, with *colormix*=0 (left) texture color output with *colormix*=1 (right)

In this application there were 11 predefined types of basin strata, with their corresponding color and seamless texture map. For each component, a weight variable defines how much it contributes to the final pixel color.

In this model of an oil reservoir, most of the strata are horizontal layers with certain thickness. To compute the weight value, two *smoothstep* (ascending and descending slopes) functions are multiplied to generate a positive weight within a certain depth range. In order to get a more realistic appearance and to distort the straight-line borders between layers, a gradient noise [15] is applied on the *position* parameter, which represents the coordinates inside the basin.

ISBN 978 -987-633-574-4

2.5 Texture Mapping

In order to apply texture mapping to the raymarched surfaces, texture coordinates *(u,v)* need to be defined for every pixel. Considering that the surfaces are generated in real-time based on the result of the Boolean operators intersecting the volume, the texture coordinates also need to be generated dynamically. The triplanar mapping technique was chosen. This algorithm described by Geiss in [10], assigns *(u,v)* based on the one of the 3 cartesian planes XY, XZ or YZ whose normal is closer to the normal of the pixel, as shown in Figure 5. The *getTriplanar* GLSL function returns the texture map pixel based on the *position* & *normal* values at the pixel.

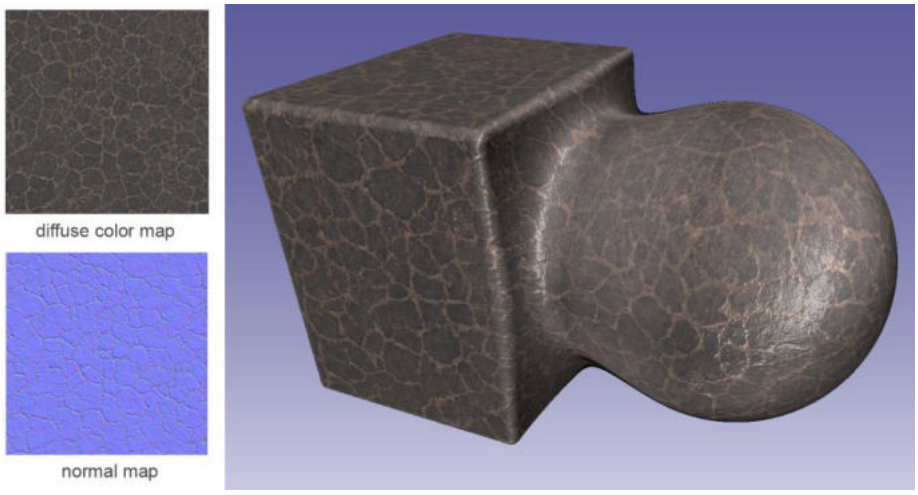


Fig. 5. The triplanar mapping algorithm applied to a dynamically generated surface.

2.6 Interactivity and User Experience

Once the VR mode is initiated, the user is immersed in a virtual scene consisting of an infinite extension floor and a colored cube of 1 meter on the side that floats at the height of the operator's line of sight. The user can walk around or can remain static and move to a different location in the virtual space, using the right controller thumbstick. By pointing his hand in the desired direction, the stick controls the speed of the motion.

ISBN 978 -987-633-574-4

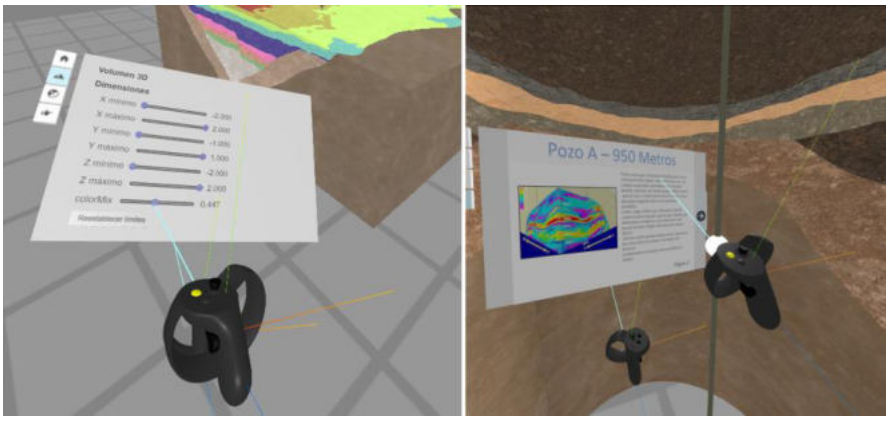


Fig. 6. Virtual control panel (left) drilling rigs “hyperlinks” (right).

The left-hand controller allows the user to show and hide a virtual control panel consisting of multiple tabs. Using the right controller, the user can point to and interact with different controls such as buttons and sliders as shown in Figure 6.

In addition, the hand controllers have side buttons associated with the grab and drop gesture, which allow the user to grab, rotate, move and drop the 3 types of Boolean operators, to interactively crop the 3D volume.

The 3D scene also features pre-designed 3D models that represent the drilling rigs and the oil well path. The user can point and select different targets along the well acting as "hyperlinks" that display related information on the panel (Figure 6 right). Finally, the user can take snapshots from his current point of view, which are automatically downloaded in JPG format, allowing a later analysis, outside of the virtual reality environment.

2.7 Virtual Control Panel

The application parameters can be controlled simultaneously from a 2D menu (on the browser window) or the virtual control panel in VR. The *MenuManager* component works as a centralized database for current application parameters. The 2D menu and the virtual control panel interact with it though getters & setters and *onValueChanged* events trigger when a parameter is modified.

The control panel is built dynamically from a JSON configuration file that defines the content of each tab, consisting of a list of controls (button, title, slider, switch or select). Each control mimics the behavior of their corresponding HTML counterparts but in 3D.

Three.js provides a class called *Raycaster* that can test the intersection of a ray (point + direction vectors) in the 3D scene, with a list of mesh objects. In some cases, invisible bounding box meshes are used, instead of the real mesh, to test the surface-ray collision (for example in the case of the sphere representing a slider’s cursor).

Every time the *Raycaster* detects a hit, it provides information on the object that was hit and the coordinates in world and local object space. The hand controller’s

location and orientation are provided by Three.js, through *VRController* class, which also provides event handlers for all the buttons and thumbsticks events.

2.8 Final Image Composition

The rendering function has two modes: desktop and VR. In VR mode, the viewport size is set to the dimensions of the VR device, provided by the WebXR API. In this mode, the viewport is divided in two halves (left & right eye view) and the rendering has to be executed twice on each frame using the corresponding view and projection matrices for each eye.

In order to generate a view from a camera, the rendering function requires two render passes. In the first pass, the ray marching algorithm is executed by a pixel shader that is assigned to a quad geometry matching the size of the viewport. The resulting image includes the basin volume, the floor and the background. This shader requires not only to output color information, but also depth for each pixel. This is important for compositing the next pass. In WebGL 1.0 it is not possible to output depth information from a fragment shader by default, so EXT_frag_depth needs to be enabled and available in the browser.

In the second pass all the objects that are represented by triangle meshes (Boolean operators, hand controllers, virtual control panel, oil rigs, etc.) are rendered. All of them are opaque with the exception of the Boolean operators that are translucent. As all transparent meshes are drawn at the end of the second pass, they blend properly with the previous content of the framebuffer.

3 Conclusions

The application developed allows the user to intuitively manipulate and explore a 3D volume, without the need to learn complex user interactions commonly used in 3D CAD software (using keyboard & mouse) thus reducing the learning curve for casual users.

The Boolean operators ease the study of the borders between strata and how they propagate inside a sedimentary basin. Moreover, the depth perception provided by the stereoscopic view, helps the user to properly interpret the shapes and scales.

The combination of ray marching and scanline rendering techniques, and the implementation of a procedural 3D texture, allowed the efficient implementation of a real-time application with low latency and a good performance, without the uncomfortable effects usually associated with virtual reality experiences.

The WebXR API, while still needs improvements, seems to be a promising alternative to build and deploy multi-platform VR experiences to wide audiences, at lower costs and reduced delivery times.

References

1. Jampeisov, Z.: Using Virtual Reality Technology in Oil and Gas Industry. 9, 124–127 (2019). https://doi.org/10.31033/ijemr.9.2.15.
2. Aveleyra, E.E.: Aportes para el debate: Las tecnologías en la enseñanza universitaria: nuevos escenarios, nuevos desafíos. En C. Nosiglia (comp.). La Universidad de Buenos Aires. Aportes para la CRES, pp.177-189. Editorial Universitaria de Buenos Aires, Buenos Aires (2018).
3. Cabero Almenara, J., Barroso Osuna, J.: Nuevos retos en tecnología educativa. Síntesis, Madrid (2015).
4. Coll, C., Monereo, C.: Psicología de la educación virtual. Aprender y enseñar con las tecnologías de la información y comunicación. Morata, Madrid (2008).
5. Martí, R., Gisbert, M., Larraz, V.: Ecosistemas tecnológicos de aprendizaje y gestión educativa. Características estratégicas para un diseño eficiente. Edutec. Revista Electrónica de Tecnología Educativa, (64), pp. 1-17 (2018).
6. WebXR Standard, https://www.w3.org/TR/webxr.
7. Hughes, J. F., van Dam, A., McGuire, M., Sklar, D. F., Foley, J. D., Feiner, S. K., Akeley, K.: Computer Graphics - Principles and Practice, 3rd Edition. Addison-Wesley (2014).
8. Menon, J.: An Introduction to Implicit Techniques. ACM SIGGRAPH 96 Course Notes: Implicit Surfaces for Geometric Modeling and Computer Graphics, pages A1–13 (1996).
9. Newman, T. S., Yi, H.: A survey of the marching cubes algorithm. 30, 854–879 (2006). https://doi.org/10.1016/j.cag.2006.07.021.
10. Nguyen, H.: GPU gems. Addison-Wesley, Upper Saddle River, NJ [u.a.] (2007).
11. Rendering Worlds with Two Triangles with raytracing on the GPU in 4096 bytes. https://www.iquilezles.org/www/material/nvscene2008/rwwtt.pdf.
12. Modeling with distance functions, https://iquilezles.org/www/articles/distfunctions/distfunctions.htm.
13. Hart, J., C.: Sphere Tracing: A Geometric Method for the Antialiased Ray Tracing of Implicit Surfaces. The Visual Computer, 12(10):527–545, (1996).
14. Phong, B.: Illumination of Computer-Generated Images, Department of Computer Science, University of Utah, UTEC-CSs-73-129 (1973).
15. Ebert, D., Musgrave, K., Peachey, D., Perlin, K., Worley. Texturing and Modeling: A Procedural Approach. Academic Press, October 1994. ISBN 0-12-228760-6

ISBN 978 -987-633-574-4

PReViMuGA: Prototipo para un Recorrido Virtual del Museo Gregorio Álvarez

Sanchez Viviana, Larrosa Norberto, Fracchia Carina, Amaro Silvia

Universidad Nacional del Comahue, Buenos Aires 1400, Argentina
{viviana.sanchez, norberto.larrosa, carina.fracchia,
silvia.amaro}@fi.uncoma.edu.ar

Resumen Para un museo contar con una aplicación que permita ofrecer experiencias de recorridos virtuales sobre sus exposiciones, constituye una herramienta importante para la didáctica del patrimonio, facilitando la observación y la experimentación. Se permite a los interesados en una muestra recorrerla de manera virtual, sin importar su ubicación geográfica, y principalmente garantizando el resguardo del patrimonio cultural. El límite se encuentra en disponer de una conexión de internet y un dispositivo que permita el despliegue de la aplicación.
La Realidad Virtual hace posible la implementación de este tipo de aplicaciones y en general sin requerir grandes costos.
En este trabajo se presenta la implementación de un prototipo web, que permite concretar una experiencia de un recorrido virtual e interactivo de una de las salas de exposición del museo Gregorio Álvarez, ubicado en la ciudad de Neuquén, Argentina. Se describen consideraciones de diseño e implementación, conclusiones y líneas de acción futuras.

Palabras claves: Realidad Virtual · Museos · TIC · recorridos virtuales.

1. Introducción

Los museos son instituciones muy importantes dentro de una sociedad dado que son los encargados de la investigación y conservación del patrimonio, además de otras funciones tales como la educación, difusión y exposición. Un museo se define como una " *institución sin fines lucrativos, permanente, al servicio de la sociedad y de su desarrollo, abierta al público, que adquiere, conserva, investiga, comunica y expone el patrimonio material e inmaterial de la humanidad y su medio ambiente con fines de educación, estudio y recreo* "[5]. Una transmisión eficaz del patrimonio permitiría acercar a un público altamente numeroso la variedad de matices que el patrimonio arqueológico posee, y de esta manera fomentar la revalorización de su significado y preservación.

La propia comprensión de la historia, sumado a la dificultad que conlleva la interpretación del patrimonio debido al alto nivel de abstracción que este ejercicio requiere, aparecen como las dificultades cognitivas más señaladas [11]. Desarrollos que datan de más de dos décadas atrás muestran las grandes posibilidades que ofrecen las TIC (Tecnologías de la Información y la Comunicación)

ISBN 978 -987-633-574-4

para abrir los espacios físicos ofrecidos en los museos al mundo entero, dando la posibilidad de acceder e interactuar con los elementos (digitalizados) allí exhibidos. Sumado a esto, el avance en el surgimiento de dispositivos periféricos tales como los Google Cardboard han abierto la posibilidad a otras tecnologías, como es el caso de la RV, ofreciendo experiencias inmersivas a un costo accesible.

En relación a nuestro contexto, se está trabajando en una propuesta de recorridos virtuales para favorecer la difusión de la muestra permanente del museo Gregorio Álvarez, localizado en la ciudad de Neuquén, Argentina. Estos desarrollos se enmarcan en los proyectos de investigación 04/F010 "Visualización de Datos y Realidad Virtual" y 04/F016 "Computación Aplicada a las Ciencias y Educación", y el proyecto de extensión "El museo va a las escuelas", todos pertenecientes a la Facultad de Informática de la Universidad Nacional del Comahue.

En el documento se presenta un prototipo web denominado PReViMuGA (Prototipo para un Recorrido Virtual del Museo Gregorio Álvarez), además de las consideraciones de diseño, herramientas utilizadas en su implementación, algunas conclusiones y líneas futuras.

1.1. Realidad Virtual

La realidad virtual (RV) se basa en computación gráfica, tecnología de simulación y tecnología de multimedia; para simular funciones visuales, auditivas, táctiles y otras funciones sensoriales humanas. A través de esta tecnología las personas pueden sumergirse en un entorno virtual con la información generada por una computadora y experimentar algún tipo de interacción. El recorrido virtual es una rama de aplicaciones de RV y consiste en la creación de información del entorno virtual en un nuevo espacio multidimensional, basándose en datos reales mediante el uso de la tecnología de RV. Con el entorno creado se puede realizar el recorrido a larga distancia del mundo real, con las características de las 3I: **i**nteracción, **i**nmersión e **i**maginación [1].

Álvaro Ulldemolins denomina recorrido virtual a "*una simulación de un lugar virtual compuesto por una secuencia de imágenes*". Considerando esta definición para poder implementar un recorrido, ya sea fijo o interactivo, se deben conocer las bases de cómo presentar el conjunto de imágenes que se quieren visualizar. En un recorrido fijo el usuario no puede interactuar con el entorno, por lo que es el montaje del recorrido lo que determina lo que el usuario verá. Por lo tanto, es importante conocer los tipos de planos que podemos utilizar a la hora de realizar un montaje de un recorrido virtual. En cambio, en un recorrido interactivo el usuario puede ver cualquier zona que desee e interactuar con los elementos del escenario, alejarse, acercarse, caminar o volar por la escena, de manera que el recorrido lo realiza según sus necesidades [10].

PReViMuGA recrea un recorrido virtual interactivo no inmersivo, característica que se describe a continuación:

Recorridos Virtuales no inmersivos Estos recorridos pueden crearse utilizando tableros temáticos basados en fotografías, información y enlaces a páginas en la red. También en algunos casos se pueden trabajar con recorridos que

ISBN 978 -987-633-574-4

permitan vistas en 360º, pero muchas veces con interactividad limitada. Como desventajas se pueden señalar la desvinculación entre lo que percibe el usuario y las características reales de los objetos. Si se logra combinar la percepción 3D de los edificios por medio de fotografías esféricas y el recorrido en 360º (ej. Street View) con la interactividad en tiempo real con las obras, además de ofrecer retroalimentación en tiempo real, se podría emular la visita real[6]. Otro aspecto a considerar es el público heterogéneo por lo que deberían pensarse diferentes propuestas.

1.2. Evolución a los museos virtuales

En Argentina se encuentran dos tipos de espacios científico/culturales: por un lado los *museos tradicionales* que, si bien han incorporado nuevas propuestas y programas, son predominantemente expositivos; y por otro lado *los centros de ciencias interactivos*. En el surgimiento de los Museos interactivos de Ciencia y Tecnología en América Latina, se identifican dos períodos. En el primero, entre la década de 1980 y finales del siglo XX, las propuestas de las instituciones pioneras de cada país tendían a ser adaptaciones de los modelos extranjeros. Estos modelos consistían de exhibiciones centradas en temas clásicos y universales de la ciencia, que permitían su utilización en demostraciones atractivas y de bajo costo. El segundo período, que se inicia a partir del año 2000, se caracteriza por programas y propuestas enfocadas en temas globales desde una perspectiva anclada en las realidades, necesidades e intereses de la población de la región. Esta característica sirvió para que los centros de ciencia creados en este período o aquellos que actualizaron sus exhibiciones en este sentido, adquirieran una personalidad local propia [2].

En el último año la Pandemia Covid-19, tuvo un gran impacto sobre diferentes ámbitos de la sociedad y los museos no fueron la excepción. Ante esta situación y dadas las disposiciones ASPO y DISPO (contempladas en el Decreto 641/2020 dispuestas por el Gobierno Nacional), los museos tuvieron que pensar cómo adaptar los servicios que ofrecían de manera presencial a la virtualidad. Esto se vio reflejado en varias de las publicaciones de las Jornadas Nacionales de Museos Universitarios, donde se describe el impacto que ha tenido la virtualidad en esta nueva realidad. Por ejemplo los autores José Lombardo, Ayelén Lenzi, Oscar Duarte, Elba Boggiano y Alberto Capparelli en su publicación *Comunicando, difundiendo y divulgando las ciencias químicas desde un museo virtual*, concluyen que la transformación en un museo virtual, surgida de manera abrupta, ha logrado una respuesta favorable de la comunidad virtual. Entienden que la virtualidad permanecerá aun cuando vuelvan tiempos de normalidad y acompañará al Museo presencial. Las propuestas virtuales se deben mejorar, incorporar nuevas tecnologías para comunicar, difundir y divulgar con mayor eficiencia las ciencias químicas, que permitan un museo más accesible [4].

Es claro que en este contexto las herramientas TIC son fundamentales para implementar un museo que pueda ser visitado y recorrido de manera no presencial. Además brindar la posibilidad a los interesados en una muestra, de poder interactuar con las piezas e informarse de las mismas, es un posibilidad que la

ISBN 978 -987-633-574-4

RV/RA (Realidad Aumentada) hacen posible. El uso de RV permite integrar objetos del mundo real con animaciones e información adicional, favoreciendo su difusión, y preservando el patrimonio cultural tangible e intangible [12]

2. Diseño del prototipo

Son varias las tareas involucradas en el proceso de desarrollo de un prototipo incremental que permita recrear de manera virtual, la experiencia de un visitante a la sala de un museo. La finalidad del prototipo es que sea el punto de partida para el desarrollo e implementación de una aplicación de RV que permita recrear las salas de exposición, visualizar las piezas, interactuar con ellas y brindar la posibilidad de recorrer las muestras en el momento que el visitante lo desee.

En la Figura 1 se visualiza el flujo de trabajo y cada una de las tareas involucradas. El cuadro celeste referenciado por *(1)Elementos del escenario*, agrupa a todas las tareas que están vinculadas con las piezas que se exponen en el museo y los contenidos de multimedia que se desean desplegar. El cuadro naranja referenciado por *(2)Modelado 3D*, agrupa a todas las tareas que están vinculadas con el modelado en 3D de las piezas que se exponen en el museo. Finalmente el cuadro verde referenciado por *(3)Recorrido Virtual*, agrupa a todas las actividades realizadas dentro del entorno de desarrollo para la implementación del recorrido virtual.

A continuación se describen las tareas comprendidas en cada etapa.

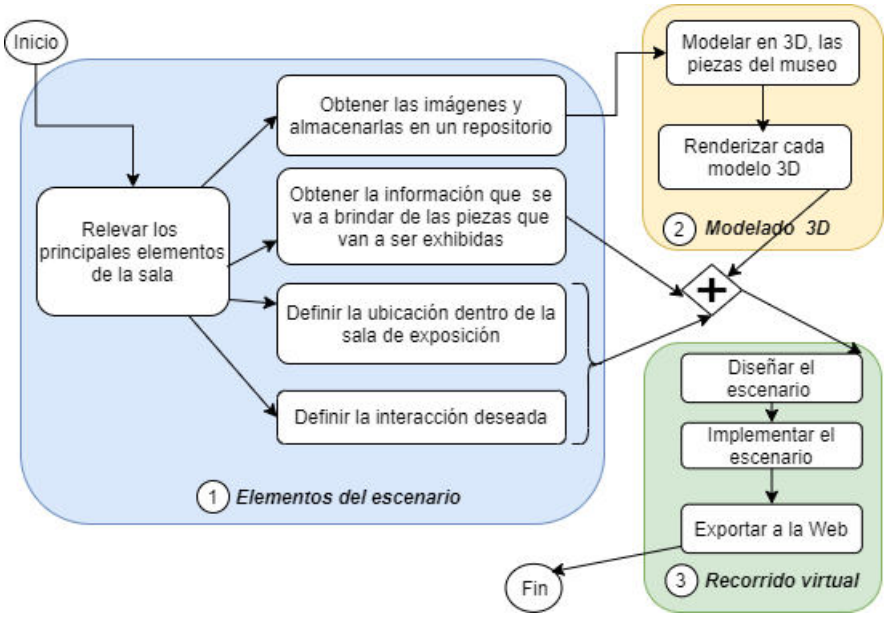


Figura 1. Tareas realizadas para el desarrollo e implementación de PReViMuGA

ISBN 978 -987-633-574-4

2.1. Elementos del escenario

El objetivo de esta etapa es relevar los elementos que se desean recrear y que van a formar parte del escenario virtual. Para ello se requiere realizar reuniones con el personal del museo, que son quienes conocen esta información.

Al pensar la sala de exposición del museo como un escenario virtual, además de las piezas que conforman la muestra, se deben identificar todos los elementos que forman parte de la sala. A continuación se describen los elementos identificados que deben ser relevados digitalmente, junto a las actividades realizadas con ellos:

- ***Las piezas:*** se deben capturar las imágenes correspondientes a las piezas que forman parte de la muestra y definir la interacción que pueda realizar el visitante con la pieza. Por ejemplo: desplegar algún texto, audio y/o video cuando el visitante selecciona la pieza. Todos los archivos de multimedia deben ser relevados y quedar accesibles al proyecto. Por último, se deben marcar aquellas piezas que se desean como parte del prototipo inicial.
- ***Los exhibidores:*** se deben capturar las imágenes correspondientes a los elementos donde las piezas son colocadas para que puedan ser exhibidas y observadas. Entre ellas encontramos los paneles, las mesas y las vitrinas móviles.
- ***Los elementos arquitectónicos:*** se deben capturar las imágenes correspondientes a los elementos que forman parte de la arquitectura de la sala de exposición como: las ventanas, las puertas, los vitrales, las columnas, el piso y el techo.
- ***El guía:*** se deben definir las interacciones que puede tener el visitante con este rol. Inicialmente se busca que sea el encargado de dar la bienvenida a la sala, describir la muestra, indicar como realizar el recorrido y brindar información cuando el visitante lo desee.

Además una actividad común a todos los elementos mencionados es la definición de la posición dentro del plano, información que es fundamental para el diseño del escenario virtual. Una vez que se definen e identifican los elementos que van a formar parte de la muestra, se deben coordinar y realizar varias visitas al museo para ejecutar cada una de las actividades descriptas.

Como resultado de la realización de esta etapa, se obtuvieron todos los archivos de multimedia (como textos, audios, imágenes, animaciones y videos) que se desplegarán en el escenario virtual. Finalmente, se define un repositorio que permite alojar todos los elementos relevados, de manera tal que puedan ser gestionados y resguardados de manera segura.

2.2. Modelado 3D

El modelado 3D está relacionado con aquellas tareas que se realizan con el propósito de manipular imágenes con un determinado software, que a través de ciertos cálculos matemáticos permite obtener una proyección visual.

ISBN 978 -987-633-574-4

El objetivo de esta etapa es el modelado de los objetos 3D correspondientes a los elementos que fueron relevados para la implementación del prototipo en la etapa anterior. El escenario principal del recorrido es la sala de exposición, que debe ser recreada a partir del plano del museo aplicando las texturas y los elementos arquitecturales.

Como resultado de la realización de esta etapa, se obtienen los modelos 3D correspondientes a: la sala de exposición, las piezas que se desean exhibir, los exhibidores y el guía del museo. Todos los modelos 3D serán almacenados en el repositorio del proyecto.

2.3. Recorrido virtual

Un recorrido virtual se destaca por la simulación de un espacio real, el cual cuenta con un conjunto de objetos 3D que pueden ser visualizados, manipulados y controlados por el usuario. Los recorridos infovirtuales son recorridos interactivos donde es posible visualizar desde cualquier zona, interactuar, controlar, alejar, y/o acercar a los diferentes elementos que componen el espacio o escenarios 3D. Es decir permiten realizar diferentes acciones en los planos o entornos específicos, que llevan al usuario a mayor interés, confianza, gusto y comodidad del recorrido. Este tipo de recorridos maximizan el propósito de un museo, brindando la posibilidad de que los interesados en recorrer una muestra lo puedan hacer en el momento que lo deseen [3].

A partir de los objetos 3D de todos los elementos que van a conformar el escenario virtual de la muestra, obtenidos en la etapa anterior, en esta etapa se diseña el escenario y se establece la organización de cada uno de los elementos dentro de la sala de exhibición. Luego de diseñado el escenario virtual, donde cada una de las piezas se encuentran ubicadas en su elemento de exhibición y este último ubicado en algún lugar estratégico de la sala, se deben definir los componentes que permiten al visitante: interactuar con las piezas de la muestra, desplazarse por toda la sala de exposición del museo y visualizar las sombras producidas por el efecto de la luz sobre cada uno de los elementos.

Como resultado de la realización de esta etapa, se obtiene el diseño del escenario de la sala de exposición del museo y la interacción que se espera ofrecer a los visitantes de la misma.

2.4. Publicación en la Web

Internet es un medio muy utilizado en la actualidad para dar a conocer un producto o un proyecto, esto se debe a que brinda la posibilidad de un acceso universal a la información.

Las aplicaciones web son cada vez más populares en todo ámbito, debido a sus características que las distinguen de otro tipo de aplicaciones [9]. Para ejecutarlas se requiere un navegador web y una conexión a internet.

Teniendo en cuenta las características y posibilidades que brinda internet, se pretende que el recorrido virtual pueda ser accedido desde el sitio web del museo.

ISBN 978 -987-633-574-4

Por ello su desarrollo debe garantizar su ejecución desde los navegadores de internet, además de contemplar a los dispositivos más utilizados en la actualidad.

Son varios los beneficios de poder contar con internet, como un medio de acceso a la información brindada por un museo. Es una herramienta que permite llevar el museo a los visitantes, que no necesariamente se encuentren en la zona, para que puedan realizar el recorrido de la muestra en el momento deseado. Fundamentalmente y pensando el museo como un entorno de educación no formal, esto facilitaría el acceso de las instituciones educativas que a veces, por motivos económicos o distancias, no logran realizar una visita de manera presencial.

3. Herramientas que se utilizaron en la Implementación

Como se mencionó anteriormente, que un museo pueda contar con aplicaciones que permitan realizar experiencias de recorridos virtuales sobre sus exposiciones, constituye una herramienta muy importante para la didáctica del patrimonio, facilitando la observación y experimentación.

La implementación de aplicaciones de recorridos virtuales se hace posible gracias a la tecnología brindada por la RV, con algunos límites tales como la disponibilidad y calidad de la conexión a internet, y contar con un dispositivo que posea las características básicas necesarias para permitir el despliegue de la aplicación.

El propósito de esta sección es describir las herramientas que se utilizaron para la implementación de cada una de las tareas descriptas en la sección 2 .

3.1. Modelado 3D

Para la implementación de las tareas de creación, edición y manipulación de los objetos 3D correspondientes a los elementos del museo, se seleccionó como herramienta de modelado a Blender[1]. Esta suite permite implementar animaciones 2D y modelado 3D y cuenta con la ventaja de ser multiplataforma, gratuita y de código abierto. Si bien es una herramienta compleja y la curva de aprendizaje es alta, se encuentra mucho material disponible para un aprendizaje correcto [7].

3.2. Recorrido virtual

Teniendo a disposición los objetos 3D correspondientes a los elementos del museo y a las piezas que se desean exhibir, se debe contar con una herramienta que permita la implementación de un recorrido virtual. El recorrido se debe desplegar sobre el escenario de la sala de exposición y brindar al visitante la posibilidad de interactuar con las piezas y acceder a la información que se desprende de cada una de ellas.

Como herramienta para la implementación del recorrido se utilizó la plataforma Unity[2] que permite la creación de contenido interactivo y en tiempo

[1] https://www.blender.org/
[2] https://unity.com/

ISBN 978 -987-633-574-4

real. Unity provee facilidades para implementar juegos y aplicaciones en 2D, 3D y RV a gran velocidad, es un entorno de desarrollo de juegos que posee un poderoso motor de renderizado totalmente integrado y brinda un conjunto de herramientas intuitivas que permiten crear contenido 3D interactivo. Además en su tienda ofrece un importante número de activos de calidad listos para usar y una comunidad donde se intercambia conocimiento.

Se utilizaron 2 paquetes de código abierto, obtenidos de la tienda oficial de Unity: *First Person All-in-One*[3] que permite el desplazamiento del visitante dentro del escenario y *Fungus*[4] que permite la interacción del visitante con las piezas y el guía.

Los elementos sobre los que se van a exhibir las muestras en la escena de la sala de exposición (como los paneles móviles, las vitrinas y las mesas), se implementan como Objetos Prefabs de Unity. Estos objetos se pueden crear, configurar y almacenar con todos sus componentes, valores de propiedad y elementos secundarios. Su característica más importante es que son objetos reutilizables, de manera tal que cualquier cambio que se implementa en este tipo de objetos, impacta sobre todas sus instancias. Es decir que actúa como una plantilla a partir de la cual se pueden crear nuevas instancias para una escena. Esto es muy importante a la hora de implementar la sala de exposición, ya que en la misma se encuentran varios de estos elementos.

En la Figura 2 se visualiza una captura del entorno de Unity, de una escena del recorrido virtual de PReViMuGA. Se pueden visualizar todos los elementos identificados y descriptos en 2.1. En particular la vitrina que contiene en su interior piezas de la muestra, el panel móvil con cuadros de la exposición y el guía desplegando un mensaje de bienvenida.

3.3. Publicación en la Web

WebGL (Web Graphics Library) es una tecnología que hace posible la visualización de gráficos 2D y 3D en el navegador, sin necesidad de instalar complementos adicionales. Deriva de la biblioteca OpenGL ES 2.0, la cual es un API de bajo nivel que incluye a los dispositivos móviles. Es un API de javascript que se puede utilizar con HTML5, está escrito dentro de la etiqueta <canvas> y permite a los navegadores de internet el acceso a unidades de procesamiento gráficos (GPU) [8].

Dado que el entorno de desarrollo Unity permite exportar los proyectos a la tecnología WebGL, se puede obtener la implementación del prototipo como una aplicación JavaScript que utiliza tecnologías HTML5 y la API de renderización WebGL. Una vez realizada la exportación del proyecto y obtenidos los archivos de la aplicación generados por Unity, se publican en un servidor web para que pueda ser accedido por el personal del museo y el equipo de desarrollo.

[3] https://assetstore.unity.com/packages/tools/input-management/first-person-all-in-one-135316

[4] https://assetstore.unity.com/packages/tools/game-toolkits/fungus-34184

ISBN 978 -987-633-574-4

Figura 2. Captura de una escena de PReViMuGA desde el entorno de desarrollo Unity

4. Conclusiones

La implementación de un recorrido virtual del museo permite recorrer la sala donde se exhiben las muestras, ver fotos y brindar la posibilidad al visitante de acceder al conocimiento ofrecido. Un recorrido virtual y accesible desde la web, permite además que los interesados puedan determinar el lugar y el momento más óptimo para su experiencia virtual. El interés que despierta este tipo de recorrido, puede ser explotado por el museo, garantizando el resguardo del patrimonio y poniendo a disposición el conocimiento de la muestra al mundo.

En la actualidad no es impensado contar con un dispositivo desde el cual se tenga acceso a internet. El incremento de ordenadores personales y la evolución de los dispositivos móviles, convierten al entorno web en un espacio privilegiado para acceder a la aplicación.

En este contexto, no se requiere ninguna instalación adicional, el único requerimiento es contar con un navegador web.

Las etapas descriptas para el desarrollo y las herramientas utilizadas para la implementación del prototipo, se pueden utilizar para implementar de manera incremental e interactiva, una aplicación de RV para una exposición completa del museo.

El campo de la RV es muy grande, así como la cantidad de herramientas y opciones que existen para que la experiencia sea lo más real posible. La utilización de las herramientas Blender y Unity cumplieron con todas las expectativas y los requerimientos definidos para la implementación de PReViMuGA. Tienen una gran comunidad de desarrolladores y esto se ve reflejado en la cantidad de información disponible en la web, además de la brindada en los dominios de las herramientas.

ISBN 978 -987-633-574-4

5. Trabajo Futuro

Se prevé utilizar el desarrollo del prototipo para la implementación del recorrido virtual de una muestra completa del museo Gregorio Álvarez. Cuando el recorrido se encuentre en producción, se proyecta realizar un estudio que permita evaluar su impacto en la sociedad.

Una vez que se abran las puertas del museo nuevamente a los visitantes, se proyecta implementar una experiencia de RV inmersiva. Se busca analizar la percepción y sensación de presencia que se ha conseguido con la aplicación, y aspectos relacionados a la usabilidad, adaptación y calidad del entorno.

El conocimiento y la experiencia adquirida se puede aplicar en la implementación de recorridos virtuales, ya sea para otros museos de la zona o de algún dominio donde la RV disminuya la brecha que pueda existir entre las personas y el conocimiento brindado por el lugar.

Referencias

1. Burdea, G., COIFFET, P.: Virtual reality technology, new yersey: John wiley&sons (2003)
2. Cambre, M.: Museos interactivos de ciencia y tecnología en américa latina. Red-POP: 25 años de popularización de la ciencia en América Latina p. 41 (2015)
3. Casas, L., Devesa, N., Ulldemolins, Á.: Animación 3D. UOC, Universitat Oberta de Catalunya (2011)
4. Franco, P.: Jornadas nacionales de museos universitarios: resúmenes
5. ICOM: Definición de museo, https://icom.museum/es/recursos/normas-y-directrices/definicion-del-museo/
6. Meraz, J.M.F., Domínguez, C.D.: Del museo sin muros, al museo como simulación fotográfica: experiencias contemporáneas en los museos en línea. Kepes **14**(16), 185–217 (2017)
7. Morelli, R.D., Ctenas, H.A.P., Nieva, L.S.: Modelado paramétrico 3d, render y animación con software libre: Interacción freecad+ blender. Geometrias & Graphica 2015 Proceedings pp. 023–036 (2015)
8. Parisi, T.: WebGL: up and running. .º'Reilly Media, Inc."(2012)
9. Ríos, J.R.M., Ordóñez, M.P.Z., Segarra, M.J.C., Zerda, F.G.G.: Comparación de metodologías en aplicaciones web. 3C Tecnología: glosas de innovación aplicadas a la pyme **7**(1), 1–19 (2018)
10. Ulldemolins, Á.: Recorridos virtuales. UOC (2010)
11. VIVANCOS, A.E., FERRER, L.A., GARCÍA, S.P.: ¿ hay vida más allá de la arqueología? la educación como una oportunidad. Revista Temporis [ação](ISSN 2317-5516) **17**(1), 20–42 (2017)
12. Xu, S.: Intangible cultural heritage development based on augmented reality technology. In: 2018 International Conference on Robots & Intelligent System (ICRIS). pp. 352–355. IEEE (2018)

ISBN 978 -987-633-574-4

Análisis y clasificación de ladrillos de hormigón celular a través de imágenes

Rodrigo Ortiz de Zarate, Lucas Rios, Gisela Roncaglia,
César Martínez, Enrique M. Albornoz

[1] Instituto de investigación en señales, sistemas e inteligencia computacional, sinc(i) UNL-CONICET, Ciudad Universitaria, Ruta Nacional Nº 168, km 472.4, (3000) Santa Fe.
Facultad de Ingenieria y Ciencias Hidricas – Universidad Nacional del Litoral
rodrigoodz@gmail.com, luucas125@outlook.com.ar, giselaroncaglia@gmail.com,
{cmartinez, emalbornoz}@sinc.unl.edu.ar

Resumen. Actualmente, el área de la construcción sostenible está enfocada en minimizar la utilización de recursos y promover el uso de técnicas constructivas innovadoras a partir de materiales que reduzcan la demanda energética, de recursos y con bajo impacto en el ambiente. El objetivo de este trabajo es generar una herramienta que permita determinar de manera sencilla y automática, si un ladrillo cumple con los requerimientos mínimos para ser un ladrillo apto para el ambiente de la construcción. La base de datos de imágenes ha sido relevada en una escena desarrollada para este trabajo. Se evaluaron diversas técnicas de procesamiento digital de imágenes para la extracción de características y se utilizó el método de KNN para clasificar las imágenes de los ladrillos. Finalmente, se compara y discute la efectividad de cada uno de estos métodos a partir de los resultados obtenidos que son prometedores.

Palabras claves: poros, hormigón, ladrillo, segmentación de imágenes, textura

1 Introducción

En la industria de la construcción, la producción de hormigón, desde la extracción hasta su transporte, genera casi un 10% del CO_2 emitido mundialmente. Esta industria, con sus procesos y productos involucrados, es una de las más nocivas para el medio ambiente [1]. En la actualidad, en Argentina, el sistema constructivo de ladrillo y cemento es el más difundido. Es por esto que surge como prioridad promover sistemas eficientes que fomenten un menor consumo de recursos, mayor rapidez en la ejecución de la obra, menor cantidad de residuos producidos, menor consumo de agua, entre otras características [2]. Hoy por hoy, los ladrillos de hormigón celular curado en autoclave (HCCA) son altamente usados en las construcciones, esto se debe a su sencillo y eficiente sistema constructivo [3], además de ser mucho más amigables con el medio ambiente. Estos se forman por una mezcla de aglomerantes (principalmente cemento y una proporción de cal), áridos finos (arena cuarcita finamente molida), agua y un agente expansor que genera millones de burbujas de aire en cada ladrillo a partir de una reacción química [4]. Esta estructura celular le otorga al HCCA muchas propiedades

ISBN 978 -987-633-574-4

que lo hacen muy eficiente, entre éstas se encuentran [3,4]: buena aislación térmica debido a la gran cantidad de “cámaras de aire” cerradas e incomunicadas que se encuentran en la masa; alta resistencia a la penetración de agua líquida, ya que la contextura cerrada tiene prácticamente nula succión capilar; mayor aislamiento acústico debido a la reducción de las ondas sonoras en el paso sucesivo a través de las “cámaras de aire”; gran resistencia al fuego; menor peso y mayor duración; entre otras. Entre el 60% y el 90% de la estructura de estos ladrillos se compone por poros y son los responsables de las propiedades mencionadas anteriormente. Por lo tanto, es fundamental que el proceso de fabricación sea correcto para lograr una adecuada presencia y distribución de poros, con el fin de obtener óptimos resultados.

Los poros presentes en el ladrillo pueden clasificarse como microporos y macroporos. Los microporos tienen tamaños que oscilan entre 100 nm y 0,1mm [5]. Para el análisis de estos se utilizan técnicas particulares propias de la mecánica como “porosimetría de intrusión de mercurio”, donde se inserta una sustancia como mercurio líquido y se obtiene a partir de esto la distribución de la porosidad en función del tamaño aparente de acceso a los poros [6]. Por otro lado, los macroporos tienen tamaños entre 0,5mm y 2mm [7], y pueden ser analizados con técnicas de procesamiento de imágenes.

En el presente trabajo, se abordan diferentes metodologías para la clasificación y distinción entre ladrillos aptos y no aptos, comenzando desde técnicas más básicas de segmentación a través de umbralización simple, hasta el análisis de propiedades estadísticas usando métodos de caracterización de texturas. Además, fue necesario generar una base de datos de imágenes para la que se configuró una escena particular y se detalla a continuación para poder replicar o extender la experimentación.

2 Metodología

En esta sección se detallarán cada uno de los pasos realizados para el desarrollo del modelo final, desde la creación de una base de datos de imágenes específica.

2.1 Base de datos de imágenes

La escena, que puede verse en las Figuras 1 y 2, se configuró posicionando una mesa apoyada contra la pared. Ambas fueron cubiertas con una tela de color negro, con el fin de lograr un contraste con el color blanco del ladrillo. Sobre la mesa se coloca cada ladrillo de tamaño 100x100mm con diferentes distribuciones de poros. Se decidió que la escena debía ser interior para poder ajustar la iluminación, y evitar que la luz solar influya en la escena. De esta manera, las fotos pueden ser registradas en distintos momentos del día y no se genera falta de homogeneidad en la iluminación, lo que dificultaría el tratamiento de las imágenes. Esto permite, además, reproducir la escena de forma precisa para realizar nuevas capturas y aumentar la base de datos con imágenes similares. En cuanto a la iluminación artificial de la escena, se utilizaron dos luces LED de 12 Watts 6500 K ubicadas a los lados del ladrillo, a 42 cm de distancia. Esta configuración fue el resultado de pruebas preliminares con las cuales se determinó

ISBN 978 -987-633-574-4

que es muy importante generar sombras en los poros, para facilitar su detección y así disminuir el post-procesamiento requerido al momento de utilizar las imágenes.

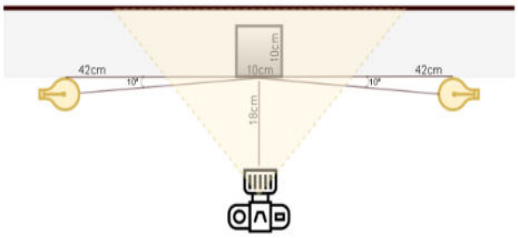


Fig. 1. Esquema de la escena con la que se capturaron las imágenes, indicando distancia, dirección y ángulo de inclinación de cada elemento.

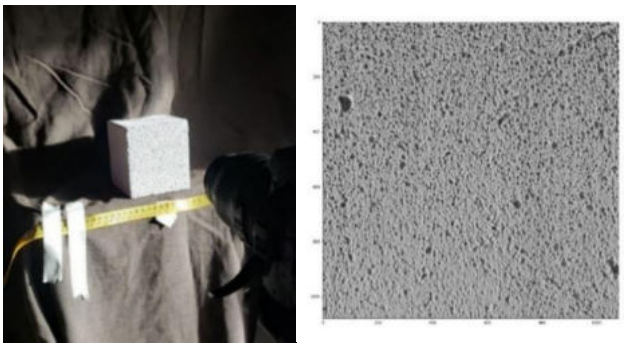


Fig. 2. Escena real (izquierda), región de interés (derecha).

Para la adquisición de imágenes se utilizó una cámara Nikon D3000 con un lente 18-55mm f/5,6 configurando un tiempo de exposición de 1/160 y cuyos resultados son archivos JPEG con una resolución de 3872x2592px. Finalmente, se obtuvieron 64 imágenes, 44 correspondientes a ladrillos aptos y 20 a ladrillos no aptos. A partir de este conjunto de imágenes y con el fin de focalizar el trabajo en el análisis de poros, se decidió generar otro conjunto de imágenes extrayendo una región de 1080x1080px centrada de cada ladrillo. Con este conjunto se realizarán los análisis que se detallan a continuación.

2.2 Procedimientos de segmentación

En un primer abordaje nos concentramos en encontrar y segmentar cada uno de los poros de la superficie, con el fin de obtener el área que estos ocupan. De esta manera podremos conseguir una estimación cercana de la proporción que está siendo cubierta por poros en la región de interés. En base a dicha proporción, determinaremos la clasificación final del ladrillo.

ISBN 978 -987-633-574-4

2.2.1 Método de umbralización adaptativa y segmentación

En este primer método, nos enfocamos en la búsqueda de cada poro en la imagen. El procedimiento consiste en los siguientes pasos (esquematizados en la Fig. 3):

- Umbralización adaptativa Gaussiana [8]: este método se basa en la binarización de una imagen mediante el cálculo de un valor de umbral en un área pequeña, obteniendo diferentes umbrales para diferentes áreas. Así, se obtiene una imagen representativa de los poros del ladrillo.
- Apertura morfológica: esta operación consiste en la aplicación de una erosión, seguida de una dilatación morfológica utilizando un mismo elemento estructurante [9]. Con esto se logran separar aquellos poros que aparecen unidos sin perder el tamaño original de los poros.
- Eliminación de elementos del borde: se realiza una limpieza de aquellos objetos que están sobre el borde de la imagen aplicando operaciones morfológicas con el fin de facilitar el conteo de poros.
- Búsqueda de componentes conectados: este método se utiliza para localizar, en una imagen binaria, cada elemento desconectado del resto (en este caso, los poros) [9]. Esto permite etiquetar cada uno como un objeto individual y brinda características útiles para la clasificación.

Finalizados estos pasos, se dispone de cada poro y su área representada en píxeles, la cual usaremos para determinar la densidad de poros en la superficie. Se calcula el área total de la zona que no se considera poro, es decir, la diferencia entre el área total de la imagen y el área sumada de todos los poros. De esta manera, se tiene un único valor de área por cada imagen, el cual se utilizará en el proceso de clasificación.

Fig. 3. Secuencia de procesamiento del método basado en umbralización adaptativa.

2.2.2 Método de umbralización de entropía y segmentación

Este segundo método no utiliza la imagen tal como fue capturada, sino que se calcula una imagen de entropía con el fin de cuantificar el 'desorden' presente en la imagen del ladrillo y es ésta imagen la que se procesa con los pasos descritos en 2.2.1 (ver Fig. 4). Para esto, se genera una imagen de entropía utilizando como kernel un disco de 3px y la siguiente fórmula [10], donde $P(x_i)$es la probabilidad de ocurrencia:

$$H(x) = -\sum_{i=1}^{n} P(x_i) log P(x_i) \quad \textbf{(1)}$$

ISBN 978 -987-633-574-4

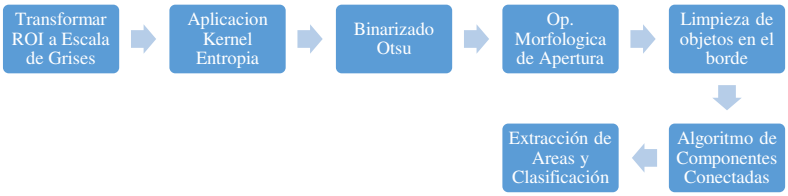


Fig. 4. Secuencia de procesamiento del método basado en entropía.

2.3 Procedimientos de segmentación basados en texturas

Debido a la forma irregular que presentan los poros en el ladrillo, se presentan casos donde segmentarlos se vuelve muy difícil, más aún si no se dispone de un sistema de iluminación que recree la escena utilizada en este trabajo. Esta es una primera apreciación con base en las pruebas iniciales, en trabajos futuros se evaluará la robustez de los métodos para diferentes configuraciones de escenas. Es esperable que los métodos basados en texturas tomen relevancia en contextos más difíciles. La textura de una superficie hace referencia a la distribución de valores de intensidad a nivel espacial, y a partir de estas se puede obtener fácilmente propiedades como: fineza, rugosidad, suavidad, etc. En base a una comparativa de estos valores a lo largo de todas las imágenes, obtendremos las propiedades que definen un ladrillo apto y uno no apto.

2.3.1 Método basado en la matriz de co-ocurrencia (GLCM)

La matriz GLCM almacena valores que explican con cuánta frecuencia se relaciona espacialmente un píxel de un nivel de gris con otro píxel con otro nivel de gris específico. A estas matrices se las considera medidas de segundo nivel, ya que involucran una relación estadística entre un píxel en cierta ubicación espacial y otro píxel desplazado cierta distancia y dirección respecto del primer píxel [11]. Aquí (ver Fig. 5), se calcula la matriz GLCM para cada una de las imágenes disponibles y a estas matrices se le calculan diferentes propiedades: ‘contraste', 'disimilitud', 'homogeneidad', 'ASM', 'energía' y 'correlación'. Cada una de estas propiedades serán utilizadas como entrada para el proceso de clasificación.

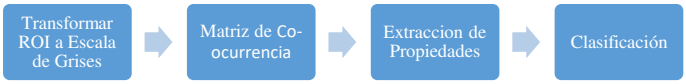


Fig. 5. Secuencia de procesamiento del método basado en GLCM.

ISBN 978 -987-633-574-4

2.3.2 Método basado en filtrado de Gabor

El filtro de Gabor es un filtro lineal que nos permite identificar el contenido frecuencial de una imagen en una dirección específica y en una zona de estudio determinada. Este se define en base a un kernel cuya expresión es la siguiente [11]:

$$Kernel = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda}+\psi\right)\right) \quad (2)$$

donde σ es la desviación estándar, γ es la relación de aspecto espacial, λ es la longitud de onda, ψ es la fase y θ es la orientación de la normal de las líneas paralelas de la función y

$$\begin{cases} x' = x\cos\theta + y\sin\theta \\ y' = -x\sin\theta + y\cos\theta \end{cases} \quad (3)$$

Teniendo en cuenta esta expresión, es posible ver que la variación de cualquiera de los parámetros provoca la generación de un kernel distinto y, por lo tanto, existe una cantidad enorme de posibilidades. Para las pruebas realizadas aquí, se establece que σ pertenece a [1:5] con paso 2, θ pertenece a [0:π] con paso π/4 y γ pertenece a [0,05:0,5], logrando así, generar una colección de 24 kernels por cada imagen. Estos kernels son aplicados sobre la zona de interés para generar 24 imágenes filtradas de una misma imagen y a cada una de estas imágenes filtradas se le calculará su media. De esta manera, tendremos 24 propiedades disponibles sobre cada una de las imágenes, y estas conforman el vector de características que se utilizaran en el proceso de clasificación (un esquema de este proceso se ve en la Figura 6).

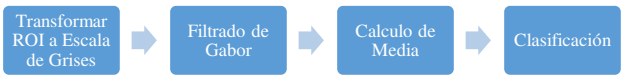


Fig. 6. Secuencia de procesamiento del método basado en filtro de Gabor.

2.3.3 Método de patrones binarios locales (LBP)

El método de LBP, al igual que GLCM, es una medida de segundo nivel, mediante la cual se genera una matriz que almacena datos correspondientes a una relación entre los píxeles [12]. En este caso se genera una matriz que almacena valores que relacionan un píxel con sus vecinos, evaluando si cada vecino es mayor o menor que el píxel central, según la siguiente expresión [13]:

$$\sum_{n=0}^{7} s(i_n - i_c)2^n = \sum_{n=0}^{7} s(z)2^n = Codigo\ LBP \quad (4)$$

donde i_n representa al píxel vecino, i_c es el píxel central, n representa el número de vecinos y $s(z)$ está definido según la siguiente expresión [13]:

ISBN 978 -987-633-574-4

$$s(z) = \begin{cases} 1, z \geq 0 \\ 0, z < 0 \end{cases} \quad (5)$$

Este resultado se guarda en la matriz. Aquí, se obtiene la matriz LBP para cada una de las imágenes disponibles y a dicha matriz se le calcula el histograma, al cual se lo conoce como "vector de características" y es la entrada para el proceso de clasificación.

2.4 Clasificación

Tal como se mencionó anteriormente, con cada uno de los procedimientos se extrajo una o varias características que se utiliza luego para el procedimiento de clasificación mediante el algoritmo de k-vecinos más cercanos (KNN) [14]. Este es uno de los algoritmos más simples de implementar y proporciona una buena precisión de clasificación. El algoritmo KNN (Fig. 7) está basado en una función de distancia euclídea y una función de voto de k-vecinos más cercanos. El procedimiento considera un cierto caso de prueba, encuentra los K vecinos más cercanos (vecinos correspondientes a los casos de entrenamiento) en el espacio de características y en base a esto se toma una decisión sobre la clase a la que pertenece este caso de prueba. En el ámbito del aprendizaje automático, este método se clasifica como un algoritmo supervisado, el cual es muy útil para conjuntos de datos pequeños y sin una cantidad grande de características, por esto fue considerado como el algoritmo a utilizar para este trabajo.

Fig. 7. Procedimiento de clasificación utilizando KNN.

3 Resultados y discusiones

De cada método propuesto, se extrajeron las características que son utilizadas en el clasificador. Para los métodos 2.2.1 y 2.2.2 se usó el valor del área obtenida por la segmentación como característica distintiva, para el método 2.3.1 se utilizaron las 6 características extraídas de la matriz GLCM. Luego para el método 2.3.2 se usó la media de las imágenes filtradas, obteniendo 24 características distintivas y, finalmente para el método 2.3.3 se utilizó el vector de 28 valores presentes en el histograma.

El desbalance presente en las clases se tuvo en cuenta a la hora de separar los ejemplos en ladrillos aptos y no aptos. Primero se dividió el conjunto de imágenes de forma aleatoria en 80% para entrenamiento y 20% para prueba, considerando el desbalance. El bloque de entrenamiento se volvió a dividir en 80% para entrenamiento

ISBN 978 -987-633-574-4

propiamente dicho y 20% para validación (evitando el sobre-entrenamiento) [15,16]. La exploración se repitió 500 veces y se consideraron distintos valores de K-vecinos (de 1 a 10), y luego se computó un promedio de los resultados, utilizando las métricas F1-Score y UAR (del inglés, *Unweighted Average Recall*). El objetivo fue encontrar el mejor K para cada método y luego realizar la evaluación sobre el conjunto de test.

En las figuras 8 y 9 se puede ver cómo se comporta el clasificador para cada uno de los métodos, variando la cantidad de vecinos utilizados en el algoritmo. De esta forma es posible encontrar el K óptimo en cada caso. Los primeros dos métodos (referidos a segmentación) dan resultados muy similares entre ellos, por debajo de los otros y mejoran su rendimiento según aumenta el K. La diferencia puede estar asociada a la cantidad de características extraídas para cada método. En un trabajo futuro se evaluará la combinación de las distintas características.

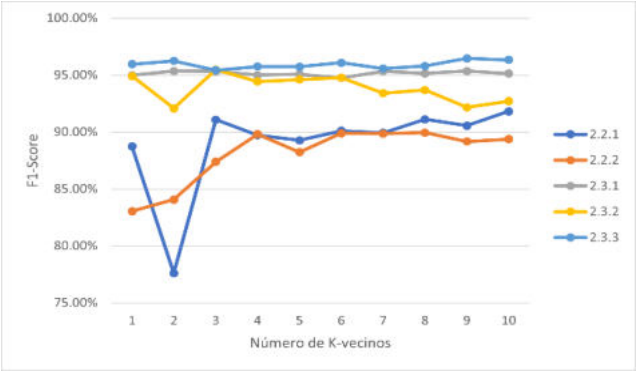


Fig. 8. Resultados promedio para la búsqueda del mejor K (F1-Score).

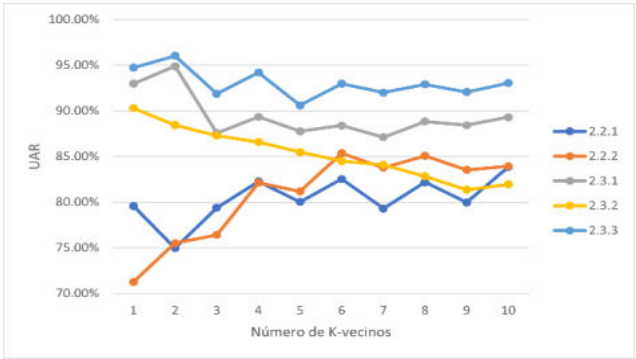


Fig. 9. Resultados promedio para la búsqueda del mejor K (UAR).

Una vez obtenidos los K óptimos de cada método, se procede a realizar la evaluación final sobre los datos de prueba que no fueron utilizados. Este proceso completo se repitió 10 veces, donde se promediaron los porcentajes resultantes. Debido a que las 10 repeticiones vuelven a generar aleatoriamente los grupos de entrenamiento y prueba, el K optimo puede variar levemente en cada repetición, por lo que se utilizó la moda estadística para determinar el resultado final, mostrado en las Tablas 1 y 2.

ISBN 978 -987-633-574-4

Tabla 1. Resultados finales utilizando K óptimo basado en F1-Score.

Método	K óptimo	F1-Score	UAR
2.2.1	10	90,51 %	85,00 %
2.2.2	5	88,14 %	84,03 %
2.3.1	2	**98,41** %	**96,94** %
2.3.2	1	96,42 %	91,25 %
2.3.3	2	96,60 %	95,28 %

Tabla 2. Resultados finales utilizando K óptimo basado en UAR.

Método	K óptimo	F1-Score	UAR
2.2.1	10	86,60 %	75,00 %
2.2.2	8	87,51 %	77,50 %
2.3.1	2	**99,00** %	**97,50** %
2.3.2	1	94,83 %	88,19 %
2.3.3	7	95,93 %	94,72 %

Con base en estos resultados, es posible decir que los métodos basados en texturas siempre dan resultados superiores a los de segmentación Cuando se utiliza F1-Score para la búsqueda del K óptimo, se obtiene al menos un 6% absoluto más sobre UAR. Mientras que, cuando se busca el K óptimo sobre UAR, se obtiene una diferencia de más del 10%. Si bien LBP funciona muy bien, el método basado en GLCM obtiene los mejores resultados. Su rendimiento es muy prometedor ya que presenta menos de un 2,5% de error para las métricas analizadas.

4 Conclusiones y trabajos futuros

En este trabajo se propusieron métodos para clasificar un ladrillo de hormigón celular en aptos para su uso o no. Para esto, se propusieron métodos de segmentación basados en umbralización, los cuales permiten interpretabilidad en los resultados, ya que se pueden inferir directamente la cantidad de poros, sus tamaños, entre otras propiedades. Los resultados obtenidos con estos métodos fueron bastante buenos, como se puede observar en los experimentos presentados. Sin embargo, sigue siendo un procedimiento problemático, debido a su excesiva dependencia a las características de la imagen respecto de la forma de captura. Por otro lado, los métodos basados en texturas hacen más difícil la interpretabilidad, ya que no es posible determinar ciertas cuestiones simples, como la cantidad de poros. No obstante, con éstos métodos se obtienen resultados más certeros y una independencia muy deseable respecto de las condiciones de luz, por lo que sería posible un margen de variación de la escena mucho mayor.

Como trabajo futuro, se prevé aumentar el número de imágenes en la base de datos para confirmar el buen comportamiento del modelo. Además, se realizarán análisis para comprobar si los métodos tienen robustez a los cambios en la escena. A mediano plazo se propone la captura de imágenes con cámaras de celular, la realización del proceso en tiempo real y validar la usabilidad en ambientes reales no controlados.

Agradecimientos

Los autores desean agradecer al instituto sinc(i) UNL-CONICET, a UNL (con CAI+D 50620190100145LI) y al ingeniero Federico Ortiz de Zarate por su apoyo y contribución como asesor temático.

Referencias

1. Macedo, L.: Construcción en seco vs tradicional de ladrillo, https://webcapp.com/blog/index.php/2020/04/07/diferencias-sistema-construccion-en-seco-vs-tradicional/.
2. Gobierno nacional argentino: Vivienda y construcción sostenible, https://www.argentina.gob.ar/ambiente/desarrollo-sostenible/vivienda.
3. Dejtiar, F.: ¿Qué es el hormigón celular curado en autoclave y cuáles son sus ventajas en la arquitectura?, https://www.plataformaarquitectura.cl/cl/918535/que-es-el-hormigon-celular-curado-en-autoclave-y-cuales-son-sus-ventajas-en-la-arquitectura.
4. Visión Técnica: Hormigón Celular Curado en Autoclave, http://www.visiontecnica.com.ar/index.php/nuevos-materiales/8-hormigon-celular-curado-en-autoclave.
5. Qu, X., Zhao, X.: Previous and present investigations on the components, microstructure and main properties of autoclaved aerated concrete – A review. Construction and Building Materials. 135, pp. 505--516 (2017).
6. Rodríguez, J.: Porosimetría por inyección de mercurio 1st ed. Universidad de Oviedo, Oviedo (2002), http://ocw.uniovi.es/pluginfile.php/4888/mod_resource/content/1/T3b-PorosimetriaMercurio.pdf.
7. Wan, H., Hub, Y., Liu, G., Qu, Y.: Study on the structure and properties of autoclaved aerated concrete produced with the stone-sawing mud. Construction and Building Materials. 184, pp. 20--26 (2018).
8. Bradski, G.: "The OpenCV Library". Dr. Dobb's Journal of Software Tools, (2000).
9. Gonzales, R., Woods, R.: Digital Image Processing 3rd ed. In: McDonald, M. (ed.). Pearson Prentice Hall, pp. 642--676 (2008).
10. Van Der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T. and the scikit-image contributors: scikit-image: Image processing in Python. PeerJ 2:e453 (2014)
11. Kak, A.: Measuring Texture and Color in Images 2nd ed. Purdue University, Indiana (2020), https://engineering.purdue.edu/kak/Tutorials/TextureAndColor.pdf.
12. Rosebrock, A.: Local Binary Patterns with Python & OpenCV, https://www.pyimagesearch.com/2015/12/07/local-binary-patterns-with-python-opencv/.
13. Sairamya, N.J., Susmitha, L., George, S.T., Subathra, M.S.P.: Intelligent Data Analysis for Biomedical Applications. In: Hemanth, D. J., Gupta, D., Balas, V. E. (eds.). Academic Press, pp.253--273 (2019).
14. Harrison, O.: Machine Learning Basics with the K-Nearest Neighbors Algorithm, https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761.
15. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825--2830 (2011).
16. Bishop, Christopher M. "Pattern recognition." Machine learning 128.9 (2006).

ISBN 978 -987-633-574-4

Post COVID-19 Cognitive disorders: Virtual Reality and Augmented Reality as mental healthcare tools

Yoselie Alvarado, Graciela Rodríguez, Nicolás Jofré,
Jacqueline Fernandez, and Roberto Guerrero

Laboratorio de Computación Gráfica (LCG)
Universidad Nacional de San Luis,
Ejército de los Andes 950
Tel: 02664 420823, San Luis, Argentina
{ymalvarado,gbrodriguez,npasinetti,jmfer,rag}@unsl.edu.ar

Abstract. Some time ago Virtual Reality and Augmented Reality were exclusively devoted to the gaming industry. Nowadays, both technologies are experiencing a deep interest from various spheres, including healthcare sector.
The new infectious disease COVID-19 has had a catastrophic effect on the world's demographics. Many patients with mild or severe COVID-19 do not recover completely and present with a wide variety of chronic symptoms after infection, often of a neurological, cognitive or psychiatric nature. The most common signs of cognitive disorder can be summarized as *mental fog*, *memory* problems and *concentration* problems.
The aim of this study was to analyze the opportunities for Virtual and Augmented Reality in the cognitive interventions related to mentioned disorders by searching for articles in scientific databases. We conclude that as these technologies and devices become cheaper and accessible worldwide, can at least be regarded as a rehabilitation therapy as effective as traditional training, and to some extent better than it.

Keywords: COVID-19, Mental Health, Cognitive Disorders, Virtual Reality, Augmented Reality.

1 Introduction

Virtual Reality (VR) and *Augmented Reality* (AR) systems are currently understood as entertainment tools, used for watching movies, playing video games, and even to immerse oneself into the digital Non-Fungible Tokens (NFT) market. Virtual reality gives people a hands-on experience of just about anything. It has not just moved the imagination of science-fiction fans, but also clinical researchers and real-life medical practitioners.

The healthcare sector is an area with fascinating possibilities. It is one of the early adopters of VR/AR and is making the most out of it. The launch of Oculus Rift and HTC Vive has further augmented the usage of Virtual Reality and

ISBN 978 -987-633-574-4

Augmented Reality [1, 2]. They triggered a huge increase in the use of VR/AR in the healthcare sector which continues to be fascinated y impressed by the diverse potential applications.

There are increasingly great examples of VR/AR having a positive effect on patients' lives and physicians' work. From developing new life-saving techniques to training the doctors of the future, VR/AR have a multitude of applications for health and healthcare, from the clinical to the consumer. At this time, the key application areas of VR/AR in healthcare are *medical training*, *patient treatment*, *medical marketing* and *disease awareness* [3, 4].

Treating Mental Health Issues like anxiety, panic attacks, and other mind-related issues such as phobias are great examples of medical patient treatment VR/AR transforming patient lives. Usually, patients are immersed in a safe and controlled environment and confronted with their fears recreated virtually. In more complicated situations such as Post-Traumatic Stress Disorder (PTSD), the Virtual Reality Exposure Therapy (VRET) technology has yielded positive results, giving patients another chance at life [5, 6].

The COVID-19 pandemic worsened the suffering of people with mental and physical ailments. The healthcare system has never been pit against an enemy such as COVID-19, forcing us to look for innovative solutions that make global healthcare more flexible and future-ready for such disruptions. Global healthcare is turning to VR/AR, which certainly make for a lucrative prospect for the future. They are helping in better preparing the healthcare systems for pandemics and global health crises, such as the one we face now [7]. And while other industries are jumping on the VR wagon, hospitals, medical institutions, and healthcare tech companies are adapting to VR space equally well [8].

Research in past epidemics has revealed a deep and wide range of psychosocial consequences at the individual and community levels during outbreaks. There are multiple associated psychological disturbances, ranging from isolated symptoms to complex disorders with marked impairment of functionality, such as insomnia, anxiety, depression, and post-traumatic stress disorder. Therefore, it is necessary for mental health services to develop strategies that allow them to react with skill and to achieve support for health personnel and the affected population, in order to reduce the development of psychological impacts and psychiatric symptoms.

In this way, the aim of this paper is to analyze and classify the use of new technologies as a therapeutic element in cognitive disorders. The work will review existing systems and applications about VR and AR technologies to achieve telerehabilitation. Research will focus on assessed applications that enhance cognitive rehabilitation interventions for sequels caused by COVID-19.

Section 2 details some statistics on post COVID-19 symptoms and lists the signs associated with cognitive disorders. Section 3 presents the benefits of applying alternative realities in cognitive therapies. Section 4 gives a brief overview of existing applications and how they can be used for the treatment of cognitive disorders according to the cognitive signs detected. Section 5 provides a small discussion and future guidelines.

ISBN 978 -987-633-574-4

2 Post Covid-19 Symptoms

Since December 2019, COVID-19 has rapidly spread worldwide, affecting people in 210 countries and territories with the current tally exceeding 200 million infected people and more than 4,265,903 deaths [9].

In December 2020, the UK National Institute for Health and Care Excellence (NICE) published guidance on the long-term consequences of COVID-19. This guidance distinguishes between *acute COVID-19* (signs and symptoms of COVID-19 last up to 4 weeks), *ongoing symptomatic COVID-19* (signs and symptoms of COVID-19 are 4 to 12 weeks in duration), and *post COVID-19* syndrome. NICE guidance defines post COVID-19 syndrome as the cluster of signs and symptoms that develop during or after a COVID-19-compatible infection, continue for more than 12 weeks, and are not explained by an alternative diagnosis. Symptoms can often occur overlapping, and fluctuate and change over time, sometimes in a flare-like fashion, and affect any body system, including cardiovascular, respiratory, gastrointestinal, neurological, musculoskeletal, metabolic, renal, dermatological, ENT (Ear, Nose and Throat Infections), and hematologic systems, in addition to psychiatric problems, generalized pain, fatigue, and persistent fever.

Seventy-six percent of those affected reported at least one symptom during follow-up and a higher percentage of prolonged symptoms was observed in patients aged 40 to 60 years and those who were hospitalized. The most common symptoms after discharge were fatigue or muscle weakness (63%) and sleeping difficulties (26%). Anxiety or depression was also reported in 23% of affected people. Studies on psychological sequelae show that almost 7 out of 10 (64.4%) of the participants reported cognitive failures (mild, moderate and severe). Half (46.3%) responded that their attention worsened and more than 4 out of 10 (43.1%) also reported that their memory worsened after being infected with COVID-19 [10, 11].

Cognitive disorders are a category of neurological disorders [12, 13]. Cognitive disorders are defined as any disorder that significantly impairs the cognitive function of an individual to the point where normal functioning in society is impossible without treatment. They primarily affect cognitive abilities including, learning, memory, perception, and problem-solving.

Cognitive disorder signs vary according to the particular disorder, but some common signs and symptoms overlap in most disorders. The most common signs of cognitive disorder can be summarized as *Mental fog*, *Memory problems* and *Concentration problems*.

- **Mental Fog**. Clouding of consciousness (also known as brain fog or mental fog) is when a person is slightly less wakeful or aware than normal. They are not as aware of time or their surroundings and find it difficult to pay attention. Typical symptoms of brain fog include poor concentration, an extra effort to focus on a task, trouble multitasking or managing too many tasks at once and trouble tracking what you are doing.
- **Memory Problems**. Memory loss can be defined as pathological forgetting: to learn something new, to recover memories from the past, or to remember

ISBN 978 -987-633-574-4

specific events. When we forget something, it is not usually that we "lose" the memory itself, but that our brain "can't find its way" to the memory we are trying to find. Our brain engages different structures to work with different types of memory. The main memory modalities are: *short-term* memory and *long-term* memory.

– **Concentration Problems**. Concentration difficulty is a decreased ability to focus your thoughts on something. Concentration difficulties can be related to difficulty staying awake, impulsiveness, intrusive thoughts or concerns, overactivity, or inattention.

The factors that have proven to be most effective in preventing and treating cognitive disorders are: *adequate sleep*, a *good diet*, *physical exercise*, an *active social life* and *cognitive activities*. The brain acts much like our muscles, so the more we use it, the better shape it will be in. **Cognitive Stimulation** seeks to stimulate, train and strengthen the different cognitive abilities of people, such as attention, perception, memory, language and executive functions.

Although the signs of cognitive impairment mentioned above are clearly different, the treatment of some of them can be addressed together. For example, for mental fog and memory problems cognitive stimulation through games, everyday tasks, and memory exercises is recommended. For concentration problems, arithmetic exercises, try to number, word search puzzles, memorizing images, among others are recommended.

3 Alternative Technologies in Cognitive Therapies

Cognitive rehabilitation encompasses a wide range of therapeutic cognitive interventions to achieve functional changes by reinforcing, strengthening, or reestablishing previously learned patterns of behavior or establishing new patterns of cognitive activity or mechanisms to compensate for impaired neurological systems.

These interventions are based on psychological theories and models of behavior and behavioral change and on neuropsychological models of brain–behavior interactions, and for many years can be conducted with paper–pencil tools, social skills training, physics skills training, cognitive rehabilitation and psychostimulant medication in some cases.

Related to this kind of therapies, many scientists suggest that it is an inadequate form of intervention. The main reasons are as follows: stimulants do not work for all people and the psychostimulant effects are limited to the period in which the drugs are physiologically active. On the other hand, some studies have shown that in order to improve functioning, interventions should include training of more specific daily-life skills and compensatory strategies. Another critical barrier for effective treatment is a lack of motivation in the patients to participate in the assigned training despite receiving encouragement and support. Perhaps because of these problems with limited transfer and motivation, attrition rates are often high in cognitive rehabilitation programmes.

ISBN 978 -987-633-574-4

In this context, it has been proven that training/therapy based on virtual and augmented technologies is feasible to accommodate these problems because of its highly engaging and gamified format. Virtual reality can be defined as a naturalistic simulated environment with which the user can interact as if the user was present. With the possibility for a fully controlled and safe environment the technology offers a more ecological valid environment for cognitive rehabilitation as it enables a multimodal setting that is quite similar to situations that patients might encounter in their daily lives. Thus, cognitive training can be integrated more easily with daily life functioning. On the other hand, AR is the integration of digital and physical information in real-time that allows the user interaction with a virtual and real world. These emerging technologies are a great promise because they motivate people with new challenges; providing rapid feedback that is tailored to their specific interests and individual needs.

4 Applications

As stated above, the signs and symptoms of COVID-19 cognitive impairment overlap in most disorders. Therefore, existing VR/AR applications can be classified into two main groups: applications that address Attention and Concentration problems; and applications that address Mental Fog and Memory problems. Figures 1 and 2 show a diagram of the systems surveyed, specifying the authors, year of implementation, purpose of the application, technology used and name of the system (if any).

4.1 Mental Fog and Memory systems

The applications include systems that were originally developed for specific diseases (such as Alzheimer's, Epilepsy, Dementia, Traumatic Brain Injury) and that are able to improve memory skills. Most of these applications are concerned with enhancing the processes of encoding, storage and recovery of information. In certain cases it has been a common practice to use external aids to promote memory, such as assistants, diaries, alarms, etc. [14–33].

4.2 Attention and Concentration Systems

The main objectives of medical intervention programs are to increase attentional capacity, decrease response time, reduce the phenomenon of hemineglect, and improve various components of the attentional system, such as sustained, selective, alternating or divided attention. The most commonly treated cases are *Attention Deficit Hyperactivity Disorder* (ADHD) and *Autism Spectrum Disorder* (ASD). In the 1990s, these programs began to include alternative technologies such as virtual reality and augmented reality in their therapies. Nowadays, it is no longer an inclusion but rather of computer-based cognitive therapy such as the following virtual/augmented reality-based cognitive rehabilitation systems [34–57].

ISBN 978 -987-633-574-4

VIRTUAL REALITY				
Authors	**Year**	**Aim**	**Technology**	**Name**
W. H. Guo et al. [14]	2004	Memory deficits	HMD	-
Veronika Brezinka [15]	2011	Cognitive-behavioural	Web	Treasure Hunt
Emmanuelle Chapoulie et al. [16]	2014	Reminiscence	Immersive IBR	IVIRAGE
Monthon Intraraprasit et al. [17]	2017	Cognitive impairment	HMD	-
Fernando A. Chicaiza et al. [18]	2018	Memory loss	HMD	-
Kiran Ijaz et al. [19]	2019	Predementia	HMD	VR-CogAssess
Andrea Vitali et al. [20]	2021	Memory loss	HMD	-
AUGMENTED REALITY				
Authors	**Year**	**Aim**	**Technology**	**Name**
S. Wood et al. [21]	2012	Memory loss	Mobile App	TARDIS
Eduardo Quintana et al. [22]	2012	Alzheimers	Mobile App	ANS
M. Carmen Juan et al. [23]	2014	Spatial memory	Mobile App	ARSM
Oscar Rosello et al. [24]	2016	Memorization	Mobile App, Headset	NeverMind
Mat Masir et al. [25]	2016	Dementia	AR Desktop	DARD
Costas Boletsis et al. [26]	2016	Dementia	Mobile App	CogARC
Leah Gilbert et al. [27]	2017	Memory impairment	Mobile App	-
Dennis Wolf et al. [28]	2018	Dementia	Mobile App, HMD	cARe
Keynes Masayoski et al. [29]	2018	Alzheimers	Mobile App	-
F. Munoz-Montoya et al. [30]	2019	Spatial memory	Mobile App	-
Jonne Schoneveld [31]	2020	Dementia	Mobile App	-
Arezou Niknam [32]	2021	Spatial Memory	Mobile App	-
Rui Silva et al. [33]	2021	Cognitive Therapy	Mobile App	SAR-ACT

Fig. 1: Mental Fog and Memory systems.

VIRTUAL REALITY				
Authors	**Year**	**Aim**	**Technology**	**Name**
Joan Mc. Comas et al. [34]	2002	ADHD	PC (3 displays)	-
N. Yan, Jue Wang et al. [35]	2008	ADHD	VR-integrated	IVA-CPT
Meghan Elizabeth Huber [36]	2008	Hemiplegia	PS3	-
Shih-Ching Yeh [37]	2012	ADHD	HMD	-
Silvia Erika Kober et al. [38]	2013	Spatial disorientation	PC, Joystick, Mic	-
Pierre Nolin [39]	2016	ADHD	HMD	ClinicaVR
Kim C M Bul [40]	2016	ADHD	PC	-
Jofre et al. [41]	2018	Attention motivating	Kinect, PC/CAVE	-
Maria Cristina Barba [42]	2019	ADHD	Kinect, HMD, EEG	BRAVO
Manish Kumar Jha [43]	2020	Alzheimer's disease	HMD	-
XRHealth [44]	2020	ADHD	HMD	XRHealth
AUGMENTED REALITY				
Authors	**Year**	**Aim**	**Technology**	**Name**
Mohd Azmidi Abdullah et al. [45]	2012	ADHD Kids	PC, webcam	ADHD-Edu
Lizbeth Escobedo [46]	2014	Attention disorientation	Mobile App	Mobis
Hendrys Tobar-Muñoz et al. [47]	2014	ADHD	PC, webcam	Gremlings
Jorge Bacca [48]	2015	Learning motivation	Mobile App	-
Chien-Yu Lin et al. [49]	2016	ADHD	Mobile App	MAR
Martín Sabarís [50]	2017	Down syndrome	Mobile App	-
I-Jui Lee et al. [51]	2018	ASD kids	PC, webcam	AR-RPG
Diego Avila-Pesantez et al. [52]	2018	ADHD kids	Mobile App	ATHYNOS
Eleni Mangina et al. [53]	2018	ADHD	Mobile App	AHA
Maria Cristina Barba et al. [42]	2019	ADHD	Kinect, HMD, EEG	BRAVO
Tasneem Khan et al. [54]	2019	Learning motivation	Mobile App	-
Saad Alqithami et al. [55]	2020	ADHD	HoloLens	AR-Therapist
Neha U. Keshav et al. [56]	2019	ADHD	HoloLens	D. Attention-Related AR
Katherine Wang et al. [57]	2020	Attention disorientation	Mobile App	MARA

Fig. 2: Attention and Concentration Systems.

ISBN 978 -987-633-574-4

5 Discussion and Conclusions

This section briefly analyzes the works surveyed in terms of the use of VR and AR technologies as therapeutic tools in the treatment of the cognitive disorders addressed and outlines some conclusions to be taken into consideration for future works.

The launch of Oculus Rift and HTC Vive, together with the emergence of smartphones, enabled access to technology with great potential for VR and AR developments. As a consequence, it triggered a boom in all kinds of research related to these technologies. Additionally, the current pandemic has forced many treatments to be performed by telehealth, further encouraging developments that use everyday devices instead of complex VR and AR devices.

Regarding the use of AR technology, both for concentration and memory, AR appears to be a more recent area of research. It was initially supported by notebooks and webcams and then became strongly dominated by the use of smartphones. Smartphones are currently the main device used in AR applications. In particular, for memory problems, implementations focus on AR over VR in line with the study by Niknam which suggests that using AR is better than VR for cognitive applications to enhance spatial working and long-term memory [32].

Regarding the use of VR technology, there is evidence of work carried out since the 1990s, both for its application to concentration and memory problems. Therefore, VR is a very robust field. In the last decades a significant growth of immersive systems using VR headsets with smartphones has been reported.

Concluding, 45 articles and texts have been included, which showed a relationship between VR, AR, and the investigated disorders. Of the included articles, 33 involved evaluations with patients, and all of them proposed their systems as therapeutic alternatives. Articles not providing additional information for the purposes of this work, as well as those more than 20 years old, were excluded.

Healthcare professionals suggested that rehabilitation therapy based on VR and AR technologies are likely more effective than conventional therapy. On the other hand, it should be noted that there is still a lot of work to be done in terms of optimization and availability of the applications surveyed. Furthermore, there is a growing number of doctors who are turning to VR for treatments for phobia, surgery simulations, skills training, etc.

A lot of the applications mentioned above are still in their infancy. In the coming years, VR/AR will be used more and more to improve the accuracy & effectiveness of current procedures, and enhance the capabilities of the human being, both as the care-giver and the patient. Quite simply, the potential for VR in the healthcare sector is huge, limited only by the creativity & ingenuity of those creating and applying the technology. However, experts predict that AR will gain greater traction in the sector in the years to come.

In a pandemic context, the global demand for medical needs over the internet sets the stage for virtual reality to step up and claim what has been long pending: a completely immersive, sensitive VR healthcare experience.

Finally, we believe that as VR/AR technologies and devices become cheaper and accessible worldwide, these technologies can at least be regarded as a rehabilitation therapy as effective as traditional training, and to some extent better than it.

References

1. Parth Rajesh Desai, Pooja Nikhil Desai, Komal Deepak Ajmera, and Khushbu Mehta. A review paper on oculus rift-a virtual reality headset. *CoRR*, abs/1408.1173, 2014.
2. Pietro Cipresso, Irene Alice Chicchi Giglioli, Mariano Alcañiz Raya, and Giuseppe Riva. The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, 9:2086, 2018.
3. Mythreye Venkatesan, Harini Mohan, Justin R. Ryan, Christian M. Schürch, Garry P. Nolan, David H. Frakes, and Ahmet F. Coskun. Virtual and augmented reality for biomedical applications. *Cell Reports Medicine*, 2(7):100348, 2021.
4. Xuanhui Xu, Eleni Mangina, and Abraham G. Campbell. Hmd-based virtual and augmented reality in medical education: A systematic review. *Frontiers in Virtual Reality*, 2:82, 2021.
5. Wenrui Deng, Die Hu, Sheng Xu, Xiaoyu Liu, Jingwen Zhao, Qian Chen, Jiayuan Liu, Zheng Zhang, Wenxiu Jiang, Lijun Ma, Xinyi Hong, Shengrong Cheng, Boya Liu, and Xiaoming Li. The efficacy of virtual reality exposure therapy for ptsd symptoms: A systematic review and meta-analysis. *Journal of Affective Disorders*, 257:698–709, 2019.
6. L.V. Eshuis, M.J. Gelderen, Mirjam Zuiden, Mirjam Nijdam, Eric Vermetten, Miranda Olff, and A. Bakker. Efficacy of immersive ptsd treatments: A systematic review of virtual and augmented reality exposure therapy and a meta-analysis of virtual reality exposure therapy. *Journal of Psychiatric Research*, 11 2020.
7. Elisa Mantovani, Chiara Zucchella, Sara Bottiroli, Angela Federico, Rosalba Giugno, Giorgio Sandrini, Cristiano Chiamulera, and Stefano Tamburin. Telemedicine and virtual reality for cognitive rehabilitation: A roadmap for the covid-19 pandemic. *Frontiers in Neurology*, 11:926, 2020.
8. Ravi Pratap Singh, Mohd Javaid, Ravinder Kataria, Mohit Tyagi, Abid Haleem, and Rajiv Suman. Significant applications of virtual reality for covid-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4):661–664, 2020.
9. World Health Organization. https://covid19.who.int/.
10. Dr Elaine Maxwell. https://evidence.nihr.ac.uk/themedreview/living-with-covid19-second-review/.
11. D. Menges, T. Ballouz, A. Anagnostopoulos, H. E. Aschmann, A. Domenghino, J. S. Fehr, and M. A. Puhan. Burden of post-covid-19 syndrome and implications for healthcare service planning: A population-based cohort study. *PloS one*, 2021.
12. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5*. Autor, Washington, DC, 5th ed. edition, 2013.
13. Carod-Artal F. J. Post-covid-19 syndrome: epidemiology, diagnostic criteria and pathogenic mechanisms involved. *Revista de neurologia*, pages 384–396, 2021.
14. WH Guo, Samuel YE Lim, Sai Cheong Fok, and GYC Chan. Virtual reality for memory rehabilitation. *International journal of computer applications in technology*, 21(1-2):32–37, 2004.

15. Veronika Brezinka. "treasure hunt" - a cognitive-behavioural computer game. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 60:762–76, 01 2011.
16. Emmanuelle Chapoulie, Rachid Guerchouche, Pierre-David Petit, Gaurav Chaurasia, Philippe Robert, and George Drettakis. Reminiscence therapy using image-based rendering in vr. In *2014 IEEE Virtual Reality (VR)*, pages 45–50. IEEE, 2014.
17. Monthon Intraraprasit, Phanuthon Phanpanya, and Chompoonuch Jinjakam. Cognitive training using immersive virtual reality. In *2017 10th Biomedical Engineering International Conference (BMEiCON)*, pages 1–5. IEEE, 2017.
18. Fernando A Chicaiza, Luis Lema-Cerda, V Marcelo Álvarez, Víctor H Andaluz, José Varela-Aldás, Guillermo Palacios-Navarro, and Iván García-Magariño. Virtual reality-based memory assistant for the elderly. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 269–284. Springer, 2018.
19. Kiran Ijaz, Naseem Ahmadpour, Sharon L Naismith, and Rafael A Calvo. An immersive virtual reality platform for assessing spatial navigation memory in predementia screening: Feasibility and usability study. *JMIR mental health*, 6(9):e13887, 2019.
20. Andrea Vitali, Daniele Regazzoni, Caterina Rizzi, and Andrea Spajani. Vr serious games for neuro-cognitive rehabilitation of patients with severe memory loss. *Computer-Aided Design and Applications*, 18:1233–1246, 02 2021.
21. S Wood and RJ McCrindle. Augmented reality discovery and information system for people with memory loss. In *Proceedings of the 9th International Conference on Disability, Virtual Reality & Associated Technologies*, pages 10–12, 2012.
22. Eduardo Quintana and Jesus Favela. Augmented reality annotations to assist persons with alzheimers and their caregivers. *Personal and ubiquitous computing*, 17(6):1105–1116, 2013.
23. M-Carmen Juan, Magdalena Mendez-Lopez, Elena Perez-Hernandez, and Sergio Albiol-Perez. Augmented reality for the assessment of children's spatial memory in real settings. *PloS one*, 9(12):e113751, 2014.
24. Oscar Rosello, Marc Exposito, and Pattie Maes. Nevermind: Using augmented reality for memorization. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16 Adjunct, page 215–216, New York, NY, USA, 2016. Association for Computing Machinery.
25. EMNE Mat Nasir, Nan Md Sahar, and AH Zainudin. Development of augmented reality application for dementia patient (dard). *ARPN Journal of Engineering and Applied Sciences*, 2016.
26. Costas Boletsis and Simon McCallum. Augmented reality cubes for cognitive gaming: Preliminary usability and game experience testing. *International Journal of Serious Games*, 3(1), Mar. 2016.
27. Leah Gilbert, Annika Hinze, and Judy Bowen. Augmented reality game for people with traumatic brain injury: Concept and prototypical exploration. In *Proceedings of the 9th International Conference on Computer and Automation Engineering*, ICCAE '17, page 51–55, New York, NY, USA, 2017. Association for Computing Machinery.
28. Dennis Wolf, Daniel Besserer, Karolina Sejunaite, Matthias Riepe, and Enrico Rukzio. Care: An augmented reality support system for dementia patients. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, UIST '18 Adjunct, page 42–44, New York, NY, USA, 2018. Association for Computing Machinery.

29. Keynes Masayoshi Kanno, Edgard Afonso Lamounier, Alexandre Cardoso, Ederaldo José Lopes, and Gerson Flávio Mendes de Lima. Augmented reality system for aiding mild alzheimer patients and caregivers. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 593–594, 2018.
30. Francisco Munoz-Montoya, M-Carmen Juan, Magdalena Mendez-Lopez, and Camino Fidalgo. Augmented reality based on slam to assess spatial short-term memory. *IEEE Access*, 7:2453–2466, 2018.
31. Jonne Schoneveld. Augmented reality photo album for people with dementia. B.S. thesis, University of Twente, 2020.
32. Arezou Niknam. An augmented reality mobile game design to enhance spatial memory in elderly with dementia. In *6th International Conference on Computer Games; Challenges and Opportunities (CGCO2021)*, 2021.
33. Rui Silva and Paulo Menezes. Sar-act: A spatial augmented reality approach to cognitive therapy. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1: GRAPP,*, pages 292–299. INSTICC, SciTePress, 2021.
34. Joan McComas, Morag Mackay, and Jayne Pivik. Effectiveness of virtual reality for teaching pedestrian safety. *Cyberpsychology and behavior: the impact of the Internet, multimedia and virtual reality on behavior and society*, 5:185–90, 07 2002.
35. Nan Yan, Jue Wang, Mingyu Liu, Liang Zong, Yongfeng Jiao, Jing Yue, Ye Lv, Qin Yang, Hai Lan, and Zhongye Liu. Designing a brain-computer interface device for neurofeedback using virtual environments. *Journal of Medical and Biological Engineering*, 28, 09 2008.
36. Meghan Huber, Bryan Rabin, Ciprian Docan, Grigore Burdea, Michelle Nwosu, Moustafa Abdelbaky, and Meredith Golomb. Playstation 3-based telerehabilitation for children with hemiplegia. volume 10, pages 105 – 112, 09 2008.
37. Shih-Ching Yeh, Chia-Fen Tsai, Yao-Chung Fan, Pin-Chun Liu, and Albert Rizzo. An innovative adhd assessment system using virtual reality. In *2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences*, pages 78–83, 2012.
38. Silvia Kober, Guilherme Wood, Daniela Schweiger, Walter Kreuzig, Manfred Kiefer, and Christa Neuper. Virtual reality in neurologic rehabilitation of spatial disorientation. *Journal of neuroengineering and rehabilitation*, 10:17, 02 2013.
39. Pierre Nolin, Annie Stipanicic, Mylène Henry, Yves Lachapelle, Dany Lussier-Desrochers, Albert "Skip" Rizzo, and Philippe Allain. Clinicavr: Classroom-cpt: A virtual reality tool for assessing attention and inhibition in children and adolescents. *Computers in Human Behavior*, 59:327–333, 2016.
40. Kim Bul, Pamela Kato, Saskia Oord, Marina Danckaerts, Leonie Vreeke, Annik Willems, Helga Oers, Ria Heuvel, Derk Birnie, Thérèse Amelsvoort, Ingmar Franken, and Athanasios Maras. Behavioral outcome effects of serious gaming as an adjunct to treatment for children with attention-deficit/hyperactivity disorder: A randomized controlled trial. *Journal of Medical Internet Research*, 18:e26, 02 2016.
41. Nicolás Jofré, Graciela Rodríguez, Yoselie Alvarado, Jacqueline Fernández, and Roberto Guerrero. Natural user interfaces: A physical activity trainer. In Armando Eduardo De Giusti, editor, *Computer Science – CACIC 2017*, pages 122–131, Cham, 2018. Springer International Publishing.
42. M. C. Barba, A. Covino, V. De Luca, L. T. De Paolis, G. D'Errico, P. Di Bitonto, S. Di Gestore, S. Magliaro, F. Nunnari, G. I. Paladini, A. Potenza, and A. Schena. Bravo: A gaming environment for the treatment of adhd. In Lucio Tommaso De Paolis and Patrick Bourdot, editors, *Augmented Reality, Virtual Reality, and*

Computer Graphics, pages 394–407, Cham, 2019. Springer International Publishing.

43. Manish Jha, Hamdi Ben Abdessalem, Marwa Boukadida, Alexie Byrns, Marc Cuesta, Marie-Andrée Bruneau, Sylvie Belleville, and Claude Frasson. *Virtual Reality Orientation Game for Alzheimer's Disease Using Real-Time Help System*, pages 13–23. 10 2020.
44. XRHealth. https://syncrovr.com/xrhealth-lanza-la-primera-clinica-de-realidad-virtual-telehealth/.
45. Mohd Azmidi Abdullah. Teaching adhd kids using augmented reality. 2013.
46. Lizbeth Escobedo, Monica Tentori, Eduardo Quintana, Jesus Favela, and Daniel Garcia-Rosas. Using augmented reality to help children with autism stay focused. *IEEE Pervasive Computing*, 13(1):38–46, 2014.
47. Hendrys Tobar-Muñoz, Ramón Fabregat, and Silvia Baldiris. Using a videogame with augmented reality for an inclusive logical skills learning session. In *2014 International Symposium on Computers in Education (SIIE)*, pages 189–194. IEEE, 2014.
48. Jorge Bacca, Silvia Baldiris, Ramon Fabregat, Sabine Graf, et al. Mobile augmented reality in vocational education and training. *Procedia Computer Science*, 75:49–58, 2015.
49. Chien-Yu Lin, Wen-Jeng Yu, Wei-Jie Chen, Chun-Wei Huang, and Chien-Chi Lin. The effect of literacy learning via mobile augmented reality for the students with adhd and reading disabilities. In *International conference on universal access in human-computer interaction*, pages 103–111. Springer, 2016.
50. Rosa-María Martín-Sabarís. Augmented reality for learning in people with down syndrome: an exploratory study. 2017.
51. I-Jui Lee, Ling-Yi Lin, Chien-Hsu Chen, and Chi-Hsuan Chung. How to create suitable augmented reality application to teach social skills for children with asd. In Nawaz Mohamudally, editor, *State of the Art Virtual Reality and Augmented Reality Knowhow*, chapter 8. IntechOpen, Rijeka, 2018.
52. Diego Avila Pesantez, Luis Rivera, Leticia Vaca-Cardenas, Stteffano Aguayo, and Lourdes Zuniga. Towards the improvement of adhd children through augmented reality serious games: Preliminary results. pages 843–848, 04 2018.
53. Eleni Mangina, Giuseppe Chiazzese, and Tomonori Hasegawa. Aha: Adhd augmented (learning environment). In *2018 IEEE International Conference on teaching, assessment, and learning for engineering (TALE)*, pages 774–777. IEEE, 2018.
54. Tasneem Khan, Kevin Johnston, and Jacques Ophoff. The impact of an augmented reality application on learning motivation of students. *Advances in Human-Computer Interaction*, 2019:1–14, 02 2019.
55. Saad Alqithami, Musaad Alzahrani, Abdulkareem Alzahrani, and Ahmed Mostafa. Ar-therapist: Design and simulation of an ar-game environment as a CBT for patients with ADHD. *CoRR*, abs/2005.02189, 2020.
56. Neha U Keshav, Kevin Vogt-Lowell, Arshya Vahabzadeh, and Ned T Sahin. Digital attention-related augmented-reality game: significant correlation between student game performance and validated clinical measures of attention-deficit/hyperactivity disorder (adhd). *Children*, 6(6):72, 05 2019.
57. Katherine Wang, Bingqing Zhang, and Youngjun Cho. Using mobile augmented reality to improve attention in adults with autism spectrum disorder. 04 2020.

ISBN 978 -987-633-574-4

Generación de mapas de calor de un partido de básquetbol a partir del procesamiento de video

Jimena Bourlot, Gerónimo Eberle, Eric Priemer, Enzo Ferrante, César Martínez, Enrique M. Albornoz
Instituto de investigación en señales, sistemas e inteligencia computacional, sinc(i)
UNL-CONICET, Ciudad Universitaria, Ruta Nac. Nº 168, km 472.4, (3000) Santa Fe
jimebourlot@gmail.com, geroo_ebeerle_11@hotmail.com, ericpriemer@yahoo.com
{eferrante,cmartinez, emalbornoz}@sinc.unl.edu.ar

Resumen.
El seguimiento de los jugadores en partidos de básquetbol para la obtención de estadísticas de los mismos representa actualmente una nueva fuente de información muy útil a la hora de la preparación de los partidos de un equipo en la temporada. En este trabajo presentamos un método para obtener el mapa de calor de dos equipos mediante una combinación de una red neuronal profunda, a fin de obtener a los jugadores dentro de la cancha, y técnicas de procesamiento digital de imágenes para convertir cada frame del video en un imagen de vista aérea (superior y perpendicular a la cancha), para luego mostrar la posición de los jugadores en un instante de tiempo especificado. Como caso de estudio, aplicamos el método para analizar un partido de los Juegos Asiáticos masculinos 2010 entre India y Afganistán. El sistema describe consistentemente la distribución de los jugadores a lo largo del partido.

Palabras claves: mapas de calor, procesamiento de video, básquetbol, redes neuronales profundas.

1 Introducción

Dentro de los distintos deportes en equipo, una buena manera de interpretar la forma de juego de cada uno es por medio de la distribución de sus jugadores en la cancha, que usualmente se observa por medios de mapa de calor que permiten visualizar las zonas donde hubo más presencias que en otras. A partir de esto, surge la motivación de poder generar mapas de calor de un partido de básquetbol televisado. Actualmente, la forma de realizar este trabajo es con costosos sensores que mediante GPS o acelerómetros miden la trayectoria del jugador en todo el partido. Estos sensores se

ISBN 978 -987-633-574-4

colocan en distintas partes del cuerpo para realizar el seguimiento [11], siendo algunos molestos para el atleta en su actividad. Algunos ejemplos son los sensores que se ubican como fajas en ciertas zonas como la rodilla [13], chips que se adhieren al cuerpo del deportista [12], o incluso insertos en protectores bucales [7]. En la Figura 1 se pueden ver imágenes de algunos de estos dispositivos. En este trabajo se pretende realizar una tarea de registro de ubicaciones utilizando únicamente la transmisión en video. Esto podría permitir que cualquier equipo, que no disponga de los recursos económicos para adquirir otra tecnología, pueda hacerlo de manera accesible, y además, es un sistema claramente no invasivo ya que no se debe "vestir" ningún sensor durante la actividad.

Figura 1. Sensores utilizados para realizar el seguimiento de personas en distintas disciplinas.

Habitualmente, las cámaras de televisión no realizan un paneo completo del campo de juego, sino que se capturan distintos sectores a partir de una rotación de las mismas. Es por esto que para realizar un mapa de calor se requiere como primer etapa la registración de los distintos frames del video, y así es posible generar una imagen panorámica que permita visualizar la cancha completa. Luego, se debe transformar el resultado en una vista perpendicular aérea, donde la visualización de los mapas de calor de los jugadores sea más sencilla.

La detección de objetos en imágenes es un campo muy incipiente, sobre todo con la llegada de modelos robustos de redes neuronales profundas [10]. Estos modelos permiten identificar objetos dentro de imágenes lo cual nos resultará útil para encontrar los jugadores y así obtener los datos que buscamos.

A continuación se describen los objetivos planteados en este trabajo, luego se introduce la metodología y los métodos utilizados y finalmente, se presentan los resultados y conclusiones.

2 Objetivos

ISBN 978 -987-633-574-4

El objetivo general es desarrollar un sistema que permita la generación de un mapa de calor de la distribución de los jugadores en un partido de básquetbol. En primer lugar se pretende definir la localización y características esperadas para el registro del video. Posteriormente, estudiar e implementar métodos de registración de imágenes para obtener una imagen de cancha entera. Luego, se procederá a estudiar e implementar rutinas para la identificación de jugadores en la cancha. Finalmente, se generarán mapas de calor a partir de las localizaciones de los jugadores en el video analizado.

3 Metodología

Se diseñó el sistema en base a una serie de procesamientos en bloques sucesivos. La Figura 2 muestra un esquema general del sistema.

Figura 2. Diagrama de bloques del sistema.

La entrada es un video que tiene una resolución de 600x400 píxeles, tomados a 30 frames por segundo. El primer paso es registrar estos frames para obtener una imagen panorámica de la cancha completa. Al realizar la registración, se obtiene la transformación que se realiza a cada frame del video y esto es útil para trasladar la posición de los jugadores del frame original a la vista panorámica en la cancha a cada instante de tiempo. Para una mejor interpretabilidad, se realiza otra transformación que lleva estos resultados a una vista aérea.

3.1 Generación de Panorámica a partir de la registración de frames

Para generar la vista de la cancha sobre la cual se mostrarán los resultados, en primer lugar se obtiene una vista panorámica a partir del video de entrada [2]. Para esto, se utilizó un algoritmo de registración de imágenes que consta en una instancia de identificación de descriptores y puntos claves mediante el *ORB detector* [5] y luego se realiza una búsqueda de correlación entre estos puntos utilizando el *BF Matcher* [5]. A partir de la obtención de los puntos relevantes entre dos frames sucesivos, es posible obtener la transformación que desplaza el frame de la vista original a la vista panorámica, utilizando un modelo de estimación lineal *RANSAC* [4]. Todos los

ISBN 978 -987-633-574-4

métodos utilizados se encuentran implementados en la librería *OpenCV* [8]. El resultado final de la registración de todos los frames del video puede verse en la Figura 3. La información calculada en este punto será relevante posteriormente para hallar la posición de los jugadores en esta vista.

Figura 3. Vista panorámica generada por la registración de los frames del video.

3.2 Identificación de jugadores y segmentación por equipos

En esta etapa se utiliza el algoritmo de detección de objetos *YOLO* [10]. El mismo es un algoritmo basado en redes neuronales profundas entrenada para segmentar una gran variedad de objetos, entre ellos animales, pelotas, personas, entre otros. Esta red es una de las más usadas ya que mediante una sola iteración de la imagen puede realizar una segmentación muy precisa de la misma, lo cual hace que se pueda implementar tanto offline como en tiempo real. Este algoritmo fue entrenado con la librería *Tensorflow* en lenguaje Python [1]. Los parámetros necesarios fueron modificados de forma tal que se obtienen sólo las personas presentes en cada frame. Luego, se segmenta la cancha para separar jugadores del público, obteniendo así los jugadores de cualquiera de los equipos, y/o los árbitros. Finalmente, se realiza un post-procesamiento que permite identificar a qué equipo pertenece cada jugador, y además descartar a los árbitros. Para tal fin, se utilizan segmentaciones utilizando distintos modelos de color [9]. El resultado de la segmentación puede verse en la Figura 4. En la misma se encuentran encuadrados con distintos colores los jugadores identificados de cada uno de los equipos, y en rojo los árbitros.

ISBN 978 -987-633-574-4

Figura 4. Resultado de la identificación de jugadores (recuadros azules y amarillos) y árbitros (recuadros rojos).

3.3 Transformación a vista aérea y formación de mapas de calor

Se pretende mostrar el mapa de calor en una vista aérea (superior y perpendicular a la cancha), ya que de esta forma se pueden ver los lados fuertes de cada jugador, por dónde realizan las jugadas principales los equipos, etc. Para ello, se realiza una transformación a la vista panorámica de la cancha. Esto se logra ubicando los puntos extremos de la cancha en la vista panorámica, y los puntos de destino que respetan las proporciones estándar de una cancha de básquetbol. Se utiliza para este proceso una matriz de transformación de perspectiva [15]. El resultado es una imagen aérea de la cancha sobre la cual se visualizarán los mapas de calor del partido, tal como puede verse en la Figura 5.

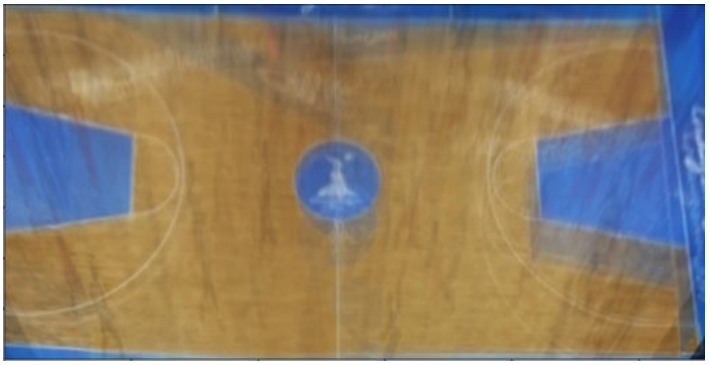


Figura 5. Vista aérea de la cancha.

ISBN 978 -987-633-574-4

Dado que la transformación se realiza sobre una imagen panorámica, construida a partir del promediado de un conjunto de reconstrucciones parciales obtenidas por la registración de los distintos frames del video, al realizar la transformación a la vista superior se vuelve notable que los jugadores no fueron eliminados por completo de la cancha. Si bien, se obtiene un resultado que se aproxima a una vista superior de una cancha de Básquetbol, tiene algunos pequeños errores que podrían mejorarse con algunas técnicas de reducción del ruido (jugadores que aún aparecen parcialmente), y mejorando el sistema de detección de los bordes de la cancha. Es por esto que se puede optar por una representación esquemática de la cancha, como la que se observa en la Figura 6, y así se logra una visualización más limpia de la información relevante, es decir, los mapas de calor del partido. La primera opción podría ser de utilidad para realizar un procesado de una transmisión televisiva, mientras la segunda tendría más utilidad para un entrenador que analice el partido.

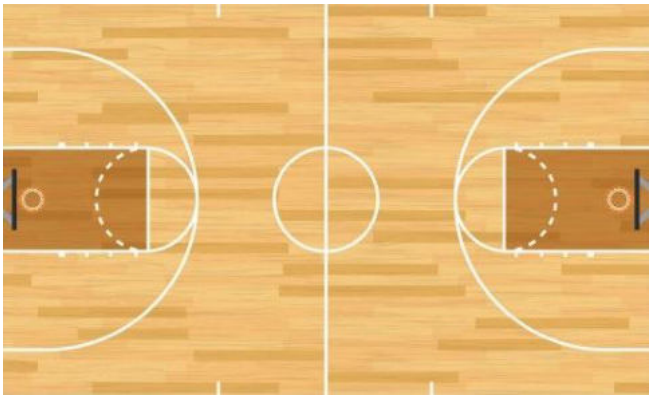


Figura 6. Representación esquemática de una cancha de básquetbol.

Una vez obtenidas las coordenadas que representan la ubicación de cada uno de los jugadores en la cancha, se procede a realizar las dos transformaciones de forma sucesiva.

La cancha se divide en celdas no solapadas de 50 cm. de lado [3]. Con esta discretización se define una matriz para cada equipo, cuyos valores contendrán la cantidad de jugadores detectados dentro de la celda correspondiente. Posteriormente, con esta información se calculan los mapas de calor de cada uno de los equipos utilizando la librería *seaborn* [14] que es utilizada para la visualización de datos estadísticos. Esta es de código abierto, está basada en *matplotlib* [6] e implementada en *Python*, y cuenta con varios modelos que permiten realizar mapas de calor con distintas variantes.

ISBN 978 -987-633-574-4

4 Resultados y Conclusiones

En este trabajo se ha desarrollado un método que permite, a partir de un video de trasmisión televisiva, obtener mapas de calor de cada uno de los equipos que juegan un partido de básquetbol. Una salida del sistema puede verse en la Figura 7. En la misma se observa el mapa de calor de uno de los equipos presentes en el video analizado. Las zonas más cálidas (rojas) representan aquellas donde el equipo tuvo más presencia de jugadores a lo largo del partido. Las referencias en los ejes de la izquierda e inferior están indicando la discretización en celdas mencionadas previamente.

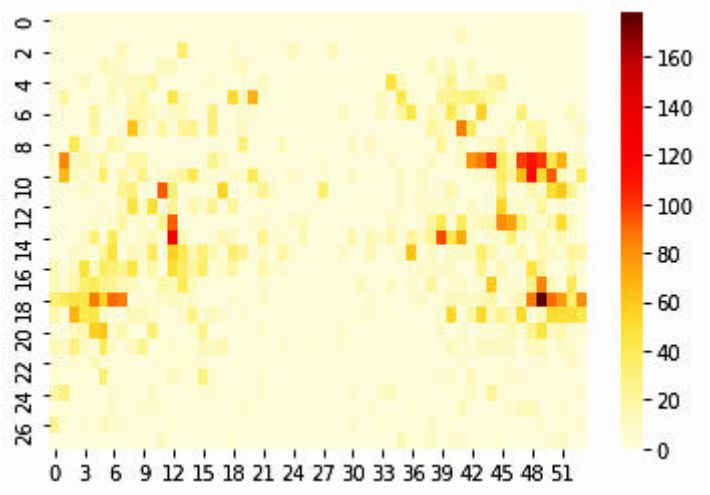


Figura 7. Mapas de calor para uno de los equipos del partido analizado.

El siguiente resultado es la representación del mapa de calor correspondiente a uno de los equipos, solapados con una cancha de básquetbol en vista superior (ver Figura 8). Este primer prototipo del sistema fue diseñado para funcionar con un video capturado con una cámara fija con rotaciones horizontales, como las utilizadas para la televisación en una vista central. La segmentación de equipos es parametrizable, dado que funciona en base a los colores de sus camisetas, y puede ser adaptado a cualquier equipo. De igual manera, la cancha puede ser parametrizada y permite la aplicabilidad en otras canchas.

ISBN 978 -987-633-574-4

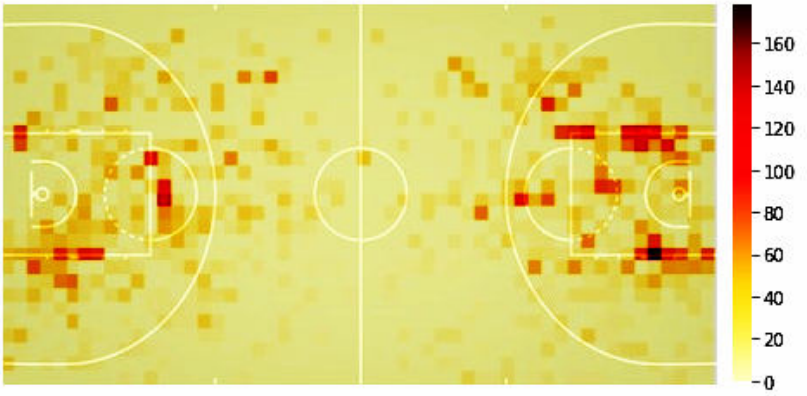


Figura 8. Mapas de calor resultante para uno de los equipos analizados, visto sobre la cancha.

Como trabajo futuro se prevé evaluar este método con videos de mayor resolución, ya que es esperable que mejoren tanto el algoritmo de registración como el algoritmo de detección de personas. Además, se pretenden implementar filtros para mejorar los resultados visuales de la cancha.
Otra de las tareas a implementar es la mejora de la generación de la vista superior de la cancha, realizando para esto la corrección de errores que se discutieron previamente: la detección de los bordes del campo de juego y la reducción del ruido presente debido a la presencia de sombras de los jugadores. A mediano plazo, se propone incorporar algoritmos de tracking que permitan el seguimiento de cada jugador para realizar estadísticas personalizadas.

Agradecimientos

Los autores desean agradecer al instituto sinc(i) UNL-CONICET, a UNL (con CAI+D 50620190100145LI)

Referencias

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. In the 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16) (pp. 265-283).
2. Chen, K., & Wang, M. (2014, July). Image stitching algorithm research based on OpenCV. In Proceedings of the 33rd Chinese Control Conference (pp. 7292-7297). IEEE.
3. Cheshire, E., Halasz, C., & Perin, J. K. (2013). Player tracking and analysis of basketball plays. In European Conference of Computer Vision.

4. Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381-395.
5. Howse, J., & Minichino, J. (2020). Learning OpenCV 4 Computer Vision with Python 3: Get to grips with tools, techniques, and algorithms for computer vision and machine learning. Packt Publishing Ltd.
6. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(03), 90-95.
7. Kim, J., Imani, S., de Araujo, W. R., Warchall, J., Valdés-Ramírez, G., Paixão, T. R., ... & Wang, J. (2015). Wearable salivary uric acid mouthguard biosensor with integrated wireless electronics. Biosensors and Bioelectronics, 74, 1061-1068.
8. Laganière, R. (2017). OpenCV 3 Computer Vision Application Programming Cookbook. Packt Publishing Ltd.
9. Lu, W. L., Ting, J. A., Little, J. J., & Murphy, K. P. (2013). Learning to track and identify players from broadcast sports videos. IEEE transactions on pattern analysis and machine intelligence, 35(7), 1704-1716.
10. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
11. Seshadri, D. R., Li, R. T., Voos, J. E., Rowbottom, J. R., Alfes, C. M., Zorman, C. A., & Drummond, C. K. (2019). Wearable sensors for monitoring the physiological and biochemical profile of the athlete. NPJ digital medicine, 2(1), 1-16.
12. Sekine, Y. et al. A fluorometric skin-interfaced microfluidic device and smartphone imaging module for in situ quantitative analysis of sweat chemistry. Lab. Chip 18, 2178–2186 (2018).
13. Stetter, B. J., Ringhof, S., Krafft, F. C., Sell, S., & Stein, T. (2019). Estimation of knee joint forces in sport movements using wearable sensors and machine learning. Sensors, 19(17), 3690.
14. Waskom, M. L. (2021). Seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021.
15. Wen, P. C., Cheng, W. C., Wang, Y. S., Chu, H. K., Tang, N. C., & Liao, H. Y. M. (2015). Court reconstruction for camera calibration in broadcast basketball videos. IEEE transactions on visualization and computer graphics, 22(5), 1517-1526.

ISBN 978 -987-633-574-4

Detección de la calidad del agua mediante imágenes satelitales: Revisión Sistemática de la literatura con análisis cuantitativo.

M. Silvia Vera Laceiras[1], Horacio Kuna[1], Norcelo G. De Miranda[1], Miryan Puchini[1], Eduardo Zamudio[1],

[1]Instituto de Investigación, Desarrollo e Innovación en Informática, Facultad de Ciencias Exactas, Químicas y Naturales, Universidad Nacional de Misiones
{vlhsilvia, eduardozamudi, hdkuna}@fceqyn.unam.edu.ar
{norcelodemiranda, puchinirm}@gmail.com

Resumen. Este artículo tiene como objetivo encontrar los estudios adecuados para realizar una RSL cuantitativa, sobre ´Detección de calidad de agua mediante imágenes obtenidas a través de teledetección satelital¨. Como resultado otorga un número de estudios que ofrecen una perspectiva representativa del conjunto de publicaciones sobre el tema respondiendo a preguntas y subpreguntas de investigación diseñadas por los investigadores. El resultado permite un conocimiento de la terminología básica utilizada, y los aspectos relevantes sobre la calidad del agua y su relación con algas tóxicas, microplásticos y sólidos en suspensión.

Palabras claves: Metaanálisis, calidad de agua, imágenes satelitales, procesamiento de imágenes, algas tóxicas, sólidos en suspensión, microplásticos.

1 Introducción

Los Objetivos de Desarrollo Sostenible (ODS) son un llamado urgente a la acción de los países para preservar nuestros océanos y bosques, reducir la desigualdad y estimular el crecimiento económico. Los ODS de gestión del agua exigen un seguimiento constante de las métricas de cobertura de la calidad del agua. Estas métricas aseguran "Garantizar la disponibilidad de agua y su gestión sostenible y el saneamiento para todos" [1] a través de acciones concretas como que de aquí a 2030, se propone mejorar la calidad del agua reduciendo la contaminación, eliminando el vertimiento y minimizando la emisión de productos químicos y materiales peligrosos, reduciendo a la mitad el porcentaje de aguas residuales sin tratar y aumentando considerablemente el reciclado y la reutilización sin riesgos a nivel mundial.
El tema de estudio es la calidad del agua a través de la observación y medición de sus contaminantes, usando para este fin el análisis de datos obtenidos a través de teledetección satelital de imágenes.

El objetivo de este artículo es encontrar los estudios adecuados para realizar un metaanálisis o Revisión sistemática de la Literatura (RSL) cuantitativa de los estudios

ISBN 978 -987-633-574-4

obtenidos utilizando metodología estadística especializada. Respondiendo a la pregunta de investigación ¿Para qué aplicamos teledetección en calidad de agua? y a las subpreguntas de investigación ¿Qué indica la presencia de algas tóxicas con respecto a la calidad del agua?, ¿Qué indica la presencia de microplásticos con respecto a la calidad del agua?, ¿Qué índices se pueden generar a través de la teledetección de microplásticos o algas en los cursos de agua?

En la sección 2 Contexto se describe brevemente el proyecto que da soporte a la línea de investigación y la facultad de referencia.

En la sección 3 Línea de Investigación y desarrollo se explica el objetivo del artículo y cuál es el procedimiento implementado para obtenerlos resultados. En la sección 4 Análisis de resultados se comparten los resultados obtenidos con el procedimiento.

En la sección 5 Conclusiones se exponen los resultados obtenidos de las prácticas metodológicas respondiendo las preguntas y subpreguntas de investigación.

2 Contexto

Esta línea de investigación se desarrolla dentro del proyecto de investigación "CIENCIA DE DATOS COMO HERRAMIENTA DE SOPORTE EN LA GESTIÓN PÚBLICA DE CALIDAD DEL AGUA" número 16/Q1224-IDP del Instituto de Investigación, Desarrollo e Innovación en Informática (IIDI) de la Facultad de Ciencias Exactas, Químicas y Naturales (FCEQyN) de la Universidad Nacional de Misiones (UNaM).

3 Línea de investigación y desarrollo

Revisiones sistemáticas de la literatura (RSL), son aquellas que resumen y analizan la evidencia respecto de una pregunta específica en forma estructurada, explícita y sistemática. Típicamente, se explica el método utilizado para encontrar, seleccionar, analizar y sintetizar la evidencia presentada. Existen 2 tipos de revisiones sistemáticas de la literatura, RSL Cualitativas: Cuando se presenta la evidencia en forma descriptiva, sin análisis estadístico. Y RSL Cuantitativas o Metaanálisis: Cuando mediante el uso de técnicas estadísticas, se combinan cuantitativamente los resultados en un sólo estimador puntual. Diferentes tipos de revisiones sirven para diferentes propósitos.

El metaanálisis de los resultados permite resumir en un solo valor numérico toda la evidencia relacionada con un tema puntual, aumentando la potencia estadística y la precisión del estimador puntual [2] Y se refiere al análisis estadístico de los datos de estudios primarios independientes enfocados en la misma pregunta, que tiene como objetivo generar una estimación cuantitativa del fenómeno estudiado, por ejemplo, la efectividad de la intervención [3]. Dichos análisis son esencialmente observacionales y utilizan estudios como unidad de investigación. Si bien puede ser controversial, pues si los estudios seleccionados tienen algún sesgo también lo tendrá la conclusión, la fuerza del metaanálisis radica en la capacidad de resumir un gran volumen de literatura en una sola publicación y producir conclusiones relevantes. El metaanálisis fue creado por Glass 1976 para combinar los resultados de varios informes diferentes en un informe,

ISBN 978 -987-633-574-4

para crear una estimación única y más precisa de un efecto [4]. Los análisis estadísticos en un metaanálisis están guiados por un modelo estadístico que debe asumirse previamente, para responder a preguntas planteadas con antelación La tarea principal del modelo estadístico es establecer las propiedades de la población del tamaño del efecto a partir de la cual se estima el tamaño del efecto individual [4,5].

Como objetivo del metaanálisis se puede "aumentar el poder estadístico; lidiar con la controversia cuando los estudios individuales no están de acuerdo; para mejorar las estimaciones del tamaño del efecto y para responder a nuevas preguntas no planteadas anteriormente en los estudios de componentes" [6,7]. También posee varias ventajas como que permite a los investigadores agrupar datos de muchos ensayos que son demasiado pequeños por sí mismos para permitir conclusiones seguras. Aunque idealmente cualquier ensayo clínico debería planificar un tamaño de muestra adecuado, históricamente la mayoría de los ensayos no han tenido el poder estadístico suficiente.

En 2002, un estudio de 5503 ensayos clínicos [8] identificó que el 69% tenía menos de 100 sujetos. Los ensayos pequeños hacen que sea más difícil rechazar la hipótesis nula porque conducen a desviaciones estándar y errores estándar más grandes. También existe el riesgo de sesgo. Un pequeño ensayo que no muestra un efecto significativo podría no ser enviado para su publicación, mientras que el ensayo del mismo tamaño que alcanzó significación (justificado o no) probablemente se publicará. [9].

En este caso las búsquedas se hicieron en los siguientes buscadores UPN (Universidad Privada del Norte, https://repositorio.upn.edu.pe/), RIA (Repositorio Institucional Abierto, https://ria.utn.edu.ar/), CIC (Comisión de Investigaciones Científicas, https://digital.cic.gba.gob.ar/), UNLP (Universidad Nacional de la Plata, http://sedici.unlp.edu.ar/), Researchgate (https://www.researchgate.net/), Ciencia Unisalle (https://ciencia.lasalle.edu.co/), UDEC(Universidad de Concepción Chile, http://repositorio.udec.cl/jspui/), Universidad EIA (https://repository.eia.edu.co/), Google scholar (https://scholar.google.com/), UNSA (Universidad Nacional de San Agustin, Peru, http://repositorio.unsa.edu.pe/), Universidad Cesar Vallejo (https://repositorio.ucv.edu.pe/), Universidad Privada San Carlos Puno (http://repositorio.upsc.edu.pe/), ACM (Association For Computer Machine, https://www.acm.org/), IEEE (Institute of Electrical and Electronics Engineers, https://www.ieee.org/), Colibri (Conocimiento Libre Reposiorio Institutcional , https://www.colibri.udelar.edu.uy/jspui/); Con una cadena de búsqueda en inglés "Remote Sensing and Water Quality Imaging" y español "Teledetección Satelital y Calidad de Agua".

La pregunta principal de la investigación que se aborda es: PI: "¿Para qué aplicamos teledetección en calidad de agua?", y las subpreguntas de investigación:

SPI: "¿Qué nos indica la presencia de algas tóxicas con respecto a la calidad del agua?"

SP2: "¿Qué nos indica la presencia de microplásticos con respecto a la calidad del agua?"

SP3: "¿Qué índices se pueden generar a través de la teledetección de microplásticos o algas en los cursos de agua?".

Definido como criterio de exclusión la antigüedad de publicación no mayor a tres años, los artículos anteriores a 2018 fueron excluidos y como criterio de selección de estudios se pondera la respuesta a 10(diez) preguntas definidas en una tabla con respuesta posible yes-no y partial que mediante la siguiente operación

matemática=CONTAR.SI(B12:K12;"Yes")+(CONTAR.SI(B12:K12;"Partial")/2) otorga un puntaje y se eligen los estudios que obtuvieron valores mayores o iguales a 4(cuatro).

4 Análisis de resultados

De 317.893 (trescientos diecisiete mil ochocientos noventa y tres) estudios encontrados de la siguiente manera en los buscadores: UPN con 24 estudios, RIA con 4.040 estudios, CIC con 49 estudios, UNLP con 14.785 estudios, Researchgate con 40 estudios, Ciencia Unisalle con 10 estudios, UDEC con 26 estudios, Universidad EIA con 7 estudios, Google Scholar con 5770 estudios, UNSA con 11028 estudios, Universidad César Vallejo con 33 estudios, Universidad Privada San Carlos Puno con 246 estudios , ACM con 281.161 e IEEE con 527 estudios.
Se procesan 33 (treinta y tres) que cumplen mayoritariamente con la cadena de búsqueda y dan respuesta a la pregunta principal y subpreguntas de investigación filtrados a través de los contenidos en el abstract o resumen.

La RSL Fig. (1) permitió diferenciar 33 (treinta y tres) documentos en inglés y español mediante la cadena de búsqueda, para luego de analizar si los estudios cumplían además de las preguntas y subpreguntas de investigación, pautas prefijadas como los criterios de inclusión y exclusión, y la valoración de calidad, se obtuvo un total de 7(siete) que son base para el desarrollo de éste artículo.

En la Fig. 1 se pueden observar algunos datos de la RSL sobre teledetección de imágenes satelitales y calidad de agua.

Titulo	Repositorio	Cadena de busqueda	AUTOR	AÑO	FUENTE
Optimización de puntos de control estelar en teledetección Rectificación geométrica de imágenes	ACM	remote sensing and water quality imaging	Xiangli Tan, Jungang Yang, Xinpu Deng	2017	ACM
Evaluación de la calidad de las imágenes de teledetección sin referencia basada en la región de interés y la similitud estructural	ACM	remote sensing and water quality imaging	Di Liu, Yingchun Li, Shaojun Chen	2018	ACM
investigación de la aplicación de la red neuronal convolucional en el registro de imágenes de teledetección	ACM	remote sensing and water quality imaging	Guohua Yue, Xiaoli Xing	2019	ACM
Un algoritmo mejorado de restauración de MTFC para la teledetección de imágene	ACM	remote sensing and water quality imaging	Yaqiong CHAI, Zhongkui FENG, Dongkai QI	2013	ACM
Evaluación de catástrofes con imágenes de teledetección de alta resolución basadas en el método de transferencia jerárquica de conocimientos	ACM	remote sensing and water quality imaging	Wen Dong , Zhanfeng Shen	2018	ACM
Mejora de la resolución de las imágenes para la teledetección Aplicaciones	ACM	remote sensing and water quality imaging	Chiman Kwan	2018	ACM
Segmentación de imágenes de teledetección multiespectral basada en el algoritmo de optimización de colonias de hormigas	ACM	remote sensing and water quality imaging	Shuo Liu, Yan-you Qiao, Qing-ke Wen	2009	ACM
Estudio sobre la variación de las aguas del lago y la fuerza motriz basada en imágenes de teledetección	ACM	remote sensing and water quality imaging	ZHONG Yanmei, LI Congzheng, ZHANG Wen	2017	ACM
UNA PRIMERA APROXIMACIÓN A LA MONITORIZACIÓN AUTOMÁTICA DE LA CALIDAD DEL AGUA DEL CASO II A PARTIR DE IMÁGENES DE SATÉLITE HJ-1	IEEE	remote sensing and water quality imaging	Yuanfeng Wu1,3, Bing Zhang1, Junsheng Li1, Hao Zhang2,3, Qian Shen2,3, Di Wu	2009	IEEE

Fig. 1. RSL sobre teledetección de imágenes satelitales y calidad de agua

En la Fig. 2 se observa el porcentaje de artículos totales publicados como resultado de una cadena de búsqueda. En el repositorio IEEE se encontraron 572 estudios que representan el 0,2% de artículos totales publicados, en ACM se encontraron 281.161 estudios que presentan el 88.4%, en Google Scholar se encontraron 5.770 estudios que representan el 1.8%, en UNLP se encontraron 14.785 estudios que representan el 4.7% y en UNSA se encontraron 11.028 estudios que representan el 3.5%.

ISBN 978 -987-633-574-4

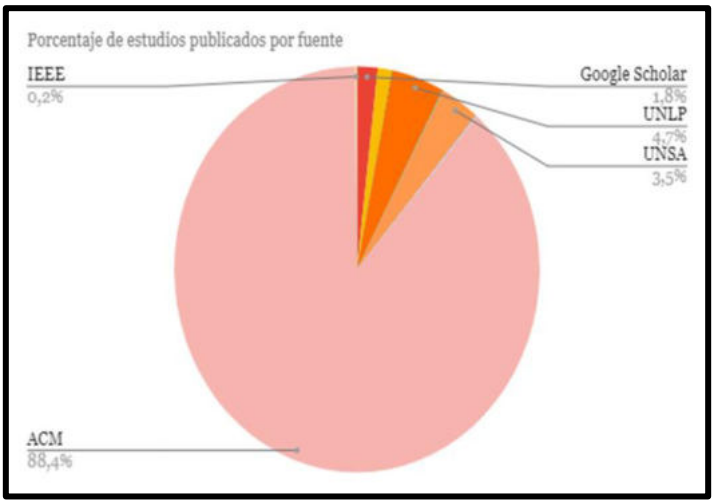


Fig. 2. Porcentaje de artículos publicados por fuente.

Se observa también la distribución de los estudios seleccionados de acuerdo al año de publicación Fig.3. De los estudios seleccionados, a los años 2019, 2017 corresponden a cada uno 2 estudios, a los años 2010, 2013, 2015, 2016, 2021, Al 2018 corresponden a cada uno 5 estudios, a los años 2019 corresponden 11 estudios, al año 2020 corresponden 8 estudios y a los años 2011, 2012 y 2014 corresponden 0 estudios.

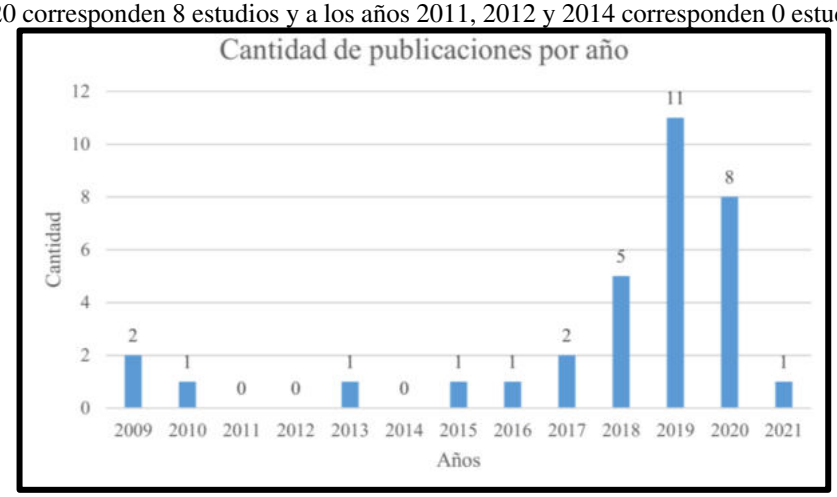


Fig. 3. Se puede observar la distribución de los estudios seleccionados, por año de publicación.

ISBN 978 -987-633-574-4

La Fig. 4 grafica la cantidad de estudios por tipo de publicación. Artículos de tésis tiene un total de 11 estudios, Artículos de conferencia tiene un total de 11 estudios y artículos

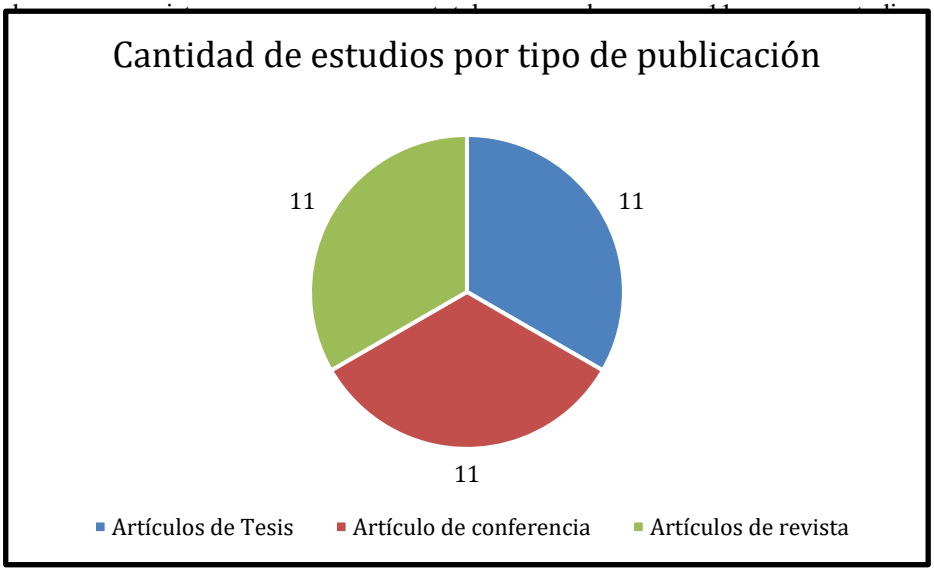


Fig. 4. Porcentaje por tipo de estudios

Ponderando luego la información de los resúmenes con un puntaje, obtenido valorando las respuestas a las 10 preguntas. Se decide como elegibles los documentos que obtengan 4 (cuatro) o más puntos y esos fueron los finalmente citados para lectura y análisis.

En el análisis se obtuvieron datos acerca de las características de los estudios, Cada estudio fue evaluado respondiendo a las siguientes preguntas:1-Los estudios mencionan ODS?, 2-El estudio está diseñado para alcanzar dichos objetivos?, 3-Responde todas las preguntas de investigación adecuadamente? 4-El estudio relaciona calidad de agua con microplásticos? 5-el estudio relaciona calidad de agua con algas tóxicas? 6-El estudio relaciona calidad de agua con sólidos en suspensión? 7-En el estudio se menciona extracción de datos de imágenes satelitales exportables? 8-En el estudio se menciona extracción de datos de drones? 9-En el estudio se presentan índices de calidad de agua? 10-El estudio refleja análisis de imágenes satelitales procesadas con software libre? Las respuestas contemplan tres posibilidades como mencionamos anteriormente, yes, no o partial y los resultados hacen a un documento elegible cuando obtiene un puntaje superior o igual a 4(cuatro).

ISBN 978 -987-633-574-4

	1. ¿Los estudios mencionan ODS?	2. ¿El estudio está diseñado para alcanzar dichos objetivos?	3. ¿Responde todas las preguntas de investigación adecuadamente?	4. ¿El estudio relaciona calidad de agua con microplásticos?	5. ¿El estudio relaciona calidad de agua con algas tóxicas?	6. ¿El estudio relaciona calidad de agua con sólidos en suspención?	7¿En el estudio se mencionan extracción de datos de imágenes satelitales exportables?	8. ¿En el estudio se menciona extracción de datos de drones?	9. ¿En el estudio se presentan índices de calidad de agua?	10. ¿El estudio refleja análisis de imágenes satelitales procesado con software libre?	Puntaje
Estudio multitem	-	Yes	Partial	-	-	Yes	Yes	-	Yes	Yes	5,5
Estimación de la	-	Partial	Partial	-	-	Yes	Yes	-	-	Yes	4
Teledetección de	-	Yes	Partial	-	-	Partial	Yes	-	Yes	Yes	5
Análisis del esta	-	Yes	Partial	-	-	-	Yes	-	Yes	-	3,5
USO DE IMÁGE	-	Yes	Partial	-	-	Yes	Yes	-	Yes	Yes	5,5
Optimización	-	-	Partial	-	-	-	Yes	-	Yes	Yes	3,5
Evaluación de	-	-	Partial	-	-	-	Yes	-	Yes		2,5
Investigación	-	-	Partial	Partial	Yes	Yes	Yes	-	Yes	Yes	6
Un algoritmo	-	-	Partial	Partial	Partial	Partial	Yes	-	Yes	Yes	5
Evaluación de	-	-	Partial	Partial	Partial	Partial	Yes	-	Yes	Yes	5
Mejora de la	-	-	Partial	Partial	Partial	Partial	Yes	-	Partial	Yes	4,5

Fig. 5. La imagen muestra cómo se pondera y asignan puntajes a los estudios encontrados. Imagen parcial del análisis metaanalítico de ponderación de estudios.

5 Conclusiones

Como fin de este estudio e inicio de una nueva de etapa de investigación, se toman de los documentos relevados y ponderados con más puntajes las ideas que nutren las conclusiones, otorgando un enfoque de claridad sobre el tema de teledetección satelital y su relación y efecto en el análisis de datos obtenidos de las imágenes, terminología y aspectos básicos de temas relevantes.

Con respecto a la aplicación de la teledetección en calidad de agua: Los estudios analizados permiten comprender y dimensionar la aplicación de la teledetección en la calidad del agua. Con la información obtenida a partir de las imágenes satelitales, se pueden analizar grandes superficies de la tierra, disminuyendo tiempos y costos operacionales. Esto incluye la posibilidad de investigar las cubiertas de agua, desde cuerpos pequeños hasta grandes masas oceánicas, considerándose como una alternativa eficaz para el estudio de dicho recurso natural. Este avance ha permitido un control y un conocimiento más ajustado de las condiciones atmosféricas, disminuyendo graves catástrofes naturales [10].

El análisis de los cursos de agua a través de teledetección se logra mediante “la interacción del flujo energético de los sensores con la superficie de la tierra” [7] hecho que recibe el nombre de radiación electromagnética. Esta interacción posibilita que “la adquisición de información por los sensores puede ser por reflexión, por emisión, y por emisión-reflexión” [11]. La adquisición de información por medio de los sensores, genera imágenes que pueden ser interpretadas mediante técnicas que permiten “analizar diversas variables biofísicas como la clorofila-a (Chl-a) y los sólidos totales en suspensión (SS), los cuales son de importancia para la calidad del agua” [12].

El avance en las técnicas de teledetección está facilitando este tipo de estudios debido a la mayor disponibilidad de imágenes y al gran desarrollo de nuevas tecnologías. Con estos métodos es posible obtener grandes cantidades de información con una resolución temporal, radiométrica y espacial elevadas, a un coste menor que con los métodos convencionales in situ [13]

Se puede observar que las imágenes producto de la interacción del flujo energético de los sensores con la superficie de la tierra brindan información de grandes áreas a un menor coste, además, la posibilidad de monitorearlas frecuentemente debido a que los satélites proporcionan datos que se encuentran en la web de forma gratuita.

ISBN 978 -987-633-574-4

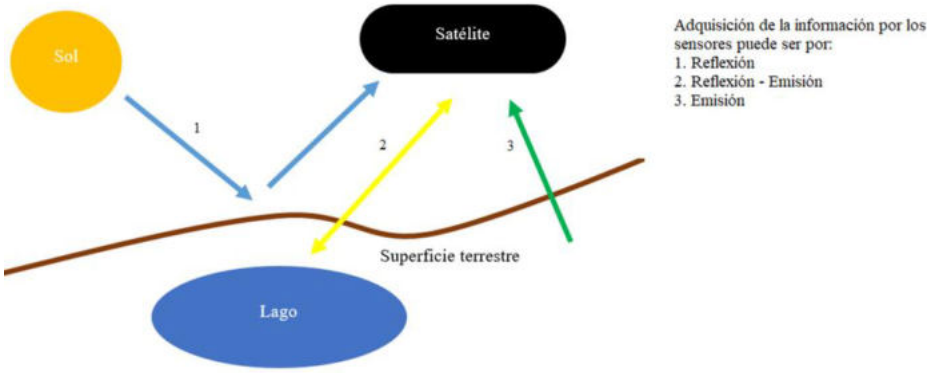


Fig. 6. Esquema de las formas de percepción remota. Adaptado de [8].

.

El monitoreo de vertimientos generados en la minería de oro requiere de estudio e innovación, pues los monitoreos realizados actualmente por las entidades competentes del recurso hídrico no proveen mediciones continuas ni de áreas extensas debido a la dificultad que se presenta en tiempo y dinero para realizar mediciones manuales o automáticas. Por tal motivo, es necesario que se cuente con nuevas herramientas para el monitoreo y el control de los vertimientos en aguas superficiales realizados por minas, cómo las técnicas de percepción remota [14].

Calidad de agua: Calidad de agua se refiere al conjunto de parámetros indicadores del estado del agua para ser usada con ciertos propósitos. Es el grupo de concentraciones, especificaciones, sustancias orgánicas e inorgánicas y la composición de la biota encontrada en el cuerpo de agua analizado, cabe destacar que la calidad de agua se ve afectada cuando el agua sufre cambios que afectan su uso real o propósito [11].

Entre los grupos de clasificación de parámetros de calidad, los parámetros físicos responden a los sentidos del tacto, olor y sabor, los parámetros químicos están relacionados con la solvencia del agua y los parámetros biológicos están asociados a la calidad del medio acuático, y se basan en los organismos que lo habitan [11].

Cuando los parámetros de calidad del agua no son adecuados para el uso real, podemos hablar de un deterioro. El deterioro de la calidad de las aguas superficiales se debe a la presencia de diversos tipos de contaminantes procedentes de actividades humanas como la agricultura, la industria, la construcción, la deforestación, etc. Así pues, la presencia de diversos contaminantes en las masas de agua puede conducir al deterioro tanto de la calidad de las aguas superficiales como de la vida acuática [14].

Análisis in situ vs Análisis a través de teledetección: El avance de la tecnología permitió la existencia de otra forma para analizar la calidad del agua mediante técnicas de teledetección. A diferencia de los análisis in situ, “que sólo pueden representar estimaciones puntuales de la calidad del agua en un tiempo y espacio determinados. Algunas limitaciones de este método consisten en que el muestreo y las mediciones requieren mano de obra, tiempo y acarrean grandes gastos; el estudio de grandes áreas es casi imposible y a esto se suma variaciones espaciales y temporales y de las tendencias que son difíciles de seguir; también la exactitud y la precisión de los datos in situ recogidos pueden ser cuestionables [15].

El monitoreo satelital ofrece una posible solución a los obstáculos de monitoreo in situ, ya que se recogen datos disponibles públicamente a escalas regionales y resoluciones temporales (es decir, repetición en el tiempo de la recolección) que son mucho más frecuentes que las campañas de muestreo de campo. La extracción de las mediciones de la calidad del agua directamente de las imágenes de satélite también puede permitir la rápida identificación de las aguas deterioradas, lo que podría dar lugar a respuestas más rápidas por parte de agencias de agua. [14]

Esto indica que los algoritmos para estudiar la clorofila que son considerados para aplicaciones a nivel global, no son apropiados para lagos y lagunas de pequeño tamaño. [15] Clorofila, presencia de algas en los cuerpos de agua: La medición de clorofila en el agua permite calcular la biomasa fitoplanctónica que constituye un indicador del estado trófico, que tiene conexión entre la concentración de nutrientes y la producción de algas. Para realizar esta medición existen grados de clasificación eutrófico como oligotrófico, mesotrófico, eutrófico e hipereutrófico [10]. El grado de eutrofia tiene como consecuencia la producción excesiva de algas, las cuales consumen oxígeno, aumentando la demanda bioquímica de oxígeno y producen la muerte de peces y animales [14]. Clorofila-a es el tipo que predomina en las algas y es un indicador de contaminación por nutrientes en los cuerpos de agua. La presencia de clorofila está relacionada con el fitoplancton, las algas determinan la estructura del ecosistema y pueden producir cambios físicos y químicos que conducen a la contaminación del agua de origen natural y antropogénico [11].

Finalmente, este estudio permite determinar las técnicas usadas, para que son usadas y qué beneficios se obtienen de su uso, proponiendo nuevos temas de interés para investigación que permitan hacer un uso eficaz de los recursos y tener datos e información para poder detectar rápidamente cambios en la calidad de agua medibles y estandarizables.

Como respuesta a las preguntas de investigación PI:“¿Para qué aplicamos teledetección en calidad de agua?”La teledetección posibilita analizar áreas pequeñas o extensas, terrestres o acuáticas, reducir el tiempo de análisis de dichas superficies y los costos de ese análisis, El uso de técnicas de teledetección facilita el análisis de parámetros como clorofila, sólidos totales en suspensión, microplásticos sin contacto físico, si bien la aplicación de teledetección no reemplaza por completo al método de campo, sino que, combinados permiten reducción de costos y tiempos.

Las evaluaciones con la incorporación de esta tecnología a nivel de parámetros de calidad se realizan mediante modelos matemáticos aplicados a los datos obtenidos de las imágenes satelitales, lo que nos permite un cálculo rápido, eficiente, económico y en tiempo real de los indicadores antes mencionados. La rápida identificación de aguas deterioradas hace posible determinar contaminantes de una fuente particular de contaminación o de múltiples fuentes, otorgando tanto al sector privado como público una herramienta de gestión y control de la calidad del agua, mediante la observación y tratamiento de una imagen satelital, que se obtiene de forma gratuita y de fácil acceso.

Bibliografía

1. Agua y saneamiento, https://www.un.org/sustainabledevelopment/es/water-and-sanitation/
2. Greenhalgh T. How to read a paper? Papers that summarize others papers (systematic reviews and metaanalyses). BMJ 1997; 315: 672-5.
3. J. Gopalakrishnan , P Ganeshkumar Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare, PMID: 24479036, PMCID: PMC3894019, DOI: 10.4103/2249-4863.109934
4. Glass, GV Primaria, secundaria y metaanálisis de la investigación. Investigador educativo, 5, 3-8.https://doi.org/10.3102/0013189X005010033, 1976
5. Xian Liu , Métodos y aplicaciones de análisis de datos longitudinales , 2016
6. Hunter, John E., & Schmidt, Frank L. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. Newbury Park, CA: Sage.1990.
7. Jiyuan Liu a,*, Mingliang Liu a,b, Hanqin Tian a,b, Dafang Zhuang a , Zengxiang Zhang c , Wen Zhang d , Xianming Tang a , Xiangzheng Deng, Spatial and temporal patterns of China's cropland during 1990–2000: An analysis based on Landsat TM data, 2000
8. Roderick P. McDonald and Moon-Ho Ringo Ho'Principles and Practice in Reporting StructuralEquation AnalysesUniversity of Illinois at Urbana–Champaign 2002
9. Stern, J.M. and Simes, R.J "Publication Bias: Evidence of Delayed Publication in a Cohort Study of Clinical Research Projects". BMJ, 13, 640-645. 1997 http://dx.doi.org/10.1136/bmj.315.7109.640
10. D. A. Gutiérrez Núñez, "Evaluación del uso de la teledetección para determinar parámetros de la calidad del agua en el embalse de la Planta Hidroeléctrica de Tacares", Trabajo de grado, Universidad de Costa Rica, 2019.
11. M. Escobar Valdivia, "Identificación de regiones contaminadas en la superficie del lago villarrica con base en imágenes sentinel en el periodo 2017-2018", Trabajo de grado, Universidad de Concepción, 2019.
12. "Estudio multitemporal de calidad del agua del embalse de Sitjar (Castelló, España) utilizando imágenes Sentinel-2", *Revista de teledetección*, vol. 56, n.º 117-130, octubre de 2020.
13. D. Uribe Ospina, "Estimación de la contaminación causada por la minería en cuerpos de agua del bajo cauca a través de imágenes satelitales", Trabajo de grado, UNIVERSIDAD EIA, 2019.
14. A. P. Cafa, "Monitorización de la calidad del agua de los lagos mediante técnicas de observación satelital", trabajo de especialización, ITBA, 2021.
15. D. C. Rivera Ruiz, "Estimación de parámetros de calidad de agua en la laguna santa elena usando imágenes satelitales", tesis maestría, Universidad de Concepción, 2020.
16. G. C. Piero Paolo & S. S. Renato Paolo, "Análisis del estado trófico mediante teledetección y datos "in situ" en la laguna de Paca, Jauja – Junín 2019", Trabajo de grado, Universidad César Vallejo, 2019.

CACIC 2021

WORKSHOP BASE DE DATOS Y MINERIA DE DATOS

COORDINADORES

Rodolfo Bertone(UNLP)
Hugo Alfonso (UNLPam)
Nora Reyes (UNSL)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978 -987-633-574-4

TreeSpark: A Distributed Tool for Progeny Analysis based on Spark

Paula López [1], Waldo Hasperué [1,2], Facundo Quiroga [1] and Franco Ronchetti [1,3]

[1] Instituto de Investigación en Informática LIDI. Facultad de Informática. Universidad Nacional de La Plata
[2] Investigador asociado - Comisión de Investigaciones Científicas (CIC-PBA)
[3] Investigador asistente - Comisión de Investigaciones Científicas (CIC-PBA)
{pdlopez,whasperue,fquiroga,fronchetti}@lidi.info.unlp.edu.ar

Abstract. Progeny analyses are useful in biological sciences for various purposes, such as improving individuals in new generations or carrying out molecular analysis of the transmission of genetic characteristics. Analyzing these data by making comparisons between individuals of a generation with their offspring is not a trivial task, and increases in complexity as more and more generations are incorporated. In this article, we present TreeSpark, an open source tool to carry out progeny analysis and provides functionality that allows simple access to the information of the individuals and their relations both as progenitors and descendants. This tool is developed as a Python module, which in turn inherits the distributed processing features of Spark, allowing it to process large volumes of progeny information. TreeSpark is compared with other similar tools, finding TreeSpark much simpler to use.

Keywords: Spark, big data, progeny analysis, genealogy, analytics.

1 Introduction

Various biological sciences carry out progeny analyzes looking for different objectives with the goal of comparing some characteristic of an individual with that of their offspring. For example, when analyzing and monitoring cattle, studies are aimed at establishing the magnitude of the improvement in milk production in the new generations, and thus be able to estimate its possible association with reproductive indicators in the offspring [1].

Progeny analyses are carried out both in animal [2][3] and plant species [4][5][6]. There are different works that range from the manual selection of breed individuals aimed at producing better individuals in the new generations based on a given characteristic of interest [7], to the molecular analysis of the transmission of genetic properties [8]. To carry out these analyses, a database prepared for this purpose is required. Above all, this database should have kinship relationship information between two individuals (descendant-parent). Crossing the information of an individual with that of its progeny or its progenitors quickly becomes complicated if many generations are included in the analysis. Researchers usually lack the required expertise in programming to use scripts developed for this types of tasks.

In this article, we present TreeSpark, an open source tool that facilitates progeny analysis by introducing a working mechanism based on Spark. TreeSpark is a Python module that includes, as part of its API, a set of variables and functions that facilitate access to information on the progeny or progenitors of any individual analyzed. TreeSpark is developed on the Spark framework, so it can be used both on individual computers as well as in distributed environments.

The present work is organized as follows: In Section 2, the problem of accessing the progeny information of an individual is described in detail. Some current tools that allow progeny analysis are detailed in Section 3. In Section 4, the TreeSpark tool is described. In Section 5, TreeSpark is compared with other state-of-the-art developments, analyzing the code that each of these requires solving various problems. Finally, conclusions are presented in Section 6.

2 Progeny Tree

In progeny analyses, the evolution of an individual with respect to its parents and progeny is studied. In other words, the focus of interest is analyzing the evolution of a branch (or a tree) of genealogical descent. To carry out these analyses, kinship relationship information between individuals is needed. All individuals that are part of the database must have information about their progenitors. In studies where it is only important to know a single progenitor (mother or father in the case of sexual species), then the set of individuals make up what is known as a progeny tree.

In this work, individuals with no parent information are called "root individuals". "Leaf" individuals are those that do not have offspring and, following this same logic, an individual is said to be "parent" of another individual, called "child". Figure 1 shows an example of a progeny database and the corresponding progeny trees of the two "root" individuals in the database.

2.1 Building Progeny Trees

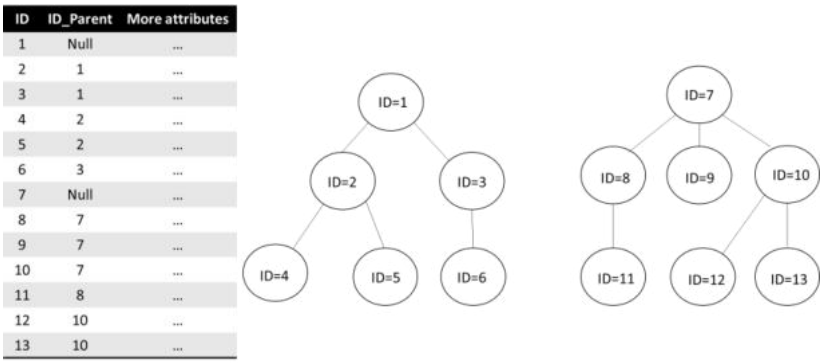
To carry out a progeny analysis for different generations of the same family from a dataset like the one shown in Figure 1, the following steps must be completed: 1) obtaining the root individuals; 2) obtaining the children from the root individuals; 3) obtaining the children of the individuals identified in the previous step; continuing recursively with this procedure until all the leaf individuals of each family are included.

In a database where progeny information is stored in a table like the one shown in Figure 1, which has the columns ID and ID_Parent, root individuals are obtained through a filter operation (SELECT), while for each generation that is to be included in the tree, a JOIN operation between the filtered result and the table of individuals must be performed. This operation is then repeated as many times as necessary until the entire family tree is formed. This way of working is inherent to the data that make up a tree structure. The following pseudo-SQL script allows obtaining the number of individuals in each family of the dataset.

ISBN 978 -987-633-574-4


```
Gen1 = SELECT ID FROM Table WHERE ID_Parent = Null
Gen2 = SELECT ID, Table.ID_Parent AS Family FROM Table
          INNER JOIN Gen1 ON Tabla.ID_Parent = Gen1.ID
Gen3 = SELECT ID, Family FROM Table INNER JOIN Gen2
          ON Table.ID_Parent = Gen2.ID
Res = SELECT Family, Count(Family) FROM Gen3
          GROUP BY Family
```

This script only allows working with three-generation trees like the ones shown in Figure 1. Developing a generic script that allows the treatment of *N* generations requires more sophisticated code, since a control structure of the WHILE type that evaluates some condition that detects if all individuals have been processed (allowing it to end the loop) is required.



ID	ID_Parent	More attributes
1	Null	...
2	1	...
3	1	...
4	2	...
5	2	...
6	3	...
7	Null	...
8	7	...
9	7	...
10	7	...
11	8	...
12	10	...
13	10	...

Fig. 1. On the left, a table with progeny information where the ID column (corresponding to the identifier of the individual) and the ID_Parent column (corresponding to the identifier of the parent individual) are highlighted. On the right, the graphic representation of the 13 individuals in the table, forming two independent trees or families.

3 Progeny Analysis Tools

Currently, there are several tools that allow analyzing progeny database. Some have the disadvantage of having a paid license, while others require a special pre-treatment of the data, and yet some others are outdated.

GraphFrames [9] and GraphLab [10] are two frameworks for treating graphs. GraphFrame is an integrated system that combines graphing algorithms, pattern matching, and relational queries. It is implemented on Spark SQL, it allows running processes in parallel, and it is compatible with the Spark dataframe API. To use it, data must be split into two tables: one with vertices (individuals) and another one with edges (kinship relationships).

On the other hand, GraphLab is a framework for machine learning written in C++ that has libraries for data transformation and manipulation, as well as model visualization. One of its various functionalities is to create graphs, with requirements similar to those of GraphFrames.

Both tools allow data to be loaded from various sources (JSON, CSV, etc.), but they only work with graphs in a generic way, i.e., they do not deal specifically with tree-shaped structures. Even though it is true that a tree is a particular type of directed graph, the functions that these tools provide, being graph-based, make it difficult to treat a tree-shaped graph. In order to work with these tools, at least two relationships (sets of edges) between individuals must be specified – parent → child and child → parent. If sibling information is also to be considered, then a third relationship has to be added between these individuals.

A previous version of the tool presented in this paper, [11] published a tool that allows using tree information to analyze progeny. This tool allows establishing ancestry and descent relationships between individuals, generating the progeny tree, providing functions to process their information. Even though the tool allows assembling and processing progeny trees, it does not support distributed execution.

Among commercial tools, ChromoSoft[1] and Breeders Assistant[2] stand out. Both work only with animal information and are designed especially for breeders. Even though they allow analyzing ancestors and descendants, in addition to calculating genetic or consanguinity coefficients, these tools support limited data formats, are tied to the payment of an annual membership, and cannot be run in distributed environments.

Finally, PedHunter[3] (which focuses on processing people information in large genealogies), PEDSYS [12] (which is designed to analyze individuals of any species), ENDOG [13] (which only focuses on analyzing information from animals), and InterHerd[4] (which is intended for dairy and meet cattle producers that wish to carry out progeny analyses, in addition to monitoring production, among other functionalities) are all tools that are currently outdated or obsolete.

4 TreeSpark

In this section, we introduce TreeSpark[5], an open source tool that allows progeny database analysis in Python using a simple and friendly syntax, as it provides variables and functions for this purpose.

The use of TreeSpark consists of, as a first stage, creating the progeny tree from a database and then, using several filtering operations “pruning” the family trees in the dataset based on the analysis to be carried out. For example, keeping individuals with more than three children, individuals that are "only children", individuals with a certain number of siblings, and so on, in addition to being able to use the data from the dataset as filters (days of longevity, milk production, number of eggs laid, etc.). Also,

[1] www.chromosoft.com/en
[2] www.tenset.co.uk/ba/
[3] www.ncbi.nlm.nih.gov/CBBresearch/Schaffer/pedhunter.html
[4] https://www.compuagro.net/interherd.htm
[5] https://gitlab.com/Danikawaii/treespark

TreeSpark allows obtaining dataframes for later analysis, made up of individuals with a given kinship relationship, such as all parents and their children, all siblings, etc.

4.1 Creating Family Trees

TreeSpark works with Spark dataframes; therefore, the first step is to create a dataframe by retrieving the data from any source supported by Spark. This dataframe must have at least two columns, as shown in the example in Figure 1: one with the identifier of the individual (ID) and the other with the identifier of the individual's parent (ID_parent). Optionally, the database can have a field that has the birth order of an individual. This must be a number that is interpreted as follows: 1 represents the first child, 2 represents the second child, and so forth. Having this information allows querying individuals sequences when they have a sibling relationship.

Once the dataframe is created, family trees must be assembled. This is done by creating the *TreeContext* object:

```
tc = TreeContext(DataFrame, "ID", "ID_PARENT", "ORDER")
```

where *DataFrame* is the dataframe with the progeny database retrieved from some data source, "ID" is the name of the column in the dataframe that stores the identification of the individuals, and "ID_PARENT" is the identifier of the parent individual. "ORDER", which is an optional parameter, is the name of the column that stores individual's birth order information. The object that represents all family trees is stored in the variable *tc*.

4.2 Filtering Family Trees

Once family trees are created, they can be "pruned" so that only those individuals that are of interest for a given analysis are retained (for example, individuals with more than four children or those that were born third), and use only the data corresponding them.

To carry out this task, TreeSpark provides a filter function called *filter* that allows “pruning” the trees. It is used as follows:

```
pruning1 = tc.filter(functionFilter1)
```

where *tc* is the *TreeContext* and *functionFilter1* is a function that will be evaluated for each of the individuals found in *tc*. *functionFilter1* is a function that takes all the information associated with an individual and returns a Boolean value. TreeSpark's *filter* function works similarly to Spark's *filter* function [14].

To simplify the code for the filters, TreeSpark incorporates special variables that refer to progeny information (Table 1) and are used with dot notation, as shown in the following examples.

```
fil1 = tc.filter(lambda ind: ind.childrenCount <= 3)
fil2 = tc.filter(lambda ind: ind.parentExists)
fil3 = tc.filter(lambda ind: ind.parent.parentExists)
```

ISBN 978 -987-633-574-4


```
fil4 = tc.filter(lambda ind: ind.childrenOrder == 1)
fil5 = tc.filter(lambda ind: ind.siblingsCount > 4)
```

In these filter functions, all the attributes found in the database can be accessed, as shown in the following example that selects all the individuals born in the year 2003:

```
pruning = tc.filter(lambda ind: ind["Year"] == 2003)
```

Pruning Results from Previous Prunings. The result of the filter function is an object that represents all the individuals that met the filter condition and therefore can be used to apply a new filter:

```
pruning2 = pruning1.filter(functionFilter2)
```

where *pruning1* is the result of a *filter* function and *functionFilter2* is another function with the characteristics mentioned above. Thus, the results from a previous “pruning” operations can be "pruned" again, and different pruning “paths” can be built as needed (Figure 2).

4.3 Lazy Evaluation

Since TreeSpark is developed using the RDDs API and Spark DataFrames, the evaluation of all the defined filters is not performed until some action is executed. Filters are not applied when the *filter* function is invoked, but the Spark RDD dependency graph (RDDs lineage) is generated internally. The graph is executed at the time of invoking any action [14]. The only action available in TreeSpark is *collect*, which retrieves all the information of the individuals that resulted from the filters applied. It is used as follows:

```
result = pruning.collect()
```

where *pruning* is the result of any previous *filter* function or the entire *TreeContext* itself. The result returned by *collect* is a Spark dataframe, or None if there are no results. The resulting dataframe will have one row for each individual that met all filter conditions, and it will also include the same columns as the original dataframe.

Table 1. Special variables that can be used in *filter* functions.

Variables	Typo of return	Description
parent	individual	Reference to the parent individual.
childrenCount	int	Number of children
siblingCount	int	Number of siblings
grandchildrenCount	int	Number of grandchildren
parentExists	bool	True if the individual has a parent individual
childrenOrder	int	Number representing sibling order
hasPrevSibling	bool	True if the individual is not the first child
hasNextSibling	bool	True if the individual is not the last child

ISBN 978 -987-633-574-4

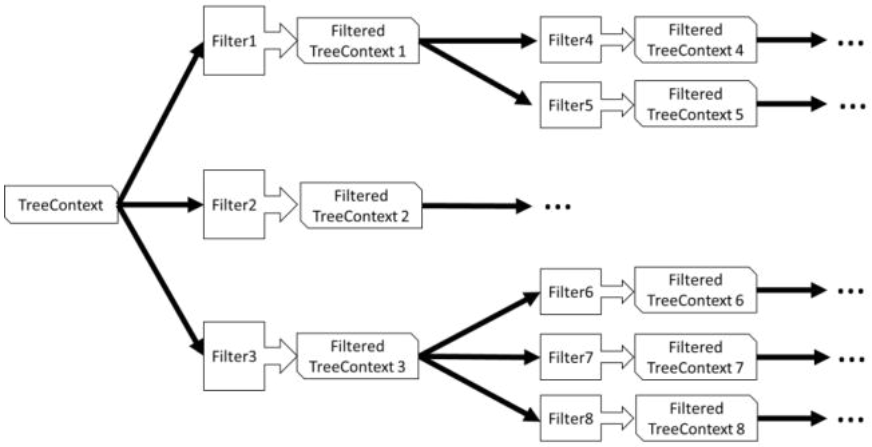


Fig. 2. Different filter "paths" stemming from a single TreeContext.

4.4 Obtaining Progeny Information

Using the results obtained after applying the filters, progeny relationships can be obtained for the resulting individuals. TreeSpark provides a set of functions that allow obtaining a collection with family relationships. These are three functions: *siblings, descendants* and *ascendants*.

The *siblings* function allows obtaining the relationships between an individual and its siblings. For example, if an individual *i,* obtained after applying a filter had three siblings h_1, h_2 and h_3 and they were born in the order h_1, i, h_2 and h_3, it would be possible to obtain the consecutive siblings in order of birth as follows:

```
result = pruning.siblings(2)
```

obtaining as a result the relations (h_1, i) and (i, h_2). The relationship (h_2, h_3) is not obtained as a result since, in this example, neither h_2 nor h_3 were the result of applying the filter. The relation (i, h_3) is not included either, because these are not consecutive siblings.

The value 2 used as a parameter in the *siblings* function indicates the number of individuals in the returned relationships. A value of 3 would return the relations (h_1, i, h_2) and (i, h_2, h_3). A value of 4 or greater would return the relationship (h_1, i, h_2, h_3). Similarly to the *siblings* function, TreeSpark provides the *descendants* function to obtain the descendants of an individual. It is used as follows:

```
result = pruning.descendants(2)
```

where the value of the parameter indicates the number of generations to be obtained. A value of 1 would only get the individual itself. With a value of 2, the function returns all children relationships; a value of 3, returns all children and grandchildren relationships, and so forth.

ISBN 978 -987-633-574-4

Finally, the tool provides the *ascendants* function, which allows obtaining the ancestors of an individual. Its use is similar to that of the *descendants* function:

```
result = pruning.ascendants(2)
```

where the value of the parameter indicates the number of generations to be obtained. Given the nature of the ancestry relationship, a relationship is obtained for each requested generation: (individual, parent), (individual, parent, grandparent), (individual, parent, grandparent, great-grandparent), etc.

5 Tool Comparison

In this section, the simplicity of the code that has to be written in TreeSpark to solve a progeny problem is analyzed. This comparison is made between TreeSpark and GraphFrames, since the latter is also a Spark-based tool.

As regards data sources, TreeSpark only needs a dataframe that contains the ID and ID_PARENT fields. On the other hand, GraphFrames requires a dataframe with the data of the vertices (the individuals) and another dataframe with the information of all the edges (parent-child relationships). For example, to get all parent-child relationships from the database in TreeContext, in TreeSpark, the following code must be run:

```
result = tc.descendants(2)
```

while in GraphFrames, the following statement must be executed:

```
result = graph.find("(n1)-[e]->(n2)")
```

where "(n1)-[e]->(n2)" is an expression that returns all edges *e* originating at node *n*1 (parent) and reaching node *n*2 (child). This way of retrieving relations from a graph becomes more complicated if we look for more complex relations such as grandparent-child-grandchild. In TreeContext, only *descendants (3)* is required, while in GraphFrames, the following statement has to be executed:

```
result = graph.find("(n1)-[e1]->(n2);(n2)-[e2]->(n3)")
```

Another example is obtaining sibling relationships. In TreeSpark, if the birth-order field is available, then it is a matter of simply executing the sentence:

```
df = tc.siblings(2)
```

while in GraphFrames, the parent-child relationship between vertices is required, as mentioned in Section 2. However, after this, the resulting graph must be converted to a DataFrame and a search for siblings using the DataFrame API (that is, externally to GraphFrame), must be carried out. Part of the code will be as follows:

```
g1 = graph.filterEdges("relationship = parent_child")
      v_df = g1.vertices() ; e_df = g1.edges()
all_df = v_vf.join(e_df, e_df("dst") === v_df("id"))
```

Then, to obtain the siblings of an individual, a search in the *all_df* dataframe using the dataframe API must be carried out. Alternatively, to carry out this search the edges corresponding to the relationships between siblings should be added to the graph, as follows:

```
g1 = graph.filterEdges("relationship = siblings")
       motifs = g1.find("(n1)-[e1]->(n2)")
```

As it can be seen, the GraphFrames code in this second example is simpler, but it requires adding more edges to the graph with the relationships between siblings. In the case of performing a search with a higher value, in TreeSpark the sentence and its complexity will remain the same, while in GraphFrames the situation will be similar to that of the descendants calculation.

An important point to note is that, even though in GraphFrames the vertex filter is supported through the *filterVertices* function, its execution not only filters individuals but also all their relationships (it eliminates those edges whose vertices no longer exist in the subgraph). This behavior means that, when applying the filter, the relationship between a child and its parent is lost if the condition provided is not met. In TreeSpark, on the other hand, the reference to the individual parent does not disappear regardless of the number of filters applied to the tree, meaning that information can be queried at any filtering instance.

6 Conclusions

We presented TreeSpark, a tool that facilitates progeny analysis through the use of specific variables and functions provided by the tool itself. TreeSpark inherits the simplicity of Python, hiding the complexity of an iterative process of multiple JOINs, thus allowing any researcher with little knowledge of Python to take advantage of all its functionality by simply searching for progenies to carry out their analyses.

TreeSpark is implemented on the Spark framework, and inherits two very important functionalities from it: On the one hand, it can retrieve data from various sources, since data must be retrieved through a DataFrame in order to use TreeSpark. On the other, and most importantly, the filters and operations carried out with TreeSpark can be executed in a distributed manner using a cluster of computers. If the database volume is very large, then TreeSpark can be run in a distributed manner. This processing is transparent to the user, since all distributed execution is carried out internally by Spark. TreeSpark tests have been carried out on a single node with a database of hundreds of individuals. As future work, we plan to study the performance of this tool in a cluster of nodes, considering how data distribution affects task execution performance. More complex filters than those shown in the examples included in this article are also

pending testing. Finally, it should be noted that TreeSpark is in development, meaning that it is still going through testing and debugging stages.

References

1. Rearte, R., LeBlanc, S. J., Corva, S. G., de la Sota, R. L., Lacau-Mengido, I. M., Giuliodori, M. J.: Effect of milk production on reproductive performance in dairy herds. Journal of Dairy Science 101(8), 7575–7584. doi: 10.3168/jds.2017-13796 (2018).
2. Lopera-Barrero, N. M., Vargas, L., Nardez-Sirol, R., Pereira-Ribeiro, R., Aparecido-Povh, J., Streit Jr, D. P., Cristina-Gomes, P.: Diversidad genética y contribución reproductiva de una progenie de *Brycon orbignyanus* en el sistema reproductivo seminatural, usando marcadores microsatélites. Agrociencia 44(2), 171-181 (2010)
3. Domínguez Viveros, J., Rodríguez Almeida, F. A., Núñez Domínguez, R., Ramírez Valverde, R., Ortega Gutierrez, J.A., Ruíz Flores, A.: Análisis del pedigrí y efectos de la consanguinidad en el comportamiento del ganado de lidia mexicano. Archivos de Zootecnia 59(225), 63-72 (2010)
4. Salomón, J. L, Castillo, J. G, Arzuaga, J. A, Torres, W, Caballero, A, Varela, M., Hernández Betancourt, V. M.: Análisis de la interacción progenie-ambiente con minitubérculos a partir de semilla sexual de papa (*Solanum tuberosum*, L.) en Cuba. Cultivos Tropicales 36(2), 83-89 (2015).
5. Kolvalsky, I. E., Solís Neffa, V. G.: Análisis de la progenie de individuos productores y no productores de gametos masculinos no reducidos de *Turnera sidoides* (Passifloraceae). Boletín de la Sociedad Argentina de Botánica 50(1), 23-33 (2015)
6. Gutiérrez Vázquez, B. N., Cornejo Oviedo, E. H., Zermeño González, A., Valencia Manzo, S., Mendoza Villarreal, R.: Conversión de un ensayo de progenies de *Pinus greggii* var. greggii a huerto semillero mediante eigen-análisis. Bosque (Valdivia) 31(1), 45-52 (2010)
7. Guitou, H. R, Monti, A., Sutz, G., Baluk, I.: Interpretación y uso correcto de las diferencias esperadas entre progenie (DEP´s) como herramienta de selección para la calidad de carne: Segunda parte. Revista Colombiana de Ciencias Pecuarias 20(3), 363-376 (2007)
8. Luévanos-Escareño, M. P., Reyes-Valdés, M. H., Villarreal-Quintanilla, J. Á., Rodríguez-Herrera, R.: Obtención de híbridos intergenéricos *Helianthus annuus* x *Tithonia rotundifolia* y su análisis morfológico y molecular. Acta botánica mexicana (90), 105-118 (2010)
9. Dave, A., Jindal, A., Li, L., Xin, R., Gonzalez, J., Zaharia, M.: GraphFrames: an integrated API for mixing graph and relational queries. 1-8. 10.1145/2960414.2960416 (2016).
10. Low, Y., Gonzalez, J., Kyrola, A., Bickson, D., Guestrin, C., Hellerstein, J.: GraphLab: A New Framework for Parallel Machine Learning. Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence, UAI 2010 (2010).
11. López, P., Hasperué, W., Rearte, R., de la Sota, R. L.: Herramienta informática para el análisis de progenie. Innovación y Desarrollo Tecnológico y Social 2(1), 35-54 (2020).
12. Dyke B.: PEDSYS: a pedigree data management system user's manual. San Antonio: Texas Southwest Foundation for Biomedical Research, Population Genetics Laboratory Technical Report No. 2. 1999;368 (1999).
13. Gutiérrez, J., Goyache, F.: A note on ENDOG: A computer program for analysing pedigree information. Journal of animal breeding and genetics. 122. 172-6. 10.1111/j.1439-0388.2005.00512.x (2005).
14. Scott, J. A.: Getting started with Apache Spark. MapR Technologies, Inc., San Jose, CA (2015)

ISBN 978 -987-633-574-4

Vehicular Flow Analysis Using Clusters

Gary Reyes[1] , Laura Lanzarini[2], César Estrebou[2] , Victor Maquilón[1]

[1]Facultad de Ciencias Matemáticas y Físicas, Universidad de Guayaquil,
Cdla. Universitaria Salvador Allende, Guayaquil 090514, Ecuador
{gary.reyesz,victor.maquilonc}@ug.edu.ec
[2]Universidad Nacional de La Plata, Facultad de Informática,
Instituto de Investigación en Informática LIDI (Centro CICPBA) 1900 La Plata,
Buenos Aires, Argentina
{laural,cesarest}@lidi.info.unlp.edu.ar

Abstract. The volume of vehicular traffic in large cities has increased in recent years, causing mobility problems, which is why the analysis of vehicular flow data becomes important for researchers. Intelligent Transportation Systems perform vehicle monitoring and control by collecting GPS trajectories, information that provides real-time geographic location of vehicles, which allows the identification of patterns on vehicle flow using clustering techniques. This paper presents a methodology capable of analyzing vehicular flow in a given area, identifying speed ranges and maintaining an updated interactive map that facilitates the identification of areas of possible traffic jams. The results obtained on a dataset from the city of Guayaquil-Ecuador are satisfactory and clearly represent the speed of vehicle displacement by automatically identifying the most representative ranges for each instant of time.

Keywords: vehicular flow, cluster, GPS trajectory

1 Introduction

Nowadays the constant increase in the volume of traffic in large cities causes problems in the vehicular flow, so the analysis of the data generated by the vehicle monitoring and control systems becomes relevant. Its study through descriptive techniques allows to identify relationships between vehicle trajectories facilitating the analysis of the flow of vehicles. They currently provide solutions in a variety of areas, such as health, finance, telecommunications, agriculture and transport, among others [1].

Data clustering is a technique widely used to identify common characteristics between instances of the same problem [2]. Over time, researchers have proposed improvements to the limitations identified in some techniques such as [3] where a correct initialization of the algorithm is achieved in a much shorter time. In other cases, techniques have been adapted to work in a specific context, such as for spatial data mining [4] [5] [6] or for GPS trajectory analysis [7]. A GPS trajectory is defined by a set of geographic locations each of which is represented by

ISBN 978 -987-633-574-4

its latitude and longitude, in an instant of time. This paper proposes a methodology for the analysis of vehicular flow in traffic through the analysis of GPS trajectories. For this purpose, each zone within an area of interest is characterized according to the average speed and the number of vehicles it contains, in a given period of time. The zones are delimited at the beginning of the process and their size depends on the precision with which the analysis is to be carried out. Then, using a variation of the dynamic clustering algorithm for data flows named Dyclee, originally defined by Barbosa et al. [8], the zones with similar characteristics are identified and an interactive map is constructed on which the ranges of speeds corresponding to the current vehicular flow and the zones where they occur can be observed. This methodology can be used, together with other tools, by traffic managers in a city to plan urban roads, detect critical points in traffic flow, identify anomalous situations, predict future mobility behavior, analyze vehicular flow, among others. The proposed methodology was used to characterize the data corresponding to GPS trajectories generated by a group of students from the University of Guayaquil, Ecuador. The obtained results allow to identify at different time instants, sectors of the city where vehicles have common speeds.

This article is organized as follows: section 2 analyzes some related work that were identified in the literature and present various solutions to the problem, section 3 describes the proposed methodology, section 4 presents the obtained results and finally, section 5 contains the conclusions and lines of future work.

2 Related work

Clustering techniques have been used in trajectory analysis for several years. They are usually adaptations of conventional algorithms using similarity metrics specially designed for trajectories [9] [10]. Such is the case of the Improved DBScan algorithm [11] which improves the traditional DBScan algorithm using its own density measurement method that suggests the new concept of motion capability and the introduction of data field theory. Another example is the Tra-DBScan algorithm [12], which uses the DBScan [13] algorithm adding a trajectory segmentation phase, in which it partitions the trajectories into sections and uses the Hausdorff distance as a similarity measure.

Yu et al. [14], an improved trajectory model is proposed and a new clustering algorithm is presented, with a similarity measure that calculates the distance between two trajectories based on multiple features of the data, achieving maximization of the similarity between them. On the other hand, Ferreira et al. [15] presents a new trajectory clustering technique that uses vector fields to represent the centers of the clusters and propose a definition of similarity between trajectories. A GPS vehicular trajectory clustering method using angular information to segment trajectories and a pivot-guided similarity function is presented by Reyes et al. [16]. Research efforts in this area continue today [17] [18].

In summary, it can be stated that clustering techniques have proven to perform well in the analysis of vehicular trajectories although their parameterization

ISBN 978 -987-633-574-4

remains an interesting challenge. This is related to the fact that they are unsupervised techniques that generally combine distance and density metrics to control the construction of the clusters.

This article used a dynamic clustering algorithm for data streams. This type of algorithms process data flows managing to overcome some of the limitations of traditional clustering algorithms, which usually iterate over the dataset more than once, causing greater memory usage and increasing execution time [19] [20]. As the distribution of the data in each stream changes continuously, it is important that these clustering algorithms that process data flows generate dynamic groups, where the number of groups depends on the distribution of the data of the flow [21] [22].

In particular, in this article a variation to the DyClee algorithm has been made, originally designed by Barbosa et al. [8]. It is a dynamic clustering algorithm for tracking evolving environments capable of adapting the clustering structure as the data is processed.

Dyclee uses a two-stage clustering approach [23]. The first stage consists of clustering the examples based on their similarity and density. The groups obtained as a result of this stage are called microclusters. Then, in the second stage, the microclusters are clustered starting with the densest ones and taking into account their overlapping and similarity in terms of their density.

3 Methodology

This paper proposes a methodology for the analysis of vehicular flow in traffic. This methodology is represented in the figure 1 and consists of three steps: the first step is to properly represent the data of the trajectories within the area of interest; the second step uses an adaptation of a dynamic clustering algorithm to identify relationships and the third step consists of creating interactive resources for the visualization of the results. Each step mentioned is described below.

Representation of vehicular flow

The first step is to provide an adequate representation of the data that make up the trajectories. To do this, first of all, the area of interest must be established. Here, it should be indicated which is the geographical area to which the trajectories to be analyzed belong. Once the area is established, it is partitioned into cells, or smaller zones, in a uniform manner. The size of each cell will depend on the precision with which it is desired to analyze the vehicular flow. In this work, 200 m2 cells were used. This is important data to take into account since the information to be analyzed corresponds to a summary of what is happening in each cell in a given period of time. The methodology proposed here consists of analyzing what is happening in each cell as a whole instead of considering each vehicle trajectory separately. This facilitates analysis and visualization.

In particular, in this article, the data corresponding to vehicular flow, represented in each cell, were analyzed in batch mode in 3-minute periods. However,

ISBN 978 -987-633-574-4

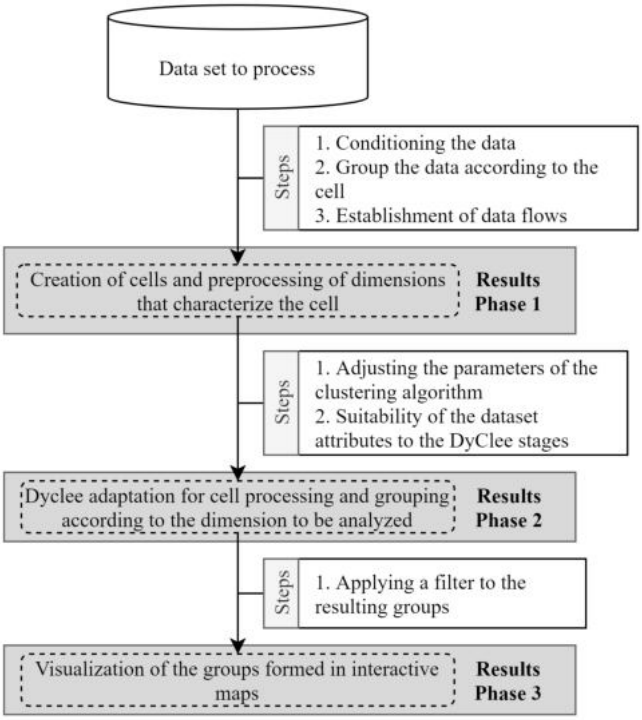


Fig. 1. Proposed methodology

if a specific analysis is desired, these periods can be shorter, for example one minute. Each period is considered an evolution since the DyClee algorithm updates the clustering with the incorporation of each block of data sequentially. In each evolution, a data flow will be entered to perform the respective calculations for each cell and to obtain characteristic information of these cells.

Adapted Dyclee algorithm

The second step uses an adaptation of the DyClee algorithm, defined in this article, to process the trajectories within each cell. The table 1 identifies the basic elements and parameters that were used in this adaptation.

As previously mentioned, the first of the two stages of the DyClee algorithm refers to the construction of microclusters. In this article, instead of using directly the GPS locations and their density, as proposed in the original article, the velocities of the trajectory sections included in each cell were considered. Thus, for a given time period, each cell will be represented by the average velocity of the trajectory sections it contains. This implies trimming the trajectories appropriately considering that speed variations may lead to a vehicle traveling at very high speed not being recorded (or having very few GPS locations) when

ISBN 978 -987-633-574-4

Table 1. Concepts and parameters associated with DyClee processing

Element	Definition
Microcluster	Represents the dataset with similar characteristics.
Hyperbox	Determines the area of the microcluster.
Relative size	Size of the microclusters with respect to the processing area.

passing through a cell. In addition, it is important to consider that speeds should be averaged considering the vehicles and not the number of locations recorded. Regarding the size of the microclusters, the value of the "relative size" parameter specifies the relative size of the "hyperbox" parameter concerning to the area to be processed. That is, as its value decreases, the number of microclusters increases and vice versa.

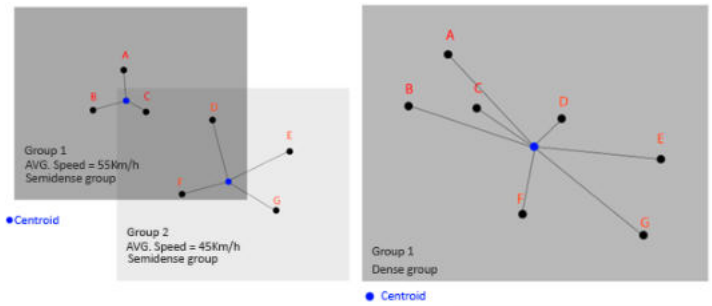


Fig. 2. Second stage operation of Dyclee algorithm. (A) Identification of directly connected microclusters and (B) Resulting expansion

In its second stage, the algorithm analyzes the densities of the microclusters formed and classifies them into two categories: dense and semi-dense. From the dense microclusters it starts to join those that are directly connected. Two microclusters will be considered directly connected if the maximum distance at which the centroids of two microclusters with similar densities can meet does not exceed the value of the "hyperbox" parameter. That is, as the value of the "hyperbox" parameter increases, the number of clusters will be smaller and therefore, the velocity ranges in the area of interest will have greater amplitude. Figure 2 illustrates the operation of this stage.

For each cluster completed in the log, a record is generated for each cell used in the clusters, which contains trajectory data within a period of analyzed time. Although each result is reflected in an interactive map as the analysis progresses, its registration allows reconstructing previous situations.

ISBN 978 -987-633-574-4

Visualization of grouping

With the result of each grouping, an interactive map is created in which the relevant information of each cell can be analyzed graphically and dynamically. Also, using the log mentioned in the previous section, it is possible to reconstruct all the maps from the beginning of the vehicular flow analysis. This provides a quick visualization of the traffic state. In the interface of each map there are layer selection controls and reference legends to interpret the results represented on the map. Two types of maps are generated: a map of the last evolution or time period analyzed and a map with all the evolutions.

In the map corresponding to a particular evolution, each cluster has been represented in a different layer and the user can select one or several layers of the map to filter the information of interest. To facilitate the visualization, a different color has been used in each layer; in this way, if more than one cluster is displayed simultaneously, it will be possible to distinguish to which of them the marked cells belong. The map also has the possibility to select the display of markers that show information of both the group and the selected cell. Figure 3 (B) illustrates the latter.

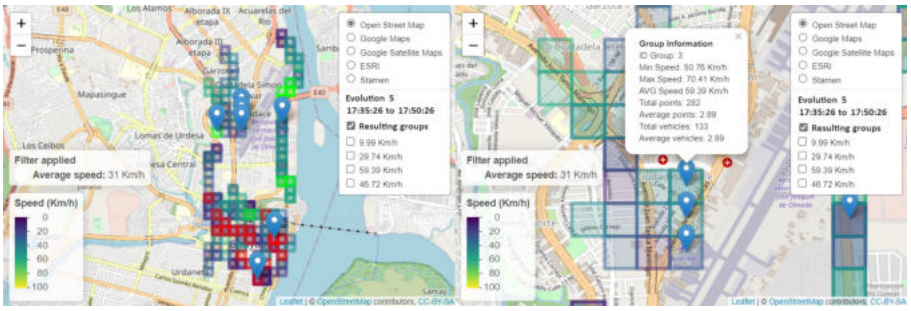


Fig. 3. Maps by evolution. (A) Layer containing the delimitation of the corresponding areas. (B) Layer on which certain markers are activated.

In the map of all the evolutions are visualized, in a single map all the evolutions made with a different color for each evolution, being able at some point to choose one or several evolutions according to the analysis to be carried out.

4 Results and Discussion

To test the performance of the methodology proposed in this article, we work with own data collected in the city of Guayaquil, Ecuador, on October 28, 2017. These data correspond to 218 trajectories made by university students traveling in some means of transportation such as cab, motorcycle and metrovia. The locations in this dataset were collected by smartphones with an average time interval

ISBN 978 -987-633-574-4

between two consecutive locations of 5 seconds. Each record contains trajectory id, latitude, longitude, time, user name, email and type of transportation.

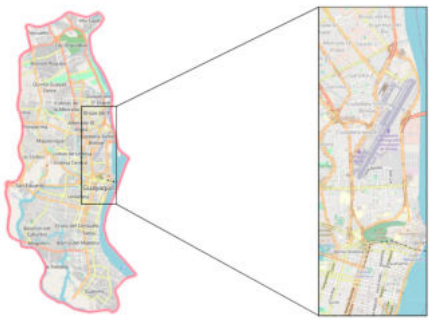


Fig. 4. Area representing the dataset for the city of Guayaquil

Given that this is a small set of trajectories, the analysis was carried out between 16:30 and 18:30 hours, as this is considered to be the time of greatest concentration of records. As a result of this filtering process, 30557 records were obtained, representing 206 trajectories of the entire data set. The area representing the selected data set is shown in figure 4.

The configuration of DyClee is done by means of the necessary parameters. The value for the "relative size" parameter was 0.2, from this value and based on the processing area, the value of the "hyperbox" dimension was defined. The 30557 records were divided into 8 blocks of 15 minutes each and analyzed consecutively. This 15-minute time period could be considered excessive, but its duration is in relation to the volume of data collected. It is important to consider that only the vehicular flow corresponding to the trajectories followed by the students who collaborated with the data collection is being analyzed. To know the vehicular flow of the city at that time, it is necessary to add the information of the rest of the vehicles that circulate in the area in that time range.

The information represented for each cell consists of the average speed of the vehicles registered inside the cell during the period of analyzed time. In Figure 5 the maps corresponding to the 8 evolutions carried out can be observed. Below each map, the information corresponding to the carried out grouping with the adaptation of the DyClee algorithm is indicated. For each group, the minimum, maximum and average speeds are indicated, as well as the deviation of the speeds belonging to that group. These values show the low overlap between them.

Figure 6 illustrates the evolution of the average speeds of the groups over time. There it can be seen that, although some change from one evolution to another, if analyzed in order, they form six common velocity ranges identified as velocity 0, 1, 2, 2, 3, 4 and 5.

ISBN 978 -987-633-574-4

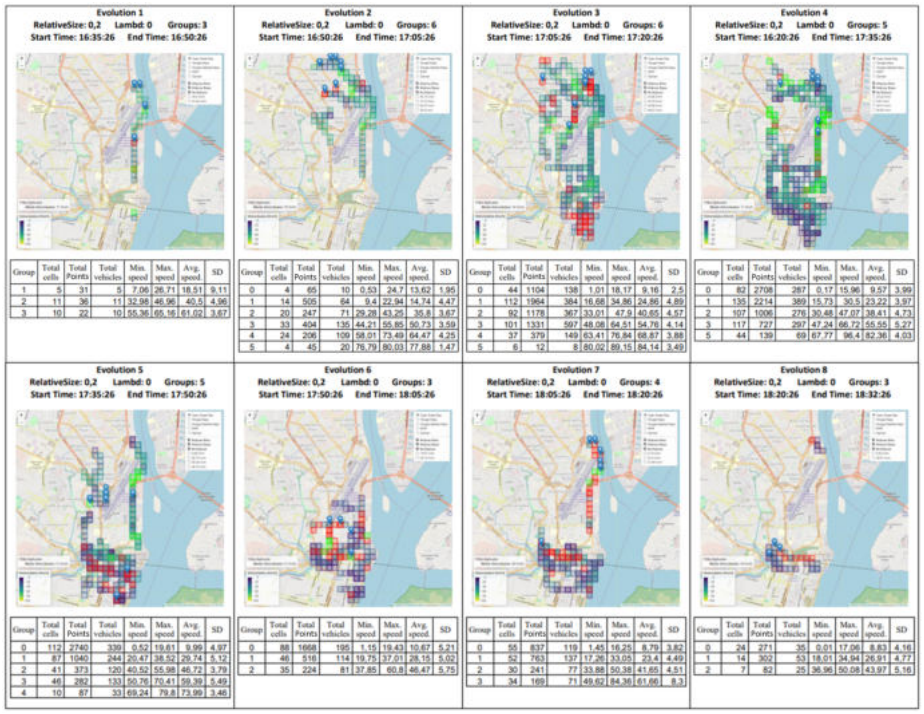


Fig. 5. Results obtained from the 8 experiments in consecutive time periods

In the case of speed rank 3 it only appears in evolutions 1, 2, 3, 4, 5 and 7 which evidences the dynamic characteristic of group conformation of the used algorithm. In addition, from evolution 3 onwards the 3 lowest speed ranges are maintained while the highest speeds are mainly concentrated between evolutions 2 to 5. This shows the speed variation that occurs in the vehicular flow over time identifying that in the first hour of the analysis the traffic reaches high speeds while in the last 45 minutes the trips are made at lower speeds. Figure 6 shows the average speeds in each group identifying the six speed ranges.

5 Conclusions

This article has proposed a methodology to identify, dynamically, the characteristics of the vehicular flow in a period of time. In order to do this, the information of the trajectories has been represented in cells and processed using an adaptation of the DyClee algorithm. As a result of the clustering, different groups that dynamically change from one evolution to another were obtained, identifying common speeds at different time instants, which allows making decisions regarding the city traffic.

ISBN 978 -987-633-574-4

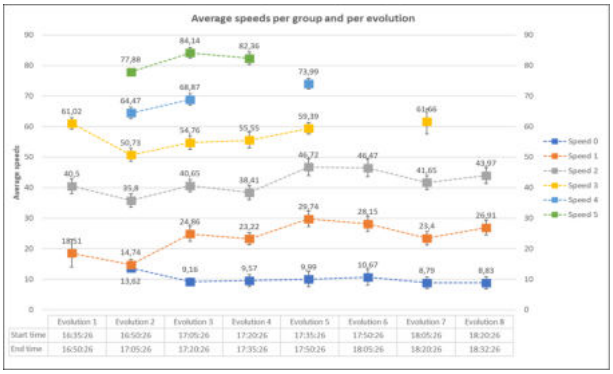


Fig. 6. Average speeds per group and evolution

Interactive maps as part of the methodology are an extremely useful tool when it comes to visualizing, according to the study area, the cells belonging to the different groupings. Through it, is possible to observe particular characteristics of each group and analyze the flow of traffic in specific sectors of the city.

As lines of future work, it is proposed to analyze the incremental incorporation of the data within the clustering together with the concept of forgetting. In this way, it will be sought to give the clustering the possibility of detecting changes in the behavior of vehicular traffic that will help to identify congestion in a more efficient manner.

References

1. A. Jain, "Data clustering: 50 years beyond k-means. 2009," *Pattern Recognition Letters*, 2009.
2. T. S. Madhulatha, "An overview on clustering methods," *arXiv preprint arXiv:1205.1117*, 2012.
3. B. Bahmani, B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii, "Scalable k-means++," 2012.
4. H. F. Tork, "Spatio-temporal clustering methods classification," in *Doctoral Symposium on Informatics Engineering*, vol. 1, pp. 199–209, Faculdade de Engenharia da Universidade do Porto Porto, Portugal, 2012.
5. J. Han, M. Kamber, and A. K. Tung, "Spatial clustering methods in data mining," *Geographic data mining and knowledge discovery*, pp. 188–217, 2001.
6. B. M. Varghese, A. Unnikrishnan, and K. Jacob, "Spatial clustering algorithms-an overview," *Asian Journal of Computer Science and Information Technology*, vol. 3, no. 1, pp. 1–8, 2013.
7. J. D. Mazimpaka and S. Timpf, "Trajectory data mining: A review of methods and applications," *Journal of Spatial Information Science*, vol. 2016, no. 13, pp. 61–99, 2016.
8. N. Barbosa Roa, L. Travé-Massuyès, and V. H. Grisales-Palacio, "Dyclee: Dynamic clustering for tracking evolving environments," *Pattern Recognition*, vol. 94, pp. 162–186, 2019.

9. M. Y. Choong, R. K. Y. Chin, K. B. Yeo, and K. T. K. Teo, "Trajectory pattern mining via clustering based on similarity function for transportation surveillance," *International Journal of Simulation-Systems, Science & Technology*, vol. 17, no. 34, pp. 19–1, 2016.
10. J. Kim and H. S. Mahmassani, "Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories," *Transportation Research Procedia*, vol. 9, pp. 164–184, 2015.
11. T. Luo, X. Zheng, G. Xu, K. Fu, and W. Ren, "An improved dbscan algorithm to detect stops in individual trajectories," *ISPRS International Journal of Geo-Information*, vol. 6, no. 3, 2017.
12. L. X. Liu, J. T. Song, B. Guan, Z. X. Wu, and K. J. He, "Tra-dbscan: A algorithm of clustering trajectories," in *Frontiers of Manufacturing and Design Science II*, vol. 121 of *Applied Mechanics and Materials*, pp. 4875–4879, Trans Tech Publications Ltd, 1 2012.
13. M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *Kdd*, vol. 96, pp. 226–231, 1996.
14. Q. Yu, Y. Luo, C. Chen, and S. Chen, "Trajectory similarity clustering based on multi-feature distance measurement," *Applied Intelligence*, pp. 2315–2338, 2019.
15. N. Ferreira, J. T. Klosowski, C. Scheidegger, and C. Silva, "Vector field k-means: Clustering trajectories by fitting multiple vector fields," 2012.
16. G. Reyes-Zambrano, L. Lanzarini, W. Hasperué, and A. F. Bariviera, "GPS trajectory clustering method for decision making on intelligent transportation systems," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 5, pp. 5529–5535, 2020.
17. H. Hu, G. Lee, J. H. Kim, and H. Shin, "Estimating Micro-Level On-Road Vehicle Emissions Using the K-Means Clustering Method with GPS Big Data," *Electronics*, 2020.
18. J. Lou and A. Cheng, "Behavior from Vehicle GPS / GNSS Data," *Sensors*, 2020.
19. B. Babcock and J. Widom, "Models and Issues in Data Stream Systems." 2002.
20. M. Garofalakis, J. Gehrke, and R. Rastogi, *Data Stream Management.* 2016.
21. M. R. Ackermann, C. Lammersen, C. Sohler, K. Swierkot, and C. Raupach, "StreamKM++: A clustering Algorithm for Data Streams," *ACM Journal of Experimental Algorithmics*, vol. 17, pp. 173–187, 2012.
22. C. C. Aggarwal, *Data Streams : An Overview and Scientific Applications.* 2010.
23. C. C. Aggarwal, P. S. Yu, J. Han, and J. Wang, "- a framework for clustering evolving data streams," in *Proceedings 2003 VLDB Conference* (J.-C. Freytag, P. Lockemann, S. Abiteboul, M. Carey, P. Selinger, and A. Heuer, eds.), pp. 81–92, San Francisco: Morgan Kaufmann, 2003.

ISBN 978 -987-633-574-4

Process Mining Applied to Postal Distribution

Victor Martinez[1], Laura Lanzarini[1,2], and Franco Ronchetti[1,2,3]

[1] *Facultad de informática, Universidad Nacional de La Plata*
[2] *Instituto de investigación en Informática LIDI (UNLP-CIC)*
[3] *Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CIC-PBA)*

martinezvictor@hotmail.com, {laural, fronchetti}@lidi.info.unlp.edu.ar

Abstract. Process mining is a technique that allows analyzing business processes through event logs. In this article, different process mining techniques are used to analyze data based on the postal distribution of products in the Argentine Republic between the years 2017 and 2019. The results obtained allow stating that 85% of the shipments made conform exactly to the model. The analysis of the situations with a low level of adjustment to the discovered process constituted a tool for quick identification of some recurring problems in the distribution, facilitating the analysis of the deviations that occurred. In the future, we expect to incorporate these techniques to build early notifications that warn about the existence of excessive deviations from the process

Keywords: Process Mining, Data Mining, Postal Distribution, Postal Processes, Business Process Management

1 Introduction

Currently most information systems record the activities carried out, whether they are business management systems (ERP, CRM, WMS, BI, etc.) or developments specific to each company. In general, the information recorded includes what was done, who did it, and when it was done, among other data. From this stored data, information that describes the process can be obtained and then used to identify improvement opportunities or solves problems.

Process Mining techniques are used to analyze these data and find patterns of behavior. The tasks with which a process begins, the sequence followed, and the tasks run to end the process can be identified. This allows "discovering" the process that is being carried out for a certain activity. With process mining, you can answer questions such as: What really happened? Why did it happen? What could happen in the future? How can process control be improved? How can the process be redesigned to improve performance? [1]. One of its main advantages is that it allows working directly on real data and obtaining the true behavior of the process, which, in some cases, is not the one originally designed.

In this article, process mining techniques will be applied to postal distribution in the Argentine Republic to analyze its operation and find operational

ISBN 978 -987-633-574-4

deviations, bottlenecks and other problems that negatively impact service quality.

In postal distribution, a record is kept of all activities, from the entry of the product to its delivery to the customer. These activities include receiving the shipment, entering a warehouse, internal transfers or delivery attempts, among others. All must be done in a specific order and within a given time frame. Occasionally, deviations occur. These may be task redundancy or inconsistency (tasks are repeated or not performed in the corresponding order), excessive time to completion, or others.

In the literature, there are works that relate the postal business, big data and process mining. Such is the case of [2], which uses data mining in a big data environment in the China Post. Due to the nature of the postal business, in that article, clustering techniques were used to group customers based on behavior, consumption habits and focus of interest, generating a more accurate and effective postal marketing strategy with very satisfactory results. Another example is [3], where process mining is applied in logistics to look for similarities and differences between different delivery processes in a changing context of manufacturing and logistics. In that work, different processes are compared using clustering techniques to achieve an automated documentation of processes in a changing context. In [4], a methodology is presented that serves as a guide for the execution of process mining projects that describes the different stages. In addition, its actual application to the IBM purchasing process is shown as a case study.

In this article, different process mining techniques will be used to analyze data based on the postal distribution of products in the Argentine Republic between the years 2017 and 2019. The authors of this work wish to state that, at the date of generation of this document, they are not aware of the existence of any similar works that implement process mining to postal distribution in the Argentine Republic.

2 Process Mining

The starting point of process mining is the event log. The process to be analyzed is assumed to require the recording of a series of sequential events pertaining to the activity and that are related to a specific case. Even though additional data can be stored for each event, recording the date of the event (day and time), the case identifier, and the type of event is mandatory.

The three basic types of process mining [5] are:

– *Discovery*: Discovery techniques take a process log as a starting point and generate a model without any additional information. A relevant example is the algorithm Alpha [6], which takes log data and generates a Petri net that explains the behavior reflected in the log. Like any technique that extracts knowledge from data, the quality of the discovered process will depend on the degree of representation of the events surveyed in relation to how the

ISBN 978 -987-633-574-4

process operates. Process parts that are not represented by events cannot be discovered.
- *Compliance Verification*: It consists of comparing an existing model (it may be the one previously discovered or a different one) with the actual sequence of events to identify deviations and verify how the process works. Applications capable of generating graphical representations and animations are usually used to observe actual behavior and see to what degree it follows to the originally defined process. Its main advantage is that it shows reality, it is not a simulation, meaning that a much more accurate analysis can be carried out.
- *Improvement*: This type of process mining is aimed at extending or improving the existing process through the underlying information in the sequence of events. Unlike the Compliance Verification type, where data are compared to the model, the goal here is to modify the process.

The degree of abstraction to be used during the analysis should be regulated considering the following aspects: fitting, accuracy, generalization and simplicity [1]. There is a relationship of compromise between these that has to be taken into account to achieve good results. Fitting refers to the ability of the model to explain the observed behavior. Accuracy refers to the accuracy with which the process is executed. In this sense, it is important that the model is not overfitted to input data because this would result in a lack of generalization, preventing the desired level of abstraction from being achieved. Simplicity is also affected by overfitting, as it is achieved by adding more detail to the process description.

Finally, it should be noted that the result obtained from the process mining analysis is highly linked to the quality of the input data. In fact, it is a known fact that there is always a certain amount of noise in the data, which can be due to incomplete tracing, intervals that have not been correctly recorded, or data duplication. This information can distort or falsify the result of the analysis [7].

In general, input data has to be verified and preprocessed to remove as much noise as possible.

3 Discovering the Postal Distribution Process

In this section, the discovery of the process will be carried out using data from the postal distribution of products in the Argentine Republic between the years 2017 and 2019.

The postal distribution process encompasses different types of products. In all cases, the process consists of receiving the product from the sender (either by physical means or a digital channel) and attempting delivering it to the recipient; actual delivery may or may not be successful. Product non-delivery does not mean that the process has failed, since there may be reasons to account for the situation.

The process records at least the following activities: receipt and identification, shipment for distribution, one or more delivery attempts, waiting at the distribution center, and return (if delivery was not possible). In all cases, each

ISBN 978 -987-633-574-4

steps carried out must be recorded with a unique shipping identifier, which allows knowing the current status of the shipment and providing that information to the customer.

Data Extraction

The first step in discovering the model consists in collecting and preprocessing the data to be used. In particular, for this case study, product shipments with two home delivery attempts were used. The current procedure establishes that those products that cannot be delivered are kept for a given time so that the recipients can pick them up; after that time, the packages are returned to their corresponding senders.

A trace is defined as a shipment. Each movement that is recorded for that shipment is an event. A trace is considered to be completed when the shipment has an entry and an end on record, with successful delivery or not.

As a result of the data collection, a table or sample of around 33,000 traces and a total of 78,000 events was generated. For each event, the minimum required fields for the analysis were recorded; namely, trace ID, event ID, event date and event description (Figure 1). Each trace can have one or more associated events. Trace identifier and identifier of each of the events is needed to build the trace history.

This allowed having available the different steps or events that occurred throughout the shipment of each product.

Trace ID | Event ID | Event description | Event date

	123 trazaID	123 EveID	ABC eveDescrip	eveFecha
31	481,053	0	INGRESADO	2017-08-16 10:45:22
32	481,053	2	1 INTENTO DE ENTREGA	2017-08-18 11:15:00
33	481,053	9	DEVOLUCION	2017-08-18 13:00:00
34	481,054	0	INGRESADO	2017-08-16 10:45:28
35	481,054	2	1 INTENTO DE ENTREGA	2017-08-22 12:05:00
36	481,054	9	DEVOLUCION	2017-08-22 17:25:00
37	481,055	0	INGRESADO	2017-08-16 10:45:27
38	481,055	1	ENTREGADO	2017-08-22 15:13:00

Fig. 1. Example of events extracted for analysis

Then, the data were then transformed using the XES format [8]. XES is a grammar for a label-based language whose objective is to provide information systems designers with a unified and extensible methodology to capture system behaviors through event logs and flows [8]. Thus, data management is streamlined and can be processed by different tools more efficiently.

ISBN 978 -987-633-574-4

To facilitate the discovery of the correct process, all incomplete traces were eliminated. An incomplete trace is a trace that does not have either a start event (reception) due to a loading error or a final event (delivery, return, recipient non-existent, deceased, refused shipment, or moved). The latter could occur due to a loading error or because the stipulated time to finish the process has not yet elapsed.

Using a filter of simple heuristic rules, complete traces were identified, and only those that had valid initial and final states remained. As a result of this filtering process, approximately 16,000 traces with 43,000 events were obtained.

3.1 Process Model

To generate the process model, a classic process mining algorithm was used, the Alpha algorithm, first proposed by van der Aalst, Weijters and Maruster[6]. The objective of this algorithm is to reconstruct causality from a set of sequences of events. It builds Petri nets with special properties (workflow nets) from event logs (such as those that an ERP system might collect). Each transition in the network corresponds to an observed task.

As regards our case study of postal distribution, even though it is possible to build the model directly from the 16,000 previously mentioned traces, some cases considered to be the most representative ones were manually selected to find the ideal process that the traces should comply with, thus simplifying this discovery stage of the process.

Figure 2 illustrates the Petri net corresponding to the discovered process that the traces must fulfill.

The process found is as follows: All traces must begin with an entry event and then go out to distribution. If a trace can be delivered, the event is logged and the process ends. If it cannot be delivered, in case of a final event (recipient deceased, missing address data, no address, unknown, etc.), the reason is recorded and the process ends. If it is not a final event (for example, the address could be reached but there was no one home) a first delivery attempt is recorded. A new visit is made at a later date. If it could be delivered, the event is logged and the process ends. If on the second visit the trace cannot be delivered, the shipment is kept at the office for a time, waiting for the recipient to come to pick it up, at which time the delivery at the office is recorded. After the waiting time has elapsed, if the recipient has not come to pick up the shipment, a return event is recorded. Those are the possible scenarios contemplated by the process.

3.2 Model Verification

The process model discovered in the previous stage was generated from a subset of previously selected traces. In this section, the compliance check carried out will be described. As previously explained, our goal is to establish how well sample traces comply with the process. Steps fulfilled, steps missed, and any additional steps not reflected in the discovered process will be considered. To do this, each of the traces obtained will be compared against the discovered found

ISBN 978 -987-633-574-4

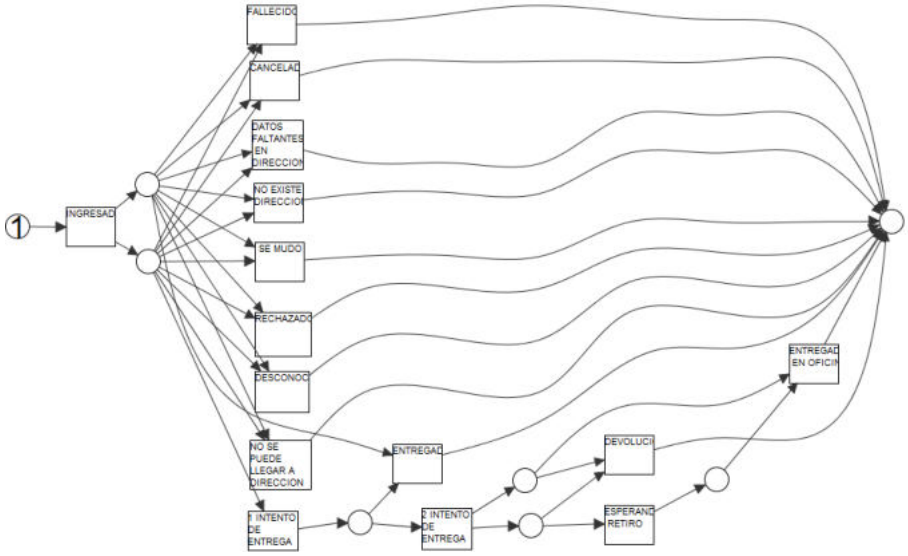


Fig. 2. Process discovered that all traces must fulfill

in the previous section. The data sample obtained in the previous section will be used, which consists of 16,811 traces with a total of 43,888 events.

The result is shown in Figure 3, with the most common events highlighted in a dark color and the most frequent runs represented with a thicker line. It can be seen that most of the traces end with the delivery, either on the first or the second attempts.

Sample statistics indicate that in 80% of the cases, the parcel is delivered either in the first or the second attempt, and that the remaining 20% is evenly distributed.

Based on these observations, it can be stated that 85% of the traces perfectly fit the process discovered, with an average number of 2.6 events in each trace.

However, some traces do not exactly follow this process, either because steps are skipped, they are not carried out in the correct order, or there are additional steps not included in the model. These traces have a compliance value that will be lower as the deviation from the model increases. Those whose compliance is below 50% are the object of analysis in this section, since this value is considered to be a significant operational deviation.

Based on these results, our analysis is focused on two paths – on the one hand, traces that have an excessive number of movements, and on the other, those that do not fit the model, either because there are missing events or because they do not follow the corresponding sequence.

In the first case, given that the number of movements per trace follows a normal distribution with a mean of 2.6 and a deviation of 0.95, it was considered appropriate to use the value of the mean plus three standard deviations

ISBN 978 -987-633-574-4

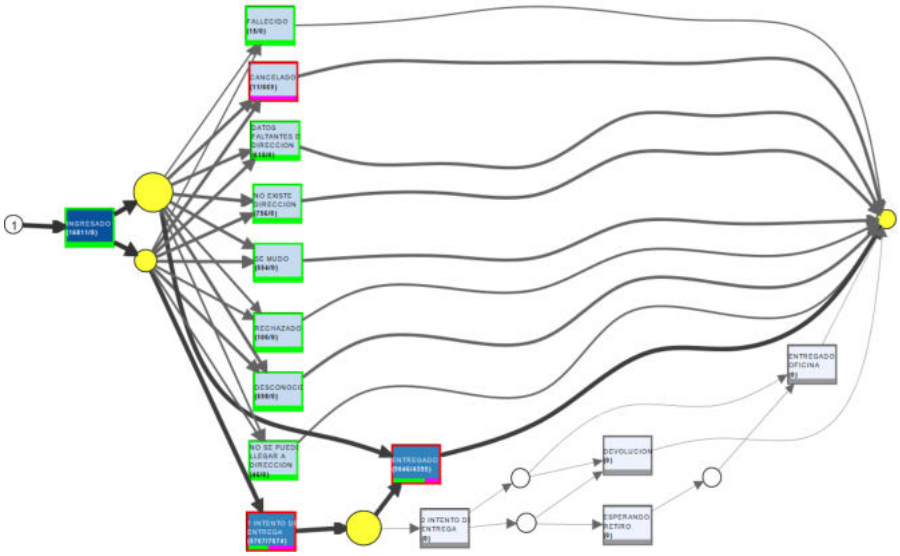


Fig. 3. Trace travel: most common events in dark and more frequent tours in thicker lines

as a representative value of an excessive number of movements. Based on this, we proceeded to filter and subsequently analyze the traces with more than 6 movements, identifying a total of 108 such cases.

The second case analyzed was that of the traces that did not exactly match the process discovered. On this occasion, the value of the compliance threshold was set at 0.5, meaning that traces with at least 50% of their movements misaligned with the process were considered to be highly deviated and required inspection. As previously stated, this can occur either because traces have events that are not part of the model or because the events in the trace do not follow the right sequence.

As a result of this, it was observed that these events were repetitions of events on different days or inconsistent records. Figure 4 illustrates both situations. The table on the left shows a case in which a first visit is recorded after a second visit; this situation can only be caused by a registration error. In the box to the right (same figure), an inconsistency is observed, since it is not possible to deliver a product that was previously returned.

Visual tools can also be used to generate animations that facilitate understanding these situations. In this particular case, with the set of traces whose compliance is below 50%, Inductive Visual Miner [9] was used to generate an animation that allows visualizing a timeline of the events for each trace. Figure 5 illustrates this animation. In this figure, each trace is symbolized with a token (or circle) that goes through the different stages of the process. This visual representation shows that some traces take longer than others and that sometimes, some erroneous behaviors occur. For example, in Figure 5, backward movements

ISBN 978 -987-633-574-4

Fig. 4. Repeated events and inconsistencies

are observed in the traces; in particular, circles have been used to highlight 25 traces that go from a second attempt to an initial attempt, and 20 that go back on themselves.

Exporting traces that do not fit the model allows carrying out a detailed analysis of the reasons for deviations and thus identifying areas of improvement in the distribution process.

4 Conclusion and future lines of research

In this article, different process mining techniques have been used to analyze data based on the postal distribution of products in the Argentine Republic between the years 2017 and 2019. At the date of generation of this document, no similar works have been found that implement process mining to postal distribution in the Argentine Republic.

Through the model generated from select traces, the process that is actually carried out was discovered. The impact that the selection criteria of these traces has on the model obtained should be noted. The first models, generated from the entire set of traces, had excessive detail and this made representation and interpretation more difficult. Currently, work is being done to identify the most frequent traces from the initial model and then automatically filter the most representative ones.

For its part, compliance verification has allowed identifying anomalous situations of interest. Cases that did not comply with the model were detected,

ISBN 978 -987-633-574-4

Fig. 5. Traces that do not comply with the model, in a circle the amount that goes backwards from a state instead of going forward (see direction of the arrow)

as well as cases that did follow it but outside the expected times or with task redundancy. Further analysis is required to determine if these cases represent manual load errors, and to look for a solution.

A fitting threshold was used to establish the minimum degree of distortion that a trace should meet so as not to affect the process. This factor should be analyzed in more detail to determine its value based on the case study at hand.

As a future line of work, we will continue working with process mining techniques not only to model situations that have already occurred, but also to be able to insert early warnings into the system when there are excessive deviations from the model.

References

1. Process Mining: Data Science in Action, Wil van der Aalst,978-3-662-49850-7,2016,Springer
2. Research of Postal Data mining system based on big data, Xia Hu1; Yanfeng Jin1; Fan Wang, 3rd International Conference on Mechatronics, Robotics and Automation, 2015, 10.2991/icmra-15.2015.124, https://www.researchgate.net/publication/300483008_Research_of_Postal_Data_mining_system_based_on_big_data
3. Context Aware Process Mining in Logistics, Mitchell M. Tseng; Hung-Yin Tsai; Yue Wang, 2017,The 50th CIRP Conference on Manufacturing Systems, https://www.sciencedirect.com/science/article/pii/S2212827117303311
4. PM2: a Process Mining Project Methodology,Maikel L. van Eck; Xixi Lu; Sander J.J. Leemans; Wil M.P. van der Aalst,Eindhoven University of Technology, The Netherlands, http://www.processmining.org/_media/blogs/pub2015/pm2_processminingprojectmethodology.pdf
5. Wil van der Aalst, The Process Mining Manifesto by the IEEE Task Force, 2012, https://www.tf-pm.org/resources/manifesto
6. Workflow mining: discovering process models from event logs, W. van der Aalst; T. Weijters; L. Maruster, 1041-4347, 2004, IEEE, https://ieeexplore.ieee.org/document/1316839

ISBN 978 -987-633-574-4

7. Process mining in flexible environments, Christian Walter Gunther,978-90-386-1964-4,2009,Technische Universiteit Eindhoven, https://research.tue.nl/en/publications/process-mining-in-flexible-environments
8. IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams, IEEE Std 1849-2016, 2016, DOI 10.1109/IEEESTD.2016.7740858
9. inductive visual miner, Sander J.J. Leemans, 2017, http://leemans.ch/leemansCH/publications/ivm.pdf
10. Reinventing the Postal Sector in an Electronic Age, Michael A. Crew; Paul R. Kleindorfer,978-1849803601,2011,Edward Elgar Publishing

ISBN 978 -987-633-574-4

Sistema de Recuperación de Información con Expansión de la Consulta Basada en Entidades

Joel Catacora[1], Ana Casali[12] y Claudia Deco[13]

[1] Facultad de Cs. Exactas, Ingeniería y Agrimensura,
Universidad Nacional de Rosario, Rosario, Argentina
[2] Centro Int. Franco Argentino de Cs. de la Información y de Sistemas
(CIFASIS: CONICET-UNR)
[3] Universidad Católica Argentina, Rosario, Argentina
{acasali,deco}@fceia.unr.edu.ar

Resumen Se propone un sistema de búsqueda semántico que expande la consulta del usuario mediante la retroalimentación por relevancia. Este sistema es aplicado al dominio legal, en particular al derecho civil, para mejorar la precisión de los resultados que puede encontrar un abogado en su búsqueda de jurisprudencia relevante para la construcción del marco legal de un caso. Para esto se utilizan las entidades pertenecientes a una base de conocimiento como medio para reformular la consulta. Se presentan dos modelos de expansión: uno automático y otro interactivo que le sugiere al usuario conceptos relacionados a su búsqueda inicial. El sistema de búsqueda propuesto se prueba a partir de un conjunto de sumarios del Sistema Argentino de Información Jurídica (SAIJ), utilizando una base de conocimiento que incluye una ontología legal desarrollada para este trabajo y el Tesauro SAIJ de Derecho Argentino. Con este sistema se pretende mejorar la experiencia de búsqueda del usuario y la precisión de los resultados.

Palabras claves: Recuperación de Información; Expansión de la Consulta; Retroalimentación por Relevancia; Base de Conocimiento; Tesauro

1. Introducción

La Recuperación de Información consiste en mostrarle al usuario documentos relevantes ante una consulta de palabras claves. Los modelos de recuperación representan formalmente el proceso de correspondencia entre la consulta y el documento. Uno de los enfoques utilizados consiste en incorporar en los modelos de búsqueda la información de una base de conocimiento mediante la expansión de la consulta. Este mecanismo añade a la consulta original otras palabras que capturan la intención del usuario o simplemente producen una consulta que permite recuperar documentos más relevantes. Las entidades o conceptos pertenecientes a la base de conocimiento pueden ser un medio para expandir una consulta.

En este trabajo se propone expandir la consulta utilizando la retroalimentación por relevancia, incorporando información semántica disponible en una base de conocimiento. En lugar de expandir la consulta con los términos de los documentos relevantes, se utilizan los términos de las entidades que se encuentran en dicha base. Se evalúan los algoritmos de búsquedas propuestos en el ámbito legal, donde

ISBN 978 -987-633-574-4

se puede asistir a los abogados en su profesión, por ejemplo, para elaborar una estrategia de defensa. Para esto, se tienen dos fuentes de información disponibles: un tesauro del dominio y una colección de documentos indexados temáticamente con el tesauro. Las fuentes de información con los que se realizó la experimentación provienen del Sistema Argentino de Información Jurídica[4] (SAIJ). El SAIJ es una base de datos documental que contiene legislación, jurisprudencia y doctrina, tanto nacional como provincial. Ofrece búsquedas por facetas a partir del Tesauro SAIJ de Derecho Argentino[5]. El usuario puede ingresar como consulta un tema del Tesauro para recuperar los documentos que se encuentran clasificados con dicho tema. Hay ciertas limitaciones en la búsqueda por facetas y búsquedas por palabras claves, por ejemplo, el usuario es el encargado de hallar los términos más cercanos a su necesidad de información y sólo puede expresar búsquedas por conjunciones de palabras claves.

Con respecto a trabajos relacionados, en Argentina hay investigaciones recientes sobre la búsqueda o recomendación de información legal y el reconocimiento de entidades nombradas en documentos legales. En [9] se propone un sistema de recomendación de normativas para construir de manera semi-automática la matriz legal de una empresa, se utiliza el algoritmo Support Vector Machine sobre leyes nacionales catalogadas con conceptos del Tesauro del SAIJ. Otro modelo de recuperación de información jurídica basado en ontologías y distancias semánticas se propone en [4] donde se utilizan vocabularios legales y generales (ConceptNet, WordReference, Banco de Vocabularios Jurídicos Argentinos) para expandir la consulta y ranquear los documentos a partir de similitudes basadas en Normalized Google Distance. En [3] se aplican técnicas de procesamiento automático del lenguaje a un conjunto de leyes de nacionales, se identifican entidades legales, utilizando el algoritmo supervisado Stanford NER y reglas manuales.

En este trabajo, se plantea un sistema de expansión semántica de búsqueda a partir de una base de conocimiento y documentos catalogados temáticamente, se lo aplica a la recuperación de información jurídica argentina. Se presentan dos modelos de expansión: uno automático y otro interactivo que le sugiere al usuario conceptos relacionados a su búsqueda inicial. El sistema propuesto se prueba sobre un conjunto de sumarios del SAIJ y una base de conocimiento integrada por una ontología legal desarrollada para este trabajo y el Tesauro SAIJ de Derecho Argentino.

La estructura de este artículo es la siguiente. En la Sección 2 se presentan conceptos preliminares. En la Sección 3 se detallan la arquitectura del sistema de búsqueda y los modelos de expansión de la consulta propuestos y en la Sección 4 se expone la experimentación realizada. Finalmente, en la Sección 5 se tienen las conclusiones.

2. Conceptos preliminares

La incertidumbre asociada a la relevancia de un documento frente a una consulta ha sido modelada probabilísticamente de diferentes maneras, entre ellas se destacan los modelos del lenguaje, estos definen una distribución de probabilidades sobre cadenas de texto representando un determinado lenguaje. En la Recuperación de

[4] http://www.saij.gob.ar
[5] http://datos.jus.gob.ar/dataset/tesauro-saij-de-derecho-argentino

ISBN 978 -987-633-574-4

Información suelen utilizarse los modelos del lenguaje más simples, los unigramas o modelos del lenguaje con distribución multinomial, donde se asume términos con independencia condicional y posicional, es decir, un modelo *bag of words*. Entre los modelos de recuperación basados en modelos se lenguaje se tiene el Query Likelihood Model (QL). Para cada documento d de la colección se define un modelo de lenguaje θ_d, el cual describe el tema del documento o las palabras claves que el usuario ingresaría si quisiera recuperar dicho documento. El puntaje del documento d es probabilidad de que la consulta $q = \langle w_1, \dots w_n \rangle$ sea una muestra o se genere de acuerdo a cada uno de estos modelos del lenguaje $P(q|\theta_d)$:

$$P(q|\theta_q) = \prod_{i=1}^{n} P(w_i|\theta_d). \tag{1}$$

El modelo del lenguaje θ_d se estima a partir del documento d, mediante Maximum Likelihood Estimation y técnicas de suavizado. Entre los métodos de suavizado se tiene la interpolación de Jelinek-Mercer (JM):

$$P(t|\theta_d) = \lambda \frac{\mathrm{tf}_{t,d}}{|d|} + (1 - \lambda) \frac{\mathrm{cf}_t}{|c|}$$

donde $\lambda \in [0, 1]$ es un parámetro del suavizado, $\mathrm{tf}_{t,d}$ es la frecuencia del término t en el documento d, cf_t es la frecuencia del término t en toda la colección de documentos y $|d|$ la longitud del documento.

Además, se han desarrollado adaptaciones que permiten incluir la estructura de los documentos (sus campos), p. ej., título, autor, en los modelos de recuperación, esto potencia el rendimiento de las búsquedas. Para el modelo QL, se tienen dos extensiones: el Mixture Language Model (MLM) [8], donde se añaden pesos como parámetros del modelo que miden la importancia de cada campo del documento, el Probabilistic Retrieval Model for Semistructured Data (PRMS) [5], donde a partir del modelo anterior se propone una forma no supervisada de estimar los pesos de los campos.

Un tipo particular de objeto que puede recuperarse es una entidad. La recuperación de entidades tiene como objetivo responder a las consultas mediante una lista ranqueada de entidades, por ejemplo "países limítrofes de Argentina". Las *entidades* o conceptos son objetos unívocamente identificables, con nombre, atributos y relaciones con otras entidades, por ejemplo, personas, localizaciones y organizaciones. Estas pertenecen a un *catálogo de entidades*, un diccionario con los nombres de las entidades junto a sus identificadores. Este catálogo puede ser una base de conocimiento modelada por una ontología. Un enfoque para ranquear a las entidades consiste en utilizar los algoritmos de recuperación tradicionales sobre representaciones documentales de las entidades, aplicando sin modificaciones los modelos diseñados para ranquear documentos. Para esto, es necesario, crear un documento, llamado *descripción de la entidad*, para cada entidad del catálogo, el cual mantiene toda la información de la entidad en la base de conocimiento. La técnica que se utiliza para la construcción de las descripciones de las entidades se denomina *predicate folding* [1, p. 69].

ISBN 978 -987-633-574-4

3. Modelo de Expansión de Consulta y Sistema Propuesto

Se propone una arquitectura de un sistema de búsqueda enriquecido semánticamente y dos modelos de búsquedas no supervisados que utilizan la información contenida en una base de conocimiento del dominio para expandir la consulta. La expansión se realiza mediante la retroalimentación por (pseudo) relevancia basada en entidades [7,1]. Por un lado se desarrolla un Modelo de Relevancia con Entidades (RE) que genera de forma automática entidades que se esperan sean relevantes a la consulta, mediante documentos que describen a las entidades. Luego, se expande la consulta con los términos más importantes de las entidades generadas. El otro método, el Modelo Iterativo de Relevancia con Entidades (IRE) requiere la asistencia del usuario para que seleccione entre las entidades sugeridas en una o más iteraciones, aquellas que considere relevantes. La arquitectura propuesta se muestra en la Figura 1, donde el usuario ingresa una búsqueda de texto libre, se la transforma en términos índice y a partir de estos términos se aplica el algoritmo de búsqueda. Luego, el sistema retorna una lista ranqueada de documentos como respuesta a su consulta (módulo *Interacción con el usuario*). Para expandir la consulta es necesaria su reformulación por parte del modelo de expansión, para luego ejecutar el algoritmo de búsqueda sobre la consulta reformulada.

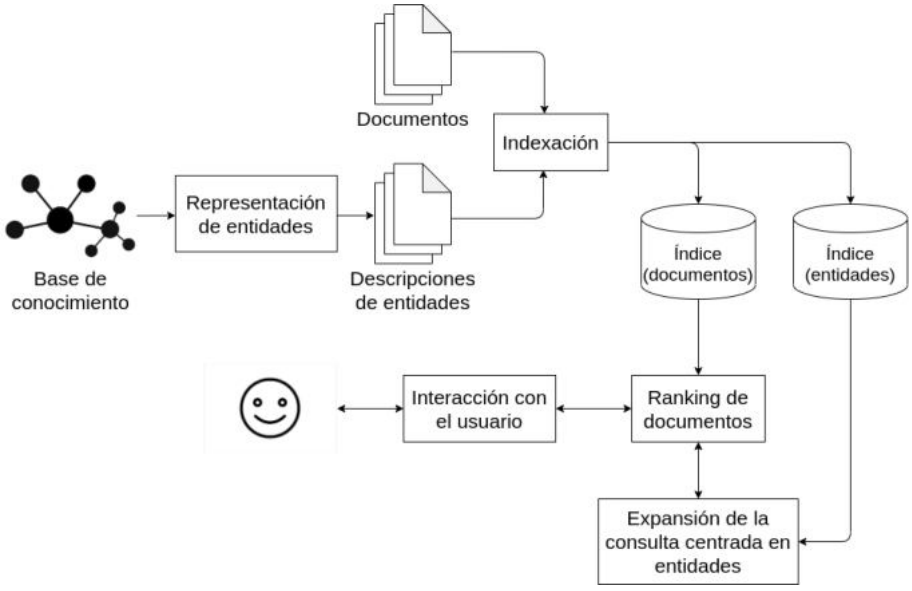


Figura 1: Arquitectura propuesta.

En este caso, el algoritmo de expansión utiliza las entidades pertenecientes a una base de conocimiento, en lugar de los documentos de la colección. El núcleo del motor de búsqueda, la implementación del modelo de recuperación y expansión de la consulta se encuentran en los módulos: *Ranking de documentos* y *Expansión de la consulta centrada en entidades*. La expansión puede realizarse de manera automáti-

ISBN 978 -987-633-574-4

ca o iterativa. Si se utiliza el Modelo IRE se genera una lista ranqueada de entidades en cada iteración del modelo y el usuario debe juzgar si son relevantes a su consulta inicial y luego se procede a la reformulación de la consulta. Este sistema tiene dos fuentes de información: la colección de documentos y la base de conocimiento. Los modelos de expansión no consultan directamente a la base de conocimiento sino que realizan sus cálculos sobre las descripciones de las entidades. Estas descripciones se crean a partir de la base de conocimiento mediante el procedimiento *predicate folding*. El módulo *Representación de entidades* implementa dicha transformación. Para realizar los cálculos del modelo de recuperación de manera eficiente se requiere precomputar ciertos valores estadísticos de la colección de documentos, los cuales se mantienen en estructuras de datos llamadas índices (módulo *Indexación*). En este caso, es necesaria la construcción de dos índices, uno por cada colección de documentos. El vocabulario de términos del sistema de búsqueda es la intersección de los vocabularios de cada índice. El modelo de expansión se encarga de consultar el índice de entidades y el modelo de recuperación de documentos consulta al índice de los documentos. Se utiliza como fuente de información una base de conocimiento legal diseñada para este trabajo y una colección de sumarios recolectados del SAIJ.

En la Tabla 1 se puede ver la descripción de la entidad "cinturón de seguridad" construida a partir de la base de conocimiento LegalBase, donde los campos *sumarios* y *sumarios-títulos* contienen todos los documentos catalogados con el concepto "cinturón de seguridad".

Campos	**Valor**
nombres	cinturón de seguridad, cinturón.
entidades-relacionadas	reglas de tránsito, tránsito automotor, automotores, vehículos, Transporte, ...
sumarios	La omisión de empleo del cinturón... La impugnación no ha de prosperar...
sumarios-títulos	Daños y perjuicios, ... Recurso de inconstitucionalidad, ...
catch-all	cinturón de seguridad, cinturón. reglas de tránsito, tránsito automotor, automotores,vehículos, Transporte, ... La omisión de empleo del cinturón... La impugnación no ha de prosperar... Daños y perjuicios, ... Recurso de inconstitucionalidad, ...

Tabla 1: Descripción de la entidad "cinturón de seguridad".

Desarrollo de una base de conocimiento legal

Para la evaluación de la propuesta se desarrollaron una base de conocimiento legal (LegalBase) y una colección de test, Sumarios20 (compuesta por 45.556 sumarios del SAIJ pertenecientes al Derecho Civil, del ámbito nacional y de la provincia de Santa Fe, junto con 10 necesidades de información). Para la creación de LegalBase se definió un tesauro de Accidentes de Tránsito reutilizando el Tesauro SAIJ y la ontología LegalOnto, la cual expresa la indexación semántica de documentos legales. Se inició la creación del tesauro identificando las palabras claves que utilizaría un abogado para describir su necesidad de información si afronta un caso

ISBN 978 -987-633-574-4

sobre accidente de tránsito. Las palabras claves se obtienen de casos ficticios y del Tesauro SAIJ. Esto se trabajó con la asistencia de un abogado. Al igual que el Tesauro SAIJ, el tesauro Accidentes de Tránsito sigue el estándar SKOS[6], define un total de 91 conceptos y 18 grupos de conceptos, de los cuales 11 son conceptos nuevos que no pertenecen al tesauro SAIJ. Por ejemplo, ante un caso de accidente de tránsito, el abogado de la víctima debería probar la existencia de los siguientes tópicos: Daño; Antijuridicidad; Relación de causalidad; Reproche o factor de atribución sobre el demandado. A partir de este ejemplo, notamos que algunos conceptos del tesauro podrían agruparse en los temas anteriores, por ej., el concepto "lucro cesante" podría formar parte del grupo "daño". Es decir, ciertos temas podrían tratarse como conjuntos de conceptos. Además, se decidió desarrollar una ontología (LegalOnto) que modele la indexación temática de documentos legales (legislación, jurisprudencia y doctrina) con conceptos SKOS. A partir de estas fuentes: el tesauro de Accidentes de Tránsito, la ontología LegalOnto y el tesauro SAIJ, se crea la base de conocimiento LegalBase para reunir el conocimiento disponible sobre las entidades legales. La integración se implementó con la directiva `owl:import` y se realizó de forma de no producir inconsistencias. La ontología LegalBase fue poblada con los sumarios de la colección Sumarios20.

Búsqueda de documentos

Se proponen dos modelos de expansión: uno que realiza una expansión automática, sin intervención del usuario, y el otro donde el usuario interviene en forma iterativa.

Expansión automática: El modelo de expansión de consultas Conceptual Language Model [7] se define de la siguiente manera para una consulta q:

$$P(t|q) \approx \sum_{e \in \mathcal{E}} P(t|e)P(e|q), \tag{2}$$

donde $\mathcal{E}$ es el catálogo de entidades. Este modelo asume que la probabilidad de seleccionar un término sólo depende del concepto una vez que se ha seleccionado ese concepto para la consulta. Deben estimarse dos componentes: la selección de términos $P(t|e)$ y la selección de entidades $P(e|q)$. Proponemos estimar: $P(t|e)$ con el modelo del lenguaje de la entidad, es decir $P(t|e) = P(t|\theta_e)$, y $P(e|q)$ con el modelo QL aplicado a la recuperación de entidades. Luego, como en [7] los documentos se ranquean con la divergencia de Kullback–Leibler (KL).

Se puede ver que el modelo de expansión propuesto es equivalente a Relevance Model [6], o también llamado RM1, con entidades en lugar de documentos, considerando todas las entidades igualmente probables ($P(e)$ se ignora). Por esto, lo llamamos Modelo de Relevancia con Entidades (RE).

Expansión interactiva: Se propone el Modelo Iterativo de Relevancia con Entidades (IRE), basado en el Iterative Relevance Model [2]. Las entidades candidatas

[6] `https://www.w3.org/2009/08/skos-reference/skos.html`

ISBN 978 -987-633-574-4

que se muestran al usuario en la i-ésima iteración se obtienen de aplicar KL:

$$score(e, q^{(i)}) = \sum_{t \in V} P(t|\theta_q^{(i-1)}) \; log \; P(t|\theta_e), \tag{3}$$

donde V es el vocabulario, $q^{(i)} = (q, \theta_q^{(i-1)})$, con $\theta_q^{(i-1)}$ la consulta expandida y reformulada de la iteración anterior, definida por la Ecuación 5, siendo $1 \leq i \leq n$. Las entidades que ya han sido mostradas en iteraciones anteriores se remueven del ranking. Las primeras k entidades de la Ecuación 3 que son juzgadas como relevantes, en la i-ésima iteración, conforman el conjunto $\mathcal{E}_q^{(i)}(k)$. Estas entidades son añadidas al conjunto de todas las revisiones hechas por el usuario hasta la i-ésima iteración $E_q^{(i)}$, o sea:

$$E_q^{(i)} = E_q^{(i-1)} \cup \; \mathcal{E}_q^{(i)}(k),$$

donde $E^{(0)} = entities(q)$ (*entity linking* sobre la consulta).

Luego, la expansión de la consulta de la i-ésima iteración es la siguiente:

$$P^{(i)}(t|\hat{\theta}_q) = \frac{1}{|E_q^{(i)}|} \sum_{e \in E_q^{(i)}} P(t|\theta_e). \tag{4}$$

Esta fórmula se obtiene de la Ecuación 2, considerando una selección de entidades por parte del usuario con distribución uniforme. En la práctica, solo se tienen en cuenta a los términos con los puntajes más altos para formar el modelo de la consulta expandida, con las probabilidades renormalizadas de modo tal que $\sum_t P^{(i)}(t|\hat{\theta}_q) = 1$. La reformulación de la consulta en la i-ésima iteración es:

$$P^{(i)}(t|\theta_q) = \begin{cases} (1-\lambda_q)P^{(0)}(t|\theta_q) + \lambda_q P^{(i)}(t|\hat{\theta}_q) & si \;\; i \geq 1 \\ \frac{\mathrm{tf}_{t,q}}{|q|} & si \; i = 0 \end{cases}, \tag{5}$$

donde $\lambda_q \in [0, 1]$ controla la influencia del modelo de expansión. En la próxima iteración, se ranquean las nuevas entidades candidatas (Ecuación 3) con $\theta_q^{(i)}$. Si $i < n$, entonces:

$$q^{(i+1)} = (q, \theta_q^{(i)}).$$

En la última iteración, $i = n$, se obtiene la reformulación de la consulta final, $\theta_q^{(n)}$. Luego, esta consulta expandida puede ser utilizada para la recuperación de documentos mediante el modelo KL. Se muestra en la Figura 2 un ejemplo de búsqueda iterativa, se sugieren 5 entidades por iteración, 5×2 (Entidades-Iteración).

En resumen, se mostraron dos modelos semánticos de búsqueda: el Modelo RE y el Modelo IRE. Ambos, utilizan una base de conocimiento para expandir la consulta a partir de los términos asociados a las entidades relacionadas a la consulta. El modelo IRE sugiere entidades ante una consulta en lugar de documentos, las cuales pueden revisarse por el usuario más rápidamente es decir, las sugerencias se relacionan semánticamente con la consulta y permiten precisar la dirección de la búsqueda ya sea para profundizarla o para moverse hacia otros aspectos de la búsqueda. Estos modelos pueden incorporan la semi-estructura de las descripciones de las entidades, mediante MLM y PRMS. Todos los modelos propuestos junto a sus extensiones son no supervisados, no requieren juicios de relevancia o *query logs*.

ISBN 978 -987-633-574-4

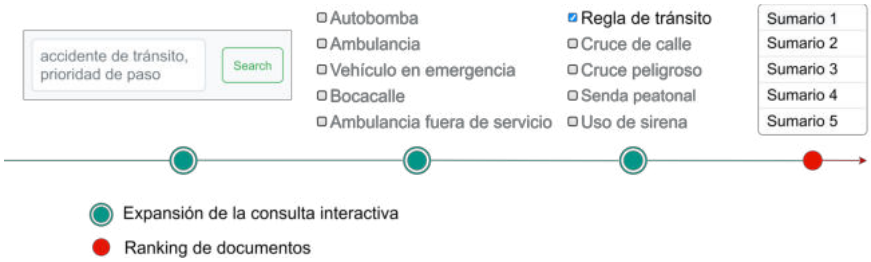


Figura 2: Línea de tiempo sobre la búsqueda "accidente de tránsito, prioridad de paso", con expansión de consulta iterativa.

4. Experimentación

Se realiza sobre la colección Sumario20 con los siguientes modelos de recuperación centrados en entidades: RE, RE+MLM, RE+PRMS, IRE, IRE+MLM, IRE+PRMS, junto con el algoritmo RM3 (con interpolación JM).

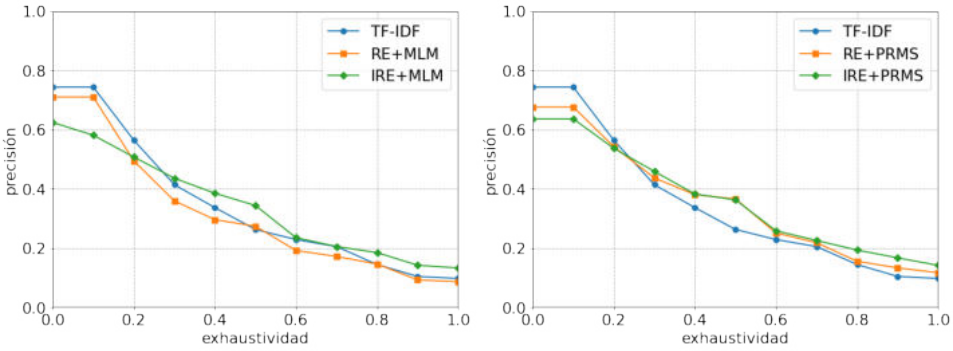
La evaluación de los algoritmos de búsqueda se realiza mediante *4-fold cross-validation*, los parámetros se ajustan con *grid search* y se utiliza *freezing ranking* para evaluar el modelo iterativo siguiendo lo propuesto en [2]. Se optimizó un conjunto determinado de variables debido al limitado poder de computo del hardware disponible. Los parámetros optimizados son: la interpolación del suavizado JM, al estimarse $P(t|\theta_e)$ en los modelos propuestos y $P(t|\theta_d)$ en RM3, y los pesos de los campos de las descripciones para los modelos PRMS y MLM, con valores en el intervalo $[0, 1]$ y saltos de 0,25. La interpolación de la consulta inicial con el modelo de la consulta expandida es $\lambda_q = 0{,}25$ en IRE (Ecuación 5) y RM3, y $\lambda_q = 0{,}5$ en RE. Se consideraron los primeros 15 términos, las primeras 10 entidades (RE), 10 documentos (RM3) y 2 iteraciones de 10 entidades, 2×10 (IRE).

En la Tabla 2 se comparan los modelos con expansión de la consulta. Los algoritmos que utilizan al modelo PRMS, alcanzan los mayores rendimientos en términos de MAP, superiores a los modelos MLM. De acuerdo a los índices de robustez, los modelos basados en MLM tienen un mayor desvío de la consulta que aquellos que usan PRMS. Es decir, los términos de expansión de la consulta generados por los modelos RE+MLM e IRE+MLM tendrían una menor relación a la consulta inicial en comparación a los términos producidos por los modelos RE+PRMS e IRE+PRMS. En la Figura 3 se muestran dos curvas de precisión y exhaustividad, donde se los compara con la implementación del modelo TF-IDF de Apache Lucene[7].

En la evaluación no se encontró diferencias significativas entre el modelo RE y modelo IRE en la colección Sumarios20. Saber cuáles son las entidades relevantes a una consulta, como lo hace el modelo IRE, no presenta en esta experimentación una diferencia frente a la expansión automática de la consulta del modelo RE. Ambos superan el rendimiento del algoritmo RM3. Entendemos que la selección interactiva de entidades no ofrece grandes beneficios ante la expansión automática, ya que las entidades seleccionadas por el usuario no aportan mejores términos para expandir la consulta. De acuerdo al proceso de creación de las descripción de las entidades, los

[7] `https://lucene.apache.org/`

ISBN 978 -987-633-574-4



(a) TF-IDF, RE+MLM, IRE+MLM. (b) TF-IDF, RE+PRMS, IRE+PRMS.

Figura 3: Curvas de precisión y exhaustividad.

Colección	Métrica	RM3	RE	RE+MLM	RE+PRMS	IRE	IRE+MLM	IRE+PRMS
Sumarios20	MAP	0.313	0.310	0.299	0.335	0.318	0.312	**0.336**
	P@10	**0.280**	0.270	0.250	0.250	0.250	0.260	**0.280**
	P@20	0.200	0.195	0.180	0.210	**0.215**	0.205	0.209
	J@20	0.994	0.845	0.875	0.880	0.980	0.985	0.990
	RI	**0.10**	-0.10	-0.20	**0.10**	0	0	**0.10**

Tabla 2: Evaluación de sistemas de recuperación con expansión de la consulta. El índice de robustez se calcula sobre TF-IDF.

términos que expanden la consulta se obtienen de documentos que tratan sobre los temas que le interesan al usuario, pero no necesariamente son los mejores términos para expandir la consulta. Además, se probaron extensiones a estos modelos: los modelos MLM y PRMS. El uso del modelo PRMS, tanto en los algoritmos RE como IRE, alcanza un mayor rendimiento que MLM. Luego, en esta experimentación los algoritmos de búsqueda con expansión de la consulta centrada en entidades que alcanzan el mayor rendimiento son los que utilizan al modelo PRMS.

5. Conclusiones

En este trabajo se propone la arquitectura de un sistema de búsqueda con expansión de la consulta para mejorar la recuperación de documentos. Este sistema permite incorporar como fuente de información a una base de conocimiento, de modo tal de expandir la consulta a partir de los términos asociados a las entidades relevantes a la consulta. Se implementaron en dicha arquitectura dos algoritmos de búsqueda no supervisados basados en el Conceptual Language Model. Uno es el Modelo de Relevancia con Entidades, el cual realiza la expansión de manera automática. El otro es el Modelo Iterativo de Relevancia con Entidades, que requiere la asistencia del usuario. Estos modelos se evaluaron en el dominio legal, utilizando una base de conocimiento legal (LegalBase) y una colección de test (Sumarios20) desarrolladas especialmente para este trabajo. Esta base de conocimiento puede

ISBN 978 -987-633-574-4

utilizarse para otras aplicaciones futuras. En la evaluación no se encontraron diferencias significativas entre estos modelos para la colección Sumarios20. Además, se probaron las extensiones MLM y PRMS sobre estos modelos, siendo PRMS el de mayor rendimiento. Se debe tener en cuenta que los resultados mostrados están sujetos a parámetros que no fueron totalmente optimizados ya que estuvo restringido por limitaciones en el hardware y que la colección de test Sumario20 es muy pequeña para ser representativa. De todas maneras los resultados mediante estos modelos de expansión semántica resultan alentadores.

A través de este sistema de búsqueda semántico se espera ayudar a los profesionales del derecho en la recuperación de documentos que les sean útiles para la redacción de una demanda. Como trabajo futuro es necesario trabajar sobre colecciones más grandes y extender el soporte a la búsqueda para otras ramas del derecho. Se propone incluir en el modelo de expansión una selección de términos que dependa de la consulta e incorporar modelos de lenguaje no supervisados como los *word embeddings*.

Referencias

1. Balog, K.: Entity-Oriented Search, The Information Retrieval Series, vol. 39. Springer (2018), `https://eos-book.org`
2. Bi, K., Ai, Q., Croft, W.B.: Revisiting iterative relevance feedback for document and passage retrieval. arXiv preprint arXiv:1812.05731 (2018)
3. Cardellino, F., Cardellino, C., Haag, K., Soto, A., Teruel, M., Alonso i Alemany, L., Villata, S.: Mejora del acceso a infoleg mediante técnicas de procesamiento automático del lenguaje. In: XVIII SID- 47 JAIIO (2018)
4. Dehner, G.A., Eckert, K.B., Lezcano, J.M., Ruidías, H.J.: Modelo de recuperación de información jurídica basado en ontologías y distancias semánticas. In: XIX Simposio Argentino de Informática y Derecho (SID 2019)-JAIIO 48 (Salta) (2019)
5. Kim, J., Xue, X., Croft, W.B.: A probabilistic retrieval model for semistructured data. In: Boughanem, M., Berrut, C., Mothe, J., Soule-Dupuy, C. (eds.) Advances in Information Retrieval. pp. 228–239. Springer Berlin Heidelberg (2009)
6. Lavrenko, V., Croft, W.B.: Relevance based language models. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 120–127. New York, NY, USA (2001)
7. Meij, E., Trieschnigg, D., de Rijke, M., Kraaij, W.: Conceptual language models for domain-specific retrieval. Inf. Processing & Management 46(4), 448–469 (2010)
8. Ogilvie, P., Callan, J.: Experiments using the lemur toolkit. In: TREC. vol. 1, pp. 103–108 (2001)
9. Perezzini, L., Casali, A., Deco, C.: Sistema de soporte para la recuperación de normativas en la ingeniería legal. In: SID 2020 - 49 JAIIO (2020)

ISBN 978 -987-633-574-4

Técnicas de Análisis de Sentimientos Aplicadas a la Valoración de Opiniones en el Lenguaje Español

Germán Rosenbrock[1], Sebastián Trossero[1], Andrés Pascal[1,2]

1 Fac. de Ciencia y Tecnología, Univ. Autónoma de Entre Ríos, Ruta 11 - Km. 11, Oro Verde, Entre Ríos, Argentina.
2 Fac. Regional Concepción del Uruguay, U.T.N, Ing Pereira 676, Concepción del Uruguay, Entre Ríos, Argentina.
rosenbrock.german@uader.edu.ar, trossero.sebastian@uader.edu.ar, andrespascal22@gmail.com

Abstract. En el presente existen grandes cantidades de datos en formato de texto escritos en el lenguaje natural, disponibles principalmente en sitios web y redes sociales, que crece día a día. El análisis manual de estos volúmenes de información es actualmente impráctico y costoso, por lo cual se hace necesario el uso de técnicas automatizadas para su procesamiento y análisis. La Minería de Opinión o Análisis de Sentimientos estudia la extracción de información a partir de datos subjetivos y es relativamente reciente. En los últimos años se han propuesto varios modelos de procesamiento del lenguaje natural para resolver el problema particular de clasificación de sentimientos. En este trabajo examinamos el rendimiento de varios de estos modelos aplicados a un caso donde los textos están escritos en el lenguaje castellano coloquial, lo que representa un desafío adicional. El caso propuesto es un conjunto de más de 50.000 reseñas de películas, extraídas del sitio www.cinesargentinos.com.ar.

Palabras claves: Minería de opinión, Análisis de sentimientos, Procesamiento del lenguaje natural en español, Data Mining, Análisis subjetivo.

1 Introducción

En un proceso de toma de decisiones, es fundamental contar con información oportuna, confiable y completa que permita un análisis real de la situación. En ciertos casos, los datos de origen son opiniones personales. En forma previa a la Web 2.0, su importancia no era alta debido a la escasa cantidad de textos que registraban opiniones. En el presente, con la disponibilidad masiva de este tipo de información, surgen nuevas oportunidades y desafíos en la búsqueda, comprensión e interpretación de la misma. Sin embargo, la búsqueda en estos sitios y la posterior valoración de las opiniones en forma manual es un trabajo intenso y costoso, por lo que es necesario contar con sistemas que automaticen este proceso.

El Análisis de Sentimientos o Minería de Opiniones estudia la interpretación automática de opiniones y sentimientos expresados mediante el lenguaje natural. Es utilizada por organizaciones, por ejemplo, para el análisis de su imagen o para determinar necesidades o también el grado de aceptación de nuevos productos. La literatura, además, muestra varios otros tipos de aplicaciones, incluyendo: valoración de películas [1], opiniones sobre deportes [2], turismo [3, 4], política [5], educación [6], salud [7], finanzas [8] y automóviles [9].

ISBN 978 -987-633-574-4

Este trabajo presenta la aplicación y comparación de distintas técnicas de aprendizaje automático como Máquinas de Vectores de Soporte (SVM), Clasificador Bayesiano Ingenuo (Naïve-Bayes), Máxima Entropía y Random Forest, con el enfoque clásico de bolsa de palabras, contra técnicas más actuales como la utilización de embeddings con redes neuronales recurrentes y Transformers, también conocidos como Modelos de Lenguaje. El caso de estudio se realiza sobre los comentarios y valoraciones de usuarios acerca de películas extraídas del sitio www.cinesargentinos.com.ar. La selección de este sitio se realizó teniendo como criterio la disponibilidad de los datos, la cantidad de opiniones, el nivel de informalidad en el uso del lenguaje, la disponibilidad de una valoración ya registrada para cada opinión (puntuaciones por estrellas), y la existencia de distintos aspectos a evaluar por cada opinión.

2 Marco Teórico

El análisis del sentimiento o la minería de opinión es el estudio computacional de opiniones, sentimientos y emociones expresadas a través de un texto. En general, las opiniones pueden centrarse en un producto, un servicio, un individuo, una organización, un evento o un tema. Utilizamos el término objeto para denotar la entidad de destino que se ha comentado. Un objeto puede además tener un conjunto de componentes (o partes) y un conjunto de atributos o propiedades). Cada componente puede tener sus propios subcomponentes y su conjunto de atributos, y así sucesivamente.

Lui [10] formaliza estos conceptos mediante las siguientes definiciones:

- **Objeto**: un objeto *o* es una entidad que puede ser un producto, persona, evento, organización o tema. Está asociado a un par, *o: (T, A)*, donde *T* es una jerarquía de componentes (o partes) y *A* es un conjunto de atributos de *o*. Cada componente tiene su propio conjunto de componentes y atributos.
- **Opinión**: una opinión sobre una característica *f* es una actitud, emoción o valoración positiva o negativa sobre *f* .
- **Orientación de una opinión**: la orientación de una opinión sobre una característica *f* indica si la opinión es positiva, negativa o neutral.

Asimismo, una opinión puede ser directa (respecto a un único objeto), o bien comparativa, que expresa una relación de similitudes, diferencias y/o preferencias entre dos o más objetos emitida por el titular de opinión sobre algunas de las características compartidas entre los objetos.

Nuestro problema es establecer si un documento expresa una opinión positiva o negativa de un objeto, aplicando diferentes técnicas de evaluación de opiniones sobre una misma base de datos, para analizar sus desempeños en forma comparativa.

Los métodos seleccionados para nuestro estudio son todos de aprendizaje supervisado, lo que significa que se requiere conocer la clase a la que pertenece la observación al momento de su entrenamiento. Los métodos son Naive Bayes, Random Forest, Regresión Logística y SVM con la representación clásica de bolsa de palabras; Redes Neuronales Recurrentes con el embedding Word2Vec y por último, para la arquitectura de Transformers se utilizó el modelo de lenguaje BETO, una versión en español del modelo original BERT.

A continuación se realiza una breve descripción de cada una de estas técnicas.

2.1 Naïve Bayes

Este algoritmo de clasificación se basa en el *Teorema de Bayes* de probabilidad condicional, además supone la independencia entre las variables predictoras. Ya que en muchos casos esta independencia no es real, se lo denomina 'Naïve' o 'Ingenuo' [11, 12, 13, 14]. La clasificación que realiza este método está dada por la probabilidad de que una observación pertenezca a una clase, dadas las probabilidades de sus variables predictoras. Es la técnica más utilizada como base de comparación.

2.2 Random Forest

Es un clasificador que consiste en un ensamble de múltiples árboles de decisión [15]. Cada uno de estos árboles se entrena con un subconjunto de registros y un subconjunto de variables del conjunto de datos tomados de forma aleatoria.

Este algoritmo puede manejar conjuntos de datos de gran dimensionalidad sin verse afectado por la colinealidad. Otra cualidad que posee este algoritmo es que se puede obtener como salida la importancia de las variables, es decir, las que más influyen en el modelo.

Es difícil de interpretar, ya que es un modelo de caja negra y dependiendo de los parámetros utilizados, en algunos casos se puede caer en overfitting.

2.3 Regresión Logística

La Regresión Logística (también conocido como clasificador de máxima entropía) [16, 17, 18], es un modelo matemático utilizado para predecir el resultado de una variable categórica, por lo general dicotómica, en función de las variables independientes o predictoras. La predicción que se obtiene es la probabilidad de pertenecer a cada clase.

Una de las ventajas fundamentales de la regresión logística sobre otras técnicas, es que el resultado del modelo entrenado se puede interpretar fácilmente. Esto se debe a que el coeficiente obtenido para cada variable dependiente, indica de qué manera influye en el modelo dicha variable. Otras ventajas son su simplicidad y eficacia.

2.4 SVM

SVM (Support Vector Machine) [2, 12, 13, 19], es un algoritmo de clasificación binario, que consiste en encontrar un hiperplano que maximice la separación entre las clases. SVM se puede utilizar con diferentes kernels dependiendo si los datos son linealmente separables o no, lo cual es un parámetro a definir. El entrenamiento de SVM con grandes conjuntos de datos no es recomendable porque no es muy eficiente.

2.5 Word2Vec+LSTM

Los *word embeddings* son una forma de representación de palabras de un documento, que además de representar la palabra aporta información de contexto dentro del documento y de similaridad con otras palabras. Word2Vec es una técnica de word embedding desarrollada en 2013 por Mikolov [20] que utiliza como representación de palabras un vector multidimensional. De esta forma, las palabras relacionadas o similares se encuentran en zonas cercanas dentro de esta representación. Estos

ISBN 978 -987-633-574-4

vectores se utilizan luego como entrada de redes neuronales para realizar tareas como clasificación, traducción o resumen de textos [21, 22].

Las redes neuronales recurrentes LSTM (Long Short Term Memory) tienen la capacidad de persistir información de estados anteriores para calcular los siguientes estados. Es por eso que son muy útiles para trabajar con secuencias, como por ejemplo en modelos de procesamiento del lenguaje natural, ya que se trata de secuencia de palabras. La limitación que tienen es que esa capacidad de "recordar" estados previos es a corto plazo. Las LSTM en cambio, son un tipo de redes neurales recurrentes que tienen ese mismo comportamiento pero a más largo plazo [23].

2.6 BERT

A finales de 2017 Google presenta una nueva arquitectura denominada Transformer [24] en la cual propone quitar las capas recurrentes y convolucionales de las redes utilizadas hasta el momento, a cambio de mecanismos o capas de atención. Estas capas de atención codifican las palabras en función de las demás palabras de la frase, permitiendo introducir información del contexto junto con la representación de cada palabra.

BERT (Bidirectional Encoder Representations from Transformers) [25] es un Modelo de Lenguaje diseñado para entrenar representaciones bidireccionales profundas a partir de textos sin etiquetar, tomando en cuenta tanto el contexto izquierdo como derecho en todas las capas. BERT ha sido pre-entrenado mediante aprendizaje no supervisado a partir de corpus de gran tamaño en idioma inglés. A diferencia de los modelos secuenciales o recurrentes tradicionales, la arquitectura de atención procesa toda la secuencia de entrada a la vez, permitiendo que todos los tokens de entrada se procesen en paralelo.

Para superar su limitación inicial de funcionamiento sólo para el inglés, han surgido versiones que soportan distintos lenguajes, o inclusive múltiples lenguajes en uno, como es el caso de mBERT [26]. Para el lenguaje español en particular, uno de los modelos más conocidos se llama BETO [27] y tiene las mismas características antes mencionadas de BERT, pero con la diferencia que el pre-entrenamiento se realizó con textos en español.

3 Experimentos Realizados

3.1 Conjunto de datos

Este estudio fue realizado sobre una base de datos de comentarios extraídos del sitio web www.cinesagentinos.com.ar; los comentarios son reseñas de distintas películas que los usuarios aportan sin ninguna estructura definida, donde además se pondera la película con un puntaje de una a cinco estrellas. Se definió que un comentario se clasifica como "positivo" si posee cuatro estrellas o más. El lote de datos final fue de 52.309 comentarios de los cuales 36.661 fueron etiquetados como positivos (aproximadamente el 70%).

ISBN 978 -987-633-574-4

3.2 Métricas utilizadas

Para evaluar la capacidad predictiva de los modelos se utilizaron las métricas usuales para estos casos de estudio, definidas de la siguiente manera:

$$Accuracy = (TP+TN) / (TP+FP+TN+FN)$$
$$Precission = TP / (TP+FP)$$
$$Recall = TP/(TP+FN)$$
$$F1\_score = 2 * (Precision* Recall) / (Precision+Recall)$$

donde: TP=True Positive, TN=True Negative, FP=False Positive, FN=False Negative.

3.3 Descripción de los experimentos

Con el fin de obtener el mejor modelo para cada uno de los algoritmos se realizó una búsqueda de hiperparámetros por medio del método Grid Search, entrenando modelos con distintos valores de los parámetros propios de cada algoritmo, quitando o dejando las “stop words” y con distintas cantidades de las palabras más frecuentes en los comentarios. A continuación se describen los hiperparámetros de ajuste:

- Naïve Bayes: ajusta un parámetro “Alpha” entre 0 (cero) y 1 (uno); es un parámetro de corrección o regularización para evitar problemas con la probabilidad cero de eventos ocultos.
- Random Forest: se define la cantidad de estimadores que corresponde a la cantidad de árboles de decisión que se utilizan. Los valores posibles van desde 1 en adelante, sin un límite superior.
- Regresión logística: se ajusta un parámetro llamado Solver con los posibles valores: "liblinear","sag" y "saga". Cada uno ajusta el modelo tomando distintas métricas de penalización.
- SVM: en el caso de este algoritmo se define el tipo de Kernel que utiliza; los posibles valores son: linear, polynomial y RBF.

Como representación del texto de entrada, para los cuatro primeros algoritmos se utilizaron “Bolsas de palabras” (en adelante BdP), que se definen mediante vectores cuyas columnas están indexadas por cada una de las palabras que se encuentran en el conjunto de datos completo, y que almacena en sus valores la concurrencia de esas palabras en el comentario. A este método también se ajustaron los siguientes parámetros para el Grid Search:

- La cantidad de palabras (o columnas de los vectores): se define *n* como la cantidad máxima de palabras a utilizar en la BdP, teniendo en cuenta que sean las *n* palabras con mayor concurrencia en el conjunto de datos. Este parámetro fue ajustado entre 1.000 y 50.000 palabras.
- Eliminar Stop Words: las Stop Words (en adelante SW) son palabras del lenguaje que no poseen riqueza semántica, por ejemplo, los conectores. En los experimentos se utilizaron dos diccionarios distintos de SW para el lenguaje español, uno incluido en la librería NLTK y el otro generado a partir del mismo. Los valores de ajuste de este hiperparámetro fueron: “No borrar SW”, “Borrar diccionario completo” y “Borrar diccionario alternativo”.

ISBN 978 -987-633-574-4

Para el caso de word2vec, el embedding fue generado a partir de las palabras de los mismos comentarios; mientras que para el modelo de Transformers, BETO ya cuenta con un embedding pre-entrenado con palabras en español.

Para cada iteración de parámetros de Grid Search se entrenaron cinco modelos distintos utilizando la técnica Monte Carlo Cross Validation. Las divisiones del conjunto de datos para entrenamiento y prueba se realizaron al 80% y 20% respectivamente. Se calcularon las métricas *Accuracy*, *Presicion*, *Recall* y *F1-Score*, tomando esta última como referencia para determinar el mejor modelo y seleccionar sus hiperparámetros como los óptimos.

3.4 Resultados

Los resultados obtenidos por los distintos algoritmos se muestran en la Tabla 1. A continuación se realiza una breve descripción de los mismos, y los hiperparámetros con los que se obtuvieron los mejores valores.

Tabla 1 Puntajes de los modelos de clasificación sobre el conjunto de datos de prueba.

Modelos	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	0.80	0.80	0.81	0.81
Random Forest	0.80	0.82	0.77	0.79
Regresión logística	0.80	0.80	0.80	0.80
SVM	0.79	0.79	0.79	0.79
LSTM + Word2Vec	0.81	0.85	0.88	0.87
BERT (BETO)	0.83	0.85	0.91	0.88

3.4.1 Naïve Bayes

El resultado de la búsqueda Grid Search para este algoritmo obtuvo el mejor F1-score de los cuatro primeros algoritmos con una ponderación aproximada del 81% de clasificación correcta (ver Tabla 1). El modelo fue entrenado con un valor 1 en el parámetro Alpha y 50.000 palabras en la BdP sin eliminar ninguna “Stop Word”.

Los distintos experimentos realizados arrojan resultados que demuestran que para esta aplicación el aumento del parámetro Alpha también aumenta la potencia predictiva del modelo resultante. Del mismo modo se observa que cuantas más palabras se utilicen para entrenar el modelo lleva a un aumento del F1-score. Por otro lado, la eliminación de SW no tiene resultados positivos en cuanto al F1-score, por el contrario, no eliminarlas mejora los resultados un 1%.

3.4.2 Random Forest

En el caso de este algoritmo se realizaron búsquedas de Grid Search ampliando la cantidad de estimadores hasta que el aumento de los puntajes no fue significativo. El modelo óptimo lo encontramos con 2.000 estimadores y una BdP de 50.000 palabras,

ISBN 978 -987-633-574-4

sin quitar las SWs. El porcentaje de comentarios correctamente clasificados por este modelo fue 79%.

3.4.3 Regresión Logística

El “Solver” que maximizó el F1-Score para este problema fue “saga” con un puntaje aproximado de 80%. En este caso la cantidad óptima de palabras fueron 40.000 para conformar la BdP, al igual que los otros, sin quitar SWs.

3.4.4 SVM

Este algoritmo se optimizó con el Kernel “linear” con 1.000 palabras en su BdP sin quitar las SWs tampoco. De los cuatro modelos que emplearon BdP, fue el que obtuvo el puntaje más bajo de F1-Score, con aproximadamente el 79% de la clasificación correcta.

3.4.5 LSTM+Word2vec

El mejor resultado obtenido fue generando un embedding de 500 palabras, sin quitar SWs, y con un learning rate de 0,02 en el entrenamiento de la red neuronal. Se obtuvo un F1-Score de 87%.

3.4.6 BETO (BERT)

El F1-score obtenido en este experimento fue de 88%, clasificando de forma incorrecta solo 1.815 del total de 10.462 comentarios. La tasa de *learning rate* óptima fue de 0,03, el *batch size* de 64, y la cantidad de palabras seleccionadas por comentario fue 150. Tampoco se quitaron las SWs.

3.5 Análisis de los Resultados

Tal como se esperaba, las dos técnicas más recientes obtuvieron los mejores resultados, alrededor de un 7% más que las primeras cuatro, aunque entre ellas no hay diferencias significativa en este caso. En cuanto al preproceso de los datos se observó que la eliminación de Stop Words tanto del diccionario original de la librería NLTK como el modificado, no generó mejores resultados si no que, por el contrario, disminuyó su rendimiento.

En la literatura reciente, existen distintos trabajos de clasificación de comentarios de películas escritos en inglés [28, 29, 30], en donde utilizando BERT se obtuvieron como resultado entre un 85% y un 94% de Accuracy, mientras que en nuestro caso de estudio el valor alcanzado fue 83%, es decir, entre un 2% y un 11% menos. Esta diferencia puede tener varias causas: diferencias propias del lenguaje, pre-entrenamiento con un corpus de menor tamaño, diferencias en el nivel de informalidad del lenguaje coloquial utilizado, o incluso, mejor ajuste de algunos hiperparámetros.

3.5.1 Comentarios mal clasificados

Para comparar los comentarios mal clasificados tomamos en cuenta solo los 2 mejores modelos obtenidos, LSTM+Word2vec y BETO. Del total de 10.462 comentarios del conjunto de testing, 1.992 fueron mal clasificados utilizando el primer algoritmo, mientras que con BERT fueron 1.815. Teniendo en cuenta que se utilizó el mismo conjunto de testing para los experimentos, se observó que 994 comentarios fueron mal clasificados por ambos algoritmos a la vez.

3.5.2 Causas de la clasificación errónea

Analizando los comentarios mal clasificados, encontramos al menos cinco posibles causas por las cuales el comentario no obtuvo la clasificación correcta:

1. Casos en los que, a pesar de que el comentario tiene una connotación positiva, la etiqueta original del mismo es negativa. Es decir, el autor del comentario escribió una opinión positiva de la película, pero la calificó negativamente.
   Por ejemplo: *"la película me pareció buena, mantiene el suspenso y está muy bien filmada, el efecto 3d está muy bien logrado"*, *"comedia entretenida, divertida, para pasar un buen rato y reírse bastante. Cameron Díaz es muy buena en la comedia y el elenco está muy bien" o "linda comedia, buenas actuaciones y los actores se complementan muy bien pero lo mejor de la película en mi opinión es la elección de música, el mejor soundtrack que he visto en mucho tiempo"*.
2. Casos de comentarios calificados positivamente por el usuario, pero acompañado de un comentario con mensaje negativo: *"No me terminó de convencer. A la peli le pasa factura todos los problemas que tuvo a la hora de realizarse. La trama a pesar de ser interesante se hace por momentos algo aburrida." o "Decepcionante. Se nota que le falta media hora. Para pasar el rato pero nada más. Está hecha sin ganas"*
3. Comentarios ambiguos, es decir, con cierto balance entre lo positivo y negativo. Por ejemplo, *"supero mis expectativas, las escenas de susto un poco predecibles" o " los primeros minutos son algo aburridos pero al pasar los minutos la pelicula es cada vez es entretenida"*.
4. Frases con sentido figurado, que probablemente no son aprendidas correctamente por el modelo: *"se paso en un suspiro", "Navegando aguas misteriosas debería ser la frase de esta saga" o "sin tramos de baches"*.
5. Negación y a veces doble o triple negación en la misma frase: es probable que los modelos tengan problemas cuando se invierte el sentido de una frase a través de la negación: "no es una pelicula de la que te arrepientas de haber visto" o "Esta nueva entrega no aporta ni suma nada".

Los primeros dos casos no están asociados a los modelos sino a los datos, y sólo son problemáticos cuando el entrenamiento se realiza sobre un corpus que posee una cantidad significativa de ellos.

Respecto a los comentarios ambiguos, una solución parcial que se presenta en distintos trabajos, es definir una tercer clase "neutral" para los casos en los cuales no está claro si el comentarios es positivo o negativo. Las últimas dos causas son

conocidas limitaciones de la mayoría de los modelos, ya que hasta el momento ningún modelo comprende realmente el significado del texto, sino que se basan en las relaciones de co-ocurrencia que encuentran entre las palabras.

4 Conclusiones y Trabajo Futuro

En este trabajo se presenta la aplicación, búsqueda de hiperparámetros, comparación y análisis de resultados de distintas técnicas de aprendizaje automático utilizadas para el Procesamiento del Lenguaje Natural. El caso de estudio fue un conjunto de más de 50.000 comentarios en lenguaje español coloquial sobre películas, extraídos del sitio www.cinesargentinos.com.ar. Los resultados indican que las técnicas más nuevas, Word2vec+LSTM y BERT, son superiores a los modelos anteriores, aunque los porcentajes de acierto obtenidos en este estudio son menores que los publicados sobre casos similares en los cuales los textos se encuentran en idioma inglés.

Algunas de las tareas que se plantean como trabajo futuro son:

- Re-etiquetar los comentarios mal etiquetados del conjunto de datos y volver a ejecutar los experimentos.
- Realizar un ajuste fino del modelo BETO, utilizando un porcentaje de los comentarios como conjunto de entrenamiento.
- Agregar una clase “neutra” en los procesos de entrenamiento y clasificación.
- Discriminar entre frases con sentido literal y figurado, y entrenar clasificadores separados para cada caso.

Referencias

1. KuatYessenov. Sentiment Analysis of Movie Review Comments. 2009.
2. N. LI and D. D. W. Using text mining and sentiment analysis for online forums hotspot detection and forecast. DecisionSupportSystems, vol. 48, nº 2, pp. 354 - 368, 2010.
3. L. C. Fiol, J. S. García, M. M. T. Miguel and S. F. Coll, «La importancia de las comunidades virtuales para el análisis del valor de marca. El caso de TripAdvisor en Hong Kong y París,» Papers de turisme, nº 52, pp. 89-115, 2012.
4. C. Henriquez, J. Guzmán and D. Salcedo. Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles. Procesamiento del Lenguaje Natural, vol. 56, pp. 25-32., 2016.
5. S. Rill, D. Reinel, J. Scheidt and R. V. Zicari. PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. Knowledge-Based Systems, vol. 69, pp. 24-33, 2014.
6. A. Ortigosa, J. M. Martín and R. M. Carro. Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior, vol. 31, pp. 527-541, 2014.
7. F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi and L. Donaldson. Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online. Journal of medical Internet research, vol. 15, nº 11, 2013.
8. X. Dong, Q. Zou and Y. Guan. Set-Similarity joins based semi-supervised sentiment analysis. Neural Information Processing. Springer Berlin Heidelberg, 2012., from Neural Information Processing, Springer Berlin Heidelberg, 2012, pp. 176-183.
9. P. D. Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th annual meeting on association for computational linguistics, Stroudsburg, PA, USA, 2002.

10. Liu B., Zhang L. (2012) A Survey of Opinion Mining and Sentiment Analysis. In: Aggarwal C., Zhai C. (eds) Mining Text Data. Springer, Boston, MA.
11. N. LI and D. D. W. Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decision Support Systems, vol. 48, nº 2, pp. 354 - 368, 2010.
12. A. Abbasi, H. Chen and A. Salem. Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. ACM Transactions on Information Systems (TOIS), vol. 26, nº 3, p. 12, 2008.
13. F. Pla and L.-F. Hurtado. Sentiment Analysis in Twitter for Spanish. Natural Language Processing and Information Systems, pp. 208 - 213, 2014.
14. Gutiérrez Esparza Guadalupe, Margain Fuentes María de Lourdes, Ramírez del Real Tania Aglaé, Canul Reich, Juana, Un modelo basado en el Clasificador Naïve Bayes para la evaluación del desempeño docente, RIED. Revista Iberoamericana de Educación a Distancia (volumen: 20, núm. 2) pp. 293 – 313, 2017.
15. Belgiu M., Dragut L. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, Volume 114, 2016.
16. Pang, Bo & Lee, Lillian & Vaithyanathan, Shivakumar. (2002). Thumbs up? Sentiment Classification Using Machine Learning Techniques. EMNLP.
17. Wang, Z. (2010). Document Classification Algorithm Based on Kernel Logistic Regression. Industrial and Information Systems (IIS), 2010 2nd International Conference on (Volume: 1) (págs. 76 - 79). Dalian: IEEE.
18. Kamran Kowsari, kiana Jafari Meimandi. Text Classification Algorithms: A Survey, 2019, Information Open Access Journals.
19. David Meyer, Support Vector Machines, The Interface to libsvm in package e1071, FH Technikum Wien, Austria, 2019.
20. T. Mikolov, I. Sutskever, K. Chen, et al., Distributed Representations of Words and Phrases and their Compositionality, arxiv:1310.4546v1, 2013.
21. A. Aubaid y A. Mishra, Text Classification Using Word Embedding in Rule-Based Methodologies: A Systematic Mapping, TEM Journal. Volume 7, Issue 4, Pages 902-914, ISSN 2217-8309, 2018.
22. T. López Solaz, J. Troyano, J. Ortega y F. Enríquez, Una aproximación al uso de word embeddings en una tarea de similitud de textos en espa˜nol, Procesamiento del Lenguaje Natural, Revista n° 57, pág. 67-74, 2016.
23. T. Sainath, O. Vinyals, A. Senior y H. Sak, Convolutional, long short-term memory, fully connected deep neural networks, 2015.
24. A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need. 2017.
25. J. Devlin, M. Chang, K. Lee y K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv:1810.04805v2, 2019.
26. T. Pires, E. Schlinger y D. Garrette, How multilingual is Multilingual BERT?, arXiv:1906.01502v1, 2019.
27. J. Cañete, G. Chaperon, R. Fuentes and J. Ho, Spanish Pre-Trained BERT Model and Evaluation Data, PML4DC at ICLR 2020, 2020.
28. M. Munikar, S. Shakya and A. Shrestha, Fine-grained Sentiment Classification using BERT, arXiv:1910.03474v1, 2019.
29. L. Maltoudoglou, A. Paisios, H. Papadopoulos, BERT-based Conformal Predictor for Sentiment Analysis, Proceedings of Machine Learning Research 128:1–16, 2020.
30. S. Garg and G. Ramakrishnan, BAE: BERT-based Adversarial Examples for Text Classification, arXiv:2004.01970v3, 2020.

ISBN 978 -987-633-574-4

A comparison of text representation approaches for early detection of anorexia

Ma. Paula Villegas[1,2], Marcelo L. Errecalde[1], Leticia C. Cagnina[1,2]

[1] Laboratorio de Investigación y Desarrollo en Inteligencia Computacional
Universidad Nacional de San Luis (UNSL)
Ejército de los Andes 950, (5700) San Luis, Argentina
[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina
{villegasmariapaula74, merreca, lcagnina}@gmail.com

Abstract. The excessive use of filters on photos, along with the hundreds of profiles on social networks that abuse retouching to reduce centimeters of their bodies, and the existing social pressure on body image, has caused an increase in Eating Disorders (ED). Thus, anorexia, bulimia nervosa and binge eating disorder are the main eating disorders that put physical and mental health at risk, especially among the very young people. Fortunately, there are technologies that allow the early detection of certain problems in different areas, in particular those related to safety and health, such as the mentioned above. Our main objective in this work is to analyze how different representations of texts behave for the early detection of people suffering from anorexia. Although we focus on ED, we believe that these results could be extended to other risks such as depression, gambling, etc. We employ *k*-TVT, an efficient and effective method used previously in the detection of signs of depression, as well as other more elaborated approaches such as Word2Vec, GloVe, and BERT. To compare the performance of these methods, we worked on a data collection provided by the eRisk 2018 laboratory to detect signs of anorexia disorder. Regarding the results, the performance of the different approaches was quite similar, with *k*-TVT and BERT being slightly better. We also conclude that *k*-TVT continues to be efficient with its flexibility, low dimension and easy computing being the more attractive characteristics.

Keywords: Early Risk Detection, Anorexia Detection, Learned Text Representations, Temporal Variation of Terms.

1 Introduction

Millions of people worldwide suffer from eating disorders such as anorexia, bulimia, orthorexia, binge eating disorders or related conditions that put their physical and mental health at risk. The incidence and prevalence of eating disorders in young people has increased in last decades. Many experts agree that the causes of these eating disorders are diverse in nature: sociocultural, psychological, hereditary, etc. However, the excessive pressure to have a "perfect" body along with the high exposure of people in social networks have contributed to make the problem worse. It is common to see messages on the web that associate happiness and success with a standard of physical perfection, and children and adolescents, in the search for their identity, are particularly

ISBN 978 -987-633-574-4

vulnerable to these messages, risking their health in order to be thin. Thus, the importance of the early detection of this type of ED.
In this sense, a task was proposed by the organizers of eRisk in 2017[1] challenge. In that year, they focused on identifying people with depression [1] and the following year (eRisk 2018[2]) anorexia screening was added [2]. In that laboratory, a collection of writings was proposed to study the relationship between anorexia and the use of language. However, this new corpus turns out to be smaller in terms of the number of documents than the one proposed for the depression task [3], increasing the difficulty of the task.

In previous studies, we presented *k*-Temporal Variation of Terms (*k*-TVT, for short) [4, 5]. This text representation was used to address the early detection of people with depression and, its effectiveness and robustness were successfully demonstrated [4-6]. In this paper, we use *k*-TVT and other more elaborated text representations such as Word2Vec, GloVe, and BERT in order to analyze the convenience of each one for the early risk detection of ED. We select those representations because they are considered state of the art for many problems, even the one studied here [7-11]

Although we focus on early detection of anorexia in this paper, we believe that the results obtained could be extended to other risks. With this work we will try to answer the following research questions:

- **RQ1:** Is *k*-TVT still performing well in the detection of anorexia?
- **RQ2:** In detecting signs of anorexia, do the state of the art methods provide optimal results?
- **RQ3:** Which can be a good representation if factors such as dimension, resources and complexity are considered?

Therefore, in Section 2 we briefly define the different text representations used in this paper. Then, in Section 3 the dataset used in our experiments is described together with the description of the task of early detection of signs of anorexia. In Section 4 an experimental study is carried out to analyze the performance of the classifiers with the selected representations. Finally, Section 5 summarizes the main conclusions obtained and future work.

2 Text representations analyzed

To compare different representations of texts for the detection of people with anorexia, we selected some classical like Bag of words and n-grams of characters, the previously presented *k*-TVT and some of the most promising learned representations such as Word2Vec, GloVe and BERT. We provide a short description of each one below.

One of the most used representations in text categorization tasks is Bag of Term (BoT), which represents each document as a bag, that is, an unordered set, of terms. We generally refer to 'term' as a word, however it can also be associated with a sequence of two or more words or a sequence of two or more characters. The first case is commonly known as **Bag of Words (BoW)** [12], which is characterized by being simple to implement and quick to obtain. However, it has the disadvantage that by

[1] https://erisk.irlab.org/2017/index.html#task
[2] https://erisk.irlab.org/2018/index.html

ignoring the order of the words in the document, a lot of semantic and conceptual information is lost. In the second case, a term is considered a sequence of words or characters, obtaining another representation called **n-grams of words** or **n-grams of characters** respectively [13]. Particularly in the English language, the 3-grams of characters are the ones that have demonstrated the best performance [14] because among others, they can capture significant suffixes such as 'ly_', 'ing', 'ed_'. In addition, the 3-gram of characters are useful for informal text collections where misspelled words or words with repetitions of characters can usually appear. Formally, in BoT a document is represented by a vector of weights of each term, according to the chosen weighting scheme (tf, tf-idf, boolean, etc.). The vectors in this representation are often huge and can be very sparse.

On the other hand, ***k*-TVT** [4] is a representation that, in addition to words, focuses on the context in which they occur. This type of approach is based on the principle that words that occur in similar contexts tend to have similar meanings [12]. Each context can be modeled through semantic elements called concepts and thus, represent words and documents with a combination of concepts. The set of concepts associated with each word can be viewed as a bag of concepts and then the document will be represented by the concepts associated with the words in that document.

In particular, *k*-TVT first represents the terms based on the different contexts in which they can occur, and then generates the document vectors from those representations. To determine the concepts, *k*-TVT does it in a simple way: using the class labels of the classification task. In this representation, two words are related if their relative frequency distributions in the documents of the different classes are similar. That is, the more frequent a word is in documents that belong to a class, the greater its membership in that class. In addition to the good performance shown by the models that use *k*-TVT, other advantages are the low dimensionality of the vectors that represent each document and the balance that it performs between the minority (or positive) class with respect to the majority (or negative) class.

Finally, the learned representations emerge, where the idea is to extend machine learning, usually used in a later step to generate the classification model, to the document representation. For the learning of representations, word embeddings are used, which are basically distributed representations of words based on dense vectors of fixed length, which are obtained from statistics of the co-occurrence of words according to a distributional hypothesis.

In the specific literature, it is possible to distinguish between representations derived from counting-based approaches such as **GloVe** (*Global Vectors*) [15], and those arising from predictive neural learning methods such as **Word2Vec** [16] and **BERT** (*Bidirectional Encoder Representations from Transformers*) [17], among others.

Generally, in predictive approaches, neural networks with many units are used and they are fed with extensive collections of texts in an unsupervised way, which enables representations to learn general concepts of languages. Thus, word embeddings capture very interesting syntactic and semantic relationships of words, such as relational meanings.

Word2Vec is a method used to obtain distributed representations of words (word embeddings). Basically, the authors in [16] proposed to learn a classifier in order to obtain word representations as a collateral effect. Embeddings are extracted from the pre-output layer of a neural network classifier. This approach considers the context in

which the word is found using the size of the context window as a parameter. For small windows, the words most similar to the target word will be semantically similar and will have the same grammatical category. On the other hand, in larger windows the words that are most similar to the target word will be semantically related without necessarily becoming similar.

The local context window parameter in combination with a global matrix factorization determines the representation of **GloVe**. The authors in [15] argue that the quotient of the probabilities of coexistence of two words is what contains the relevant information. First, the co-occurrence matrix between all words in the vocabulary in the text collection is calculated. Then, based on that matrix, the algorithm learns a vector for the representation of each word and another vector in which the context of that word is modeled. Finally, both vectors can be averaged to obtain the vector representation of each word and then, as in Word2Vec, they will be used to represent a document.

In recent years, the state of the art in the field of natural language processing has focused on predictive approaches based on transformers [18]. Transformers are a type of deep neural network architecture that includes an attention mechanism. These mechanisms encode each word of a sentence as a function of the rest of the sequence, thus allowing context to be introduced into the representation (contextual embeddings).

BERT is a powerful language representation model which uses the transformer architecture. This model pre-trains deep bidirectional representations from unlabeled text and can be fine-tuned with just adding an output layer to the neural network of the model. BERT can process a document as a sequence of sentences of tokens; analyzing left and right contexts of each token it produces a vector representation for each word as the output, considering a pre-trained model. BERT is pre-trained on a large corpus of unlabeled text: 2,500 million words extracted from Wikipedia and 800 million from Book Corpus. Although we can perform the fine-tuning of the model on a specific task and task-specific data, in this work we only use the embeddings computed from BERT, that is, the text representation of the model. We explain better this issue in Section 4.

3 Data Set and Pilot Task

The data sets were provided for the eRisk 2018 laboratory. From the two task proposed by the organizers, we only used here the early prediction of anorexia [6]. Each set is a collection of writings (posts or comments) of social media users taken from Reddit. There are two categories of users: those that have been diagnosed with anorexia (positive class) and a control group (non-anorexia or negative class). For each user, the dataset contains a sequence of writings (in chronological order) divided into 10 chunks. The first chunk contains the oldest 10% of the messages, the second chunk contains the second oldest 10%, and so forth. Table 1 summarizes, for both sets, the number of users for each class.

Then, the task Early Detection of Signs of Anorexia consists of determine if a user has some signs compatible with anorexia disorder, indicating if there is a person at risk or not. The challenge is to process sequentially the texts of each user and detecting the first traces of anorexia as early as possible.

ISBN 978 -987-633-574-4

Table 1. Data set for the anorexia detection task.

Training		Test	
Anorexia	Non-Anorexia	Anorexia	Non-Anorexia
20	132	41	279

The task is mainly focused on evaluating text mining solutions and therefore concentrates on texts written on social media. Texts must be processed in the order they were created. In this way, systems that perform well on this task could be applied to sequentially monitor user interactions on blogs, social networks, or other types of online media.

4 Experimental Study

In this study we use the corpus described above. First, in Subsection 4.1, the results obtained with all the representation are shown, including *k*-TVT and the approaches of the state of the art Word2vec, GloVe and BERT. Note that, for each metric reported, the bag of words (BoW) or character trigrams (C3G) models will be taken as the baseline. Finally, in Subsection 4.2 we compare with the best values obtained in eRisk 2018 competence and analyze the different text representations trying to answer the research questions asked before.

The executions were carried out using programming language Python 3. The scripts were originally created on the Anaconda platform that runs offline, however we move our experiments to the online Google Colaboratory platform.

Four learning algorithms were used in all experiments: Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF) and Logistic Regression (LR). The implementations of these algorithms correspond to those provided in the *Python scikit-learn* library with the default parameters.

The performance of the classifiers was evaluated using the measures F1, precision, recall and Early Risk Detection Error ($ERDE_\sigma$) [3]. $ERDE_\sigma$ simultaneously evaluates the precision of the classifiers and the delay in making a prediction. The parameter σ is a time limit for decision making, that is, if a correct positive decision is made at time t > σ, it will be considered as incorrect (false positive). In this way, the parameter σ allows specifying the level of urgency required for the task, that is, the lower the value of σ, the sooner a user at risk must be detected. For that, we vary the parameter σ considering the values: 5, 10, 25, 50 and 75 and analyze the performance of classifiers in the different levels of urgency. Note that σ = 5 means high urgency (a quick decision must be made) and σ = 75 represents the lowest urgency (there is more time to make a decision) in detecting positive cases.

Regarding the word embedding models Word2Vec and GloVe, we use the pre-trained vectors[3], given the computational effort required to obtain new ones. Then, the document vectors were obtained by averaging the embeddings of the words that appear in each one (we use vectors of 300 dimensions).

[3] Pre-trained vectors were obtained from https://nlp.stanford.edu/projects/glove/ and https://s3.amazonaws.com/dl4j-distribution/GoogleNews-vectors-negative300.bin.gz

In the case of BERT, the texts of the users were divided into sentences and the bert-as-service[4] library was used to extract the corresponding vectors from the model. We set a maximum sentence length of 40 words. The service returns the vector of the sentence, condensing the information from the second to the last hidden layer of the neural network. Finally, for BERT, the document text representation vector was generated by averaging the embeddings of the sentences which have 768 characteristics. We use BERT pre-trained vectors corresponding to the *Base* version available for download in the official Google repository[5].

4.1 Performance of the different text representations

We will compare the results obtained with each representation combined with the classifiers. Regarding the baseline (shown in the table in italic), we have taken the classical representations such as bag of words (BoW) and trigrams of characters (C3G) and we have selected the best of those as reference value. For both, different weighting scheme were used: Boolean, Term-Frequency (TF) and Term Frequency - Inverse Document Frequency (TF-IDF), where the best results were achieved with the latter.

As we said before, *k*-TVT defines concepts that capture the sequential aspects of early risk detection problems and the vocabulary variations observed in the different stages of writings. Therefore, a different number *k* of chunks that will enrich the minority (positive) class could have an impact on the $ERDE_\sigma$ measure. As for detecting depression, in this study the value of *k* was varied in the range [0, 5] (integer) and we show the best value obtained with that specific *k* for each metric.

In each chunk, classifiers usually produce their predictions with some confidence, generally the estimated probability of the predicted class. Therefore, we can select different thresholds τ considering that an instance is assigned to the target class when its associated probability p is greater (or equal) than a certain threshold τ ($p \geq \tau$). Our study considered 4 different scenarios for the assigned probabilities for each classifier: $p \geq 0.9$, $p \geq 0.8$, $p \geq 0.7$ and $p \geq 0.6$. Note that once a classifier determines that an instance is positive in a specific chunk, that decision remains unchanged until chunk 10 (that is, all information is processed).

In next tables, we use the notation X-TVT, where X references to number of chunks (*k*) used for the representation *k*-TVT. For the learned text representations, we use Word2Vec (w2v), GloVe and BERT acronyms.

In Table 2 we can observe the representations in the first column, followed by a column that references to the classifier used (these values are SVM, NB, RF and LR), and then, the probability in the third column. Next, we list the values of all metrics and, due to space limitations, we select those configurations that obtained the best values for each metric and with each text representation. The bests are highlighted in bold. In addition, we wanted to emphasize the best value among the best, for which we highlighted the model which performs the best (representation with classifier-probability configuration).

[4] https://github.com/hanxiao/bert-as-service
[5] https://github.com/google-research/bert

ISBN 978 -987-633-574-4

Table 2. Comparison between all models for Early Detection of Anorexia.

		Classifier	p	ERDE 5	ERDE 10	ERDE 25	ERDE 50	ERDE 75	F1	Pre	Re
Best $ERDE_5$	***BOW***	***LR***	***0.6***	***11.52***	*11.14*	*10.23*	*8.68*	*8.68*	*0.62*	*0.78*	*0.51*
	1-TVT	RF	0.6	11.55	**10.58**	9.18	7.47	7.47	0.67	0.77	0.59
	w2v	SVM	0.7	11.77	11.37	10.28	8.72	8.54	0.50	0.70	0.39
	GloVe	SVM	0.6	11.67	11.22	9.68	7.55	7.51	0.62	0.73	0.54
	BERT	RF	0.6	11.68	11.08	9.61	7.67	7.37	0.66	0.68	0.63
Best $ERDE_{10}$	*BOW*	*LR*	*0.6*	***11.52***	*11.14*	*10.23*	*8.68*	*8.68*	*0.62*	*0.78*	*0.51*
	1-TVT	**RF**	**0.6**	11.55	**10.58**	9.18	7.47	7.47	0.67	0.77	0.59
	w2v	LR	0.6	11.83	11.15	9.94	8.52	8.37	0.53	0.59	0.49
	GloVe	SVM	0.6	11.67	11.22	9.68	7.55	7.51	0.62	0.73	0.54
	BERT	RF	0.6	11.68	11.08	9.61	7.67	7.37	0.66	0.68	0.63
Best $ERDE_{25}$	*C3G*	*SVM*	*0.6*	*11.64*	*11.26*	*10.03*	*8.49*	*8.19*	*0.59*	*0.70*	*0.51*
	5-TVT	**RF**	**0.7**	11.96	10.82	**9.07**	7.36	7.13	0.68	0.69	0.66
	w2v	LR	0.6	11.83	11.15	9.94	8.52	8.37	0.53	0.59	0.49
	GloVe	SVM	0.6	11.67	11.22	9.68	7.55	7.51	0.62	0.73	0.54
	BERT	RF	0.6	11.68	11.08	9.61	7.67	7.37	0.66	0.68	0.63
Best $ERDE_{50}$	*BOW*	*RF*	*0.6*	*12.79*	*12.08*	*10.03*	*7.53*	*7.27*	*0.74*	*0.93*	*0.61*
	5-TVT	**RF**	**0.7**	11.96	10.82	9.07	**7.36**	7.13	0.68	0.69	0.66
	w2v	RF	0.6	12.01	11.34	10.11	8.08	7.94	0.58	0.66	0.51
	Glove	SVM	0.6	11.67	11.22	9.68	7.55	7.51	0.62	0.73	0.54
	BERT	RF	0.6	11.68	11.08	9.61	7.67	7.37	0.66	0.68	0.63
Best $ERDE_{75}$	*BOW*	*RF*	*0.6*	*12.79*	*12.08*	*10.03*	*7.53*	*7.27*	*0.74*	*0.93*	*0.61*
	5-TVT	**RF**	**0.6**	12.16	11.02	9.27	7.56	**7.09**	0.64	0.61	0.66
	w2v	RF	0.6	12.01	11.34	10.11	8.08	7.94	0.58	0.66	0.51
	Glove	SVM	0.6	11.67	11.22	9.68	7.55	7.51	0.62	0.73	0.54
	BERT	RF	0.6	11.68	11.08	9.61	7.67	7.37	0.66	0.68	0.63
Best F1	*BOW*	*RF*	*0.6*	*12.79*	*12.08*	*10.03*	*7.53*	*7.27*	*0.74*	*0.93*	*0.61*
	5-TVT	LR	0.6	11.80	10.67	9.22	7.51	7.28	0.68	0.76	0.61
	w2v	RF	0.6	12.01	11.34	10.11	8.08	7.94	0.58	0.66	0.51
	Glove	SVM	0.6	11.67	11.22	9.68	7.55	7.51	0.62	0.73	0.54
	BERT	**NB**	**0.6**	13.08	12.59	9.93	7.75	7.51	**0.76**	0.71	**0.83**
Best Precision	***BOW***	***RF***	***0.9***	*12.81*	*12.81*	*12.58*	*12.50*	*12.50*	*0.26*	***1.00***	*0.15*
	2-TVT	RF	0.9	12.40	12.22	11.02	10.71	10.71	0.41	0.85	0.27
	w2v	**RF**	**0.9**	12.81	12.81	12.81	12.81	12.81	0.05	**1.00**	0.02
	Glove	RF	0.8	12.54	12.20	11.92	11.20	10.67	0.42	0.92	0.27
	BERT	**RF**	**0.9**	12.81	12.81	12.81	12.50	12.50	0.09	**1.00**	0.05
Best Recall	*BOW*	*RF*	*0.6*	*12.79*	*12.08*	*10.03*	*7.53*	*7.27*	*0.74*	*0.93*	*0.61*
	5-TVT	RF	0.7	11.96	10.82	9.07	7.36	7.13	0.68	0.69	0.66
	w2v	RF	0.6	12.01	11.34	10.11	8.08	7.94	0.58	0.66	0.51
	Glove	SVM	0.6	11.67	11.22	9.68	7.55	7.51	0.62	0.73	0.54
	BERT	**NB**	**0.6**	13.08	12.59	9.93	7.75	7.51	**0.76**	0.71	**0.83**

Considering the good performance of *k*-TVT obtained previously for the early detection of depression, for this case of study we can observe similar results. Note that *k*-TVT obtained good results for all ERDEs (quite similar in $ERDE_5$ to the best that is the baseline). We found that an adequate configuration of the *k* parameter is: $k = 1$ when the level of urgency is high ($ERDE_5$ and $ERDE_{10}$) and $k = 5$ (more information is taken into account in the representation) when the level of urgency is lower ($ERDE_{25}$, $ERDE_{50}$

and $ERDE_{75}$). Its combination with the Random Forest classifier and low probabilities are suitable configurations for obtaining good ERDE metrics.

When we analyze the learned representations, for F1, precision and recall, BERT obtained the best values. For precision, bag of words and Word2Vec also obtained the best value.

As we concluded in previous works, *k*-TVT is a good alternative to represent documents when urgency is the main factor to be considered in the classification task, meanwhile learned representations seem to be better when precision and recall metrics matter.

Finally, we analyze the pros and cons of each representation. If we consider the execution time[6] necessary to obtain each text representation, BERT takes more time (90 minutes approximately) than GloVe and Word2vec (25 minutes approximately); being *k*-TVT the fastest computed (2 minutes on average). Regarding the size of the representations (dimension), *k*-TVT has the lowest dimension. For example, 2-TVT uses vectors of size 4 (k + 2) while the others consider large vectors: GloVe and Word2Vec with dimension 300 and BERT with 768. Therefore, the larger the size, the more computing power is required and, consequently, the response delay is greater.

4.2 Performance comparison with previous results

To complement our analysis, we compared the best results in Table 2 with the results obtained in the eRisk2018 laboratory. In the proposed task, only F1, Precision, Recall, $ERDE_5$ and $ERDE_{50}$ were considered; thus, we only show the best values for those metrics.

In the first four rows of Table 3 we can see the best values obtained in the experiments that we are carrying out in this study (BOW best $ERDE_5$, 5-TVT best $ERDE_{50}$, BERT with NB best F1 and recall, BERT with RF best precision). While in the last four rows we can see the best values (final results) published by the organizers for each metric reported.

Regarding error metrics, the lowest (best) value for $ERDE_5$ was achieved by our team (**UNSLB**), that is, our participation in the competition; while the lowest value for $ERDE_{50}$ was achieved by the German team **FHDO-BCSGD**, as well as the highest recall measure. Regarding the remaining metrics, the best F1 was obtained for the German team with the variant **FHDO-BCSGE**, while for precision, our team obtained the highest (0.91) using another method called Sequential Incremental Classification (SIC) (**UNSLD**) [6], although this value is now outperformed by the best precision reached with BERT (1.0).

Even though our UNSLB approach used *k*-TVT as the text representation, it is worth noting that the lab result is slightly lower (better) than the best values obtained with all *k*-TVT settings showed in Subsection 4.1 (see Table 2). The reason is that in the laboratory we adjust the parameters of the classifier used (the logistic regressor) in order to obtain the best results for the competition. On the contrary, in this study we left the parameters of each classifier with the default because we were interested in the analysis of the different representations without any adjustment of the classifiers used.

[6] As it ran on a virtual platform Google Colab, memory and disk resources depend on the allocation of the moment, having only chosen GPU environment.

ISBN 978 -987-633-574-4

Considering the results of eRisk2018 for the rest of the metrics, except precision, neither *k*-TVT nor the learned approaches could exceed the values obtained by the FHDO-BCSG (versions D and E) and UNSLB teams.

Table 3. Comparison with the best results in 2018 Lab for Early Detection of Anorexia.

	Classifier	p	ERDE		F1	Pre	Re
			5	50			
BOW	LR	0.6	11.52	8.68	0.62	0.78	0.51
5-TVT	RF	0.7	11.96	7.36	0.71	0.62	0.82
BERT	NB	0.6	13.08	7.75	0.76	0.71	0.83
BERT	RF	0.9	12.81	12.50	0.09	**1.00**	0.05
UNSLB	LR	0.6	**11.40**	7.82	0.61	0.75	0.51
UNSLD	-	-	12.93	9.85	0.79	0.91	0.71
FHDO-BCSGD	-	-	12.15	**5.96**	0.81	0.75	**0.88**
FHDO-BCSGE	-	-	11.98	6.61	**0.85**	0.87	0.83

Finally, we can conclude that although the analyzed approaches have not achieved the best performance obtained by the winner in the eRisk 2018 laboratory, their results are acceptable and if factors such as size, available resources and computational complexity are considered, *k*-TVT represents a suitable model to face this type of task where the most important thing is the time of delay in answering.

5 Conclusions and Future Work

In this article we have evaluated the performance of the *k*-temporal variation of terms (*k*-TVT), as well as that of several state-of-the-art approaches, for the early detection of anorexia.

As summary we can observe that the different approaches have demonstrated to be competitive to solve the problem, obtaining good results but failing to match the results published in eRisk2018 lab.

Regarding the comparison of *k*-TVT with learned representations, we can observe that the temporal variation approach outperformed the other representations when urgency was taken into account. On the other hand, we observe that BERT representation is preferable when the measures to be optimized do not have urgency factor requirements. Finally, we can conclude that beyond the effectiveness and efficiency of *k*-TVT, there is an advantage in terms of computational complexity and dimensionality over the other representations considered.

As future work, we want to modify the *k*-TVT method to process the data considering one writing at a time instead of taking fragment by fragment (chunk approach) as we have been working until now. In this way, we can use the datasets proposed in the last editions of the eRisk laboratories.

References

1. Losada D.E., Crestani F., Parapar J. eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental Foundations. In: Jones G. et al. (eds) Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2017. Lecture Notes in Computer Science, vol 10456, pp. 346-360. Springer, Cham. 2017.
2. Losada, D. E., Crestani F., Parapar J. Overview of eRisk: Early Risk Prediction on the Internet. In: Bellot P. et al. (eds) Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2018. Lecture Notes in Computer Science, vol 11018, pp. 343-361. Springer, Cham. 2018.
3. Losada, D. E., Crestani, F.: A test collection for research on depression and language use. In: Fuhr N. et al. (eds) Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2016. Lecture Notes in Computer Science, vol 9822, pp 28-39. Springer, Cham. 2016.
4. Cagnina, L., Errecalde, M. L., Garciarena Ucelay, M. J., Funez, D. G., Villegas, M. P. *k*-TVT: a flexible and effective method for early depression detection. In XXV Congreso Argentino de Ciencias de la Computación (CACIC). UNRC, Córdoba. 2019.
5. Errecalde, M. L., Villegas, M. P., Funez, D. G., Garciarena Ucelay, M. J., Cagnina, L. C.: Temporal variation of terms as concept space for early risk prediction. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Vol 1866, 2017.
6. Funez, D.G., Ucelay, M.J., Villegas, M.P., Burdisso, S., Cagnina, L.C., Montes-y-Gómez, M., Errecalde, M. UNSL's participation at eRisk 2018 Lab. CLEF. 2018.
7. Wang, Y. T., Huang, H. H., Chen, H. H. A Neural Network Approach to Early Risk Detection of Depression and Anorexia on Social Media Text. In *CLEF (Working Notes)*. 2018.
8. Aguilera, J., Farías, D. I. H., Ortega-Mendoza, R. M., Montes-y-Gómez, M. Depression and anorexia detection in social media as a one-class classification problem. *Applied Intelligence*, 1-16. 2021.
9. Shah, F. M., Ahmed, F., Joy, S. K. S., Ahmed, S., Sadek, S., Shil, R., Kabir, M. H. Early Depression Detection from Social Network Using Deep Learning Techniques. In 2020 IEEE Region 10 Symposium (TENSYMP) pp. 823-826. IEEE. 2020.
10. Bucur, A. M., Cosma, A., Dinu, L. P. Early Risk Detection of Pathological Gambling, Self-Harm and Depression Using BERT. arXiv preprint arXiv:2106.16175. 2021.
11. Ramiandrisoa, F., Mothe, J. Early detection of depression and anorexia from social media: A machine learning approach. In Circle 2020 (Vol. 2621). 2020.
12. Harris, Z. S. Distributional structure. Word, 10(2-3), 146-162. 1954.
13. Cavnar, W. B., Trenkle, J. M. N-gram-based text categorization. Ann Arbor MI, 48113(2), pp. 161-175. 1994.
14. Blumenstock, J. E. Size matters: word count as a measure of quality on wikipedia. En Proceedings of the 17th international conference on World Wide Web, pp. 1095-1096. ACM. 2008.
15. Pennington, J., Socher, R., Manning, C. D. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pp. 1532-1543. 2014.
16. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J. Distributed representations of words and phrases and their compositionality. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K. Q. Advances in Neural Information Processing Systems 26, pp. 3111-3119. Curran Associates, Inc. 2013.
17. Devlin, J., Chang, M-W., Lee, K., Toutanova, K. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". arXiv:1810.04805v2 [cs.CL]. 2018.
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Polosukhin, I. Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008). 2017.

ISBN 978 -987-633-574-4

A Comparative Study of the Performance of the Classification Algorithms of the Apache Spark ML Library

Genaro Camele[1,2], Waldo Hasperué[1,3], Ronchetti Franco[1,4] and Quiroga Facundo Manuel[1]

[1] Instituto de investigación en informática (III-LIDI), Facultad de informática - Universidad Nacional de La Plata
[2] Becario UNLP
[3] Investigador asociado - Comisión de Investigaciones Científicas (CIC-PBA)
[4] Investigador asistente - Comisión de Investigaciones Científicas (CIC-PBA)
{gcamele,whasperue}@lidi.info.unlp.edu.ar
http://weblidi.info.unlp.edu.ar/wp/

Abstract. Classification algorithms are widely used in several areas: finance, education, security, medicine, and more. Another use of these algorithms is to support feature extraction techniques. These techniques use classification algorithms to determine the best subset of attributes that support an acceptable prediction. Currently, a large amount of data is being collected and, as a result, databases are becoming increasingly larger and distributed processing becomes a necessity. In this sense, Spark, and in particular its Spark ML library, is one of the most widely used frameworks for performing classification tasks in large databases. Given that some feature extraction techniques need to execute a classification algorithm a significant number of times, with a different subset of attributes in each run, the performance of these algorithms should be known beforehand so that the overall feature extraction process is carried out in the shortest possible time. In this work, we carry out a comparative study of four Spark ML classification algorithms, measuring predictive power and execution times as a function of the number of attributes in the training dataset.

Keywords: Big Data, Machine Learning, Classification Models, Apache Spark, Spark ML

1 Introduction

As technology becomes faster and more accessible, it is possible to collect much more information; however, that comes with increasing database volume. This phenomenon can be observed in various research areas. Astronomy, marketing, and social, economic, biological and medical sciences, among others, now have the possibility of storing and analyzing large volumes of information [9] [20] [12]

ISBN 978 -987-633-574-4

[8] [6]. This growing data size gives rise to the need for algorithms that allow extracting useful information in a reasonable time.

Among the tasks carried out in data mining, classification is one of the most commonly used; there are various works where it is used for some purpose, from predicting market behavior [7], to image classification [16] and the detection of pathologies in functional medicine [11].

Training a classification model on large volumes of data is a computationally expensive task, and training time is critical in techniques that require performing this task multiple times. Attribute or feature selection techniques are some examples. These techniques consist of selecting a subset of attributes from the available datasets to obtain with this subset the same predictive power that is available when using the entire database.

Selecting a subset of attributes that obtain a predictive power similar to that achieved with the full set of attributes is not a trivial task. There are many techniques that allow selecting an optimal subset of attributes [10] [5] [3] [18] [22]. Most of the techniques for this task consist of an iterative process during which several subsets of attributes are ranked and the best of them is selected as the final result. To measure the predictive power of a subset of attributes, a classification algorithm must be executed with that subset of attributes. Therefore, feature selection techniques must run these ranking algorithms a significant number of times until the optimal attribute subset is reached. This is why the execution time required by a classification algorithm is a critical factor for feature selection techniques, and it becomes increasingly critical as the number of attributes in the database increases.

To process large volumes of data, tools are available that allow distributing computation tasks among different nodes in a cluster of computers, so that the workload is balanced and processing times decrease. In this regard, tools such as Apache Hadoop or Apache Spark allow algorithms to be run in a distributed paradigm, abstracting the developer from all the inconveniences that this entails, such as synchronization, data transfer, fault tolerance, and so forth. In particular, Apache Spark has the Spark ML library, which contains the implementation of several machine learning algorithms such as neural networks, decision trees, Random Forest, regressions, SVM, and others. It also provides the functionalities of transformation, filtering and other utilities to pre-process the information that will be used to train the models.

In this work, a comparison between four of the classification algorithms implemented in Spark ML is presented: Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB) and MultiLayer Perceptron (MLP). Different predictive metrics from the models are measured and compared along with the running time required for training. In the experiments carried out, an ovarian cancer classification database [14] with 15154 attributes was used. The different experiments carried out have variations in the number of attributes selected to measure classification algorithm performance based on the number of attributes in the database used for training.

ISBN 978 -987-633-574-4

The remaining sections of this article are organized as follows: in Section 2, some similar publications that evaluate some of the features of Spark or Spark ML are mentioned. In Section 3, the experiments carried out are described, and in Section 4 the results obtained are presented. Finally, in Section 5, conclusions and possible future works are presented.

2 State of the Art

Several studies have been carried out comparing the performance of the classification algorithms in MLlib and Spark ML libraries. These investigations focus on the number of rows in the databases or on specific problems. No studies have been found that carry out a thorough study on the performance of classification algorithms based on the number of columns in the database, which is the objective proposed in this work.

In [15] a comparison of the NB, RF, Decision Tree (DT), SVM and Logistic Regression (LR) classifiers implemented in MLlib (version 1.6.2) is carried out. The only performance metric evaluated is accuracy. These authors run trials with different sizes (number of rows) of the training data set. In addition, since the classification is carried out with Amazon product reviews, they also evaluate the performance of the classifiers by varying the number of n-grams extracted from each review. The conclusion of the work in general lines is that logistic regression is the classifier that achieves the best results.

The experiments carried out in [21] do not focus on the characteristics of the database, but rather measure the execution time of the algorithms with 128 different configurations of a 16-node Spark cluster. They evaluate LR, SVM, RF and Gradient Boosted Trees (GBT) for Decision Trees under different settings for their hyperparameters. They use variants of the "Higgs" database from the UCI repository. The conclusions of the work detail the RAM and node vCPUs configuration that achieves the best execution time with each algorithm, but the predictive power of the models obtained is not evaluated.

In [13], the authors carry out experiments to detect which classification algorithm yields the best classification results with databases of patients suffering from a mental illness. The algorithms studied are LR, DT, RF and MLP using several sets of values for their respective hyperparameters. F1-measure, accuracy, recall, and precision are measured using two different datasets. The results obtained show that the RF algorithm is the one that yields the best classification models for the two databases analyzed.

Very similar to the previous case, but with the aim of predicting stock price fluctuations in the stock market, the authors of [19] study NB, RF, DT and LR measuring accuracy, ROC curves and PR curves together with the execution time with various configurations of a Spark cluster where the number of worker nodes is varied. The database used contains information on the United States market for the last 20 years, with a total volume of information of 1.7 GB. They determined that RF and DT are the algorithms with the best predictive power.

ISBN 978 -987-633-574-4

As for computation time, NB was the fastest algorithm followed by DT, while RF and LR were the algorithms that needed the most time to obtain a model.

In [1], the authors analyze tweets to determine patients at risk of heart attacks. They use several hyperparameter settings from the DT, SVM, RF, and LR algorithms. Since the focus of the work is on the online detection of patients at risk, only the accuracy to get the best model used in production was measured. Two experiments were carried out, one with the entire database and another one with a database with a smaller number of columns as a result of selecting the best characteristics, resulting in half the attributes. DT and RF turned out to be the best models using the database with half the characteristics. In the case of the entire database, RF was the best model, followed by LR and SVM and lastly DT.

Recently, the authors of [17] measured the performance of LR, DT and SVM in a case study of intrusion detection in computer networks. The area under the ROC and PR curves was measured, as well as accuracy and training times. LR turned out to be the algorithm that obtained the best models when measuring the area under the ROC and PR curves, while DT was the one that obtained the models with the best accuracy. LR also turned out to be the fastest algorithm, followed by DT and lastly SVM.

In [2], the performance of the LR, RF, SVM and PM algorithms in natural language processing applied to posts about the COVID-19 pandemic on Twitter is compared. The variability of the experiments lies in the different number of records used to train the models. As regards precision, recall, F1-measure, accuracy and execution time, LR was the fastest algorithm and SVM the slowest. On average, SVM and LR had better models than RF and PM.

In general terms, it can be seen that RF and SVM are the algorithms that achieve the best models in terms of prognostic power, while RF and LR are the fastest ones. In most works, the performance of Spark stands out – it scales very well in terms of number of rows, thanks to the power offered by its internal structure (RDD - Resilient Distributed Datasets). Therefore it is interesting to see how Spark performs when the parameter that is modified is the number of columns in the dataset, as is discussed below.

3 Experiment

As previously mentioned, the objective of this work is to evaluate the performance, in terms of execution time, of four classification algorithms of the Spark ML library of Spark, Naïve Bayes, Random Forest, Support Vector Machine and MultiLayer Perceptron. In the experiments carried out, different numbers of columns in the dataset used for training were used, measuring performance as the number of columns in the dataset increased. In each experiment, the training time and accuracy, precision, recall and F1-measure metrics were measured. Since our goal was not finding the best model by tuning the corresponding hyperparameters of the classification algorithms, the default values of the library itself were used, with the exception of the multilayer perceptron whose structure

ISBN 978 -987-633-574-4

was defined as two 4- and 5-neuron hidden layers. We used predictive power metrics with the simple purpose of comparing one model against another, being aware that neither of them might be the best model to allows solve the real problem.

All the experiments were carried out in a cluster made up of a single master node and three worker nodes. All four nodes had Ubuntu 20.04 LTS, an Intel(R) Core(TM) i3-4160 CPU @ 3.60GHz, and 8GB of RAM. As regards the software, the Hadoop and Spark versions used were 3.2.2 and 3.1.1, respectively. Spark ML version 3.1.1 was used.

The database used corresponds to expression data of 15,154 genes in 253 patients [14]. The inference class for this dataset corresponds to *Normal* if the patient does not present evidence of an ovarian tumor, or *Cancer* if the presence of the tumor has been detected. Each gene corresponds to a column in the dataset. Experiments were carried out with different numbers of attributes: 10, 50, 100, 500, 750, 1000, 2000, 3000, 4000, 5000 and 10000. For each of the experiments, attributes were selected at random, since as already mentioned, our goal was not obtaining the best model to solve the problem of tumor detection, but rather to measure execution time with different numbers of dataset features.

To avoid any bias, a five-fold cross-validation step was performed. To obtain these folds, stratified sampling [4] was used. Additionally, the entire process of randomly selecting attributes, dividing into folds, training and evaluating the models was performed 30 times with each classification algorithm.

The Spark class CrossValidator[1] was used to carry out the measurements. However, since this class only allows evaluating models with a single metric at a time, a custom cross-validation executor was developed. This executor ensures that the four analyzed algorithms share the same random sample of characteristics extracted from the dataset, in addition to the same samples generated in each fold of the cross-validation.

The algorithm used is detailed below and the full code, along with the evaluated dataset, can be found in a public Github repository [2].

Pseudocode for the custom cross-validation executor

```
program StratisfiedCrossValidator
    parameters
        num_of_folds, models, metrics, spark_df
    begin
        for k in 1 .. num_of_folds do
            train, validation = generate_stratisfied_subsets(spark_df)

            for model in models do
                start_time = get_current_time()
                trained_model = train_model(model, train)
                store_training_time(model, start_time)
```

[1] https://spark.apache.org/docs/3.1.1/api/scala/org/apache/spark/ml/tuning/CrossValidator.html
[2] https://github.com/midusi/classification-models-spark

ISBN 978 -987-633-574-4

```
                for metric in metrics do
                    store_metric_for_model(model, metric, validation)
                end
            end
        end

        {Returns, for each model, the best result obtained for each metric
        and the training time in all folds}
        return get_best_cross_results()
    end
end
```

Pseudocode for the main program

```
program ClassificationModelComparison
    {The previous definition for the models and metrics variables is assumed};
    df = read_local_ovarian_dataframe()
    ncols = [10, 50, 100, 500, 750, 1000, 2000, 3000, 4000, 5000, 10000]
    for n_features in ncols do
        x_and_y = extract_random_features(df, n_features)
        spark_df = convert_spark_dataframes(x_and_y)

        all_results = []
        for i in 1..30 do
            result = StratisfiedCrossValidator(spark_df, models, metrics)
            store_result(result)
        end

        {Reports the 30 separate values for each model and each metric,
        and training time}
        report_results(all_results)
    end
end
```

4 Results

Figure 1 shows the average and standard deviation of the execution times for each algorithm for the different datasets used. As it can be seen, the SVM and MLP algorithms require the longest execution time as the number of attributes used for training increases. The RF algorithm also shows increased execution time in experiments that include thousands of attributes, but it is approximately 10 times smaller than SVM and MLP. Even though this cannot be seen in the figure, Naïve Bayes also has an increase in execution time, reaching 17 seconds for the dataset with 10,000 attributes.

ISBN 978 -987-633-574-4

Figure 2 shows the average and standard deviation of the 30 separate runs for the accuracy (figure 2a), precision (figure 2b), recall (figure 2c) and F1-measure (figure 2d) metrics achieved by each of the algorithms. It can be seen that NB was the algorithm with the worst models, followed by RF, which achieved very good results in experiments with large numbers of attributes. SVM and MLP are the algorithms that best model datasets with fewer attributes. In fact, the results obtained by both models in the datasets with 10, 50 and 100 attributes are very similar to each other. In datasets with a larger number of attributes, SVM begins to differ from MLP and, although RF improves the metrics as the number of attributes increases, it can be seen that the models obtained with SVM are the best in all experiments.

The values for the four metrics that were obtained with the MLP executions using 500 attributes are intriguing. Even though a detailed study of the case was not done, we believe that these values are the product of the 500 randomly selected attributes. Apparently, those 500 attributes make up a scenario where the MLP algorithm finds local minima from which it cannot escape. Out of the 30 runs, it achieved accuracy values higher than 0.96, only in eight runs, the remaining 22 are below 0.91.

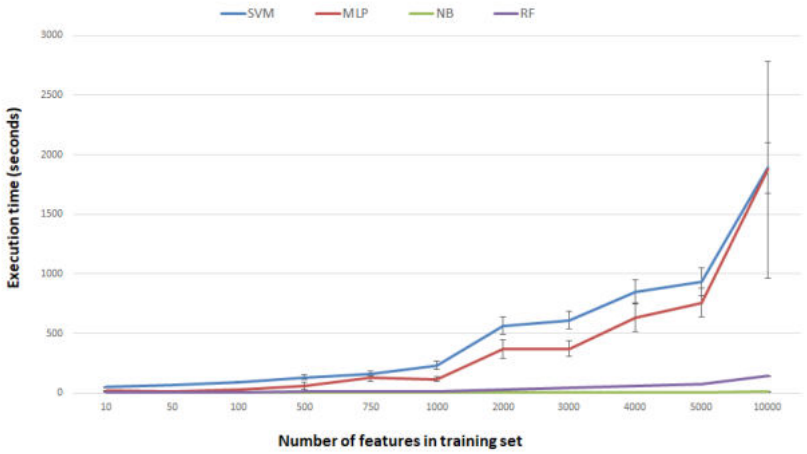
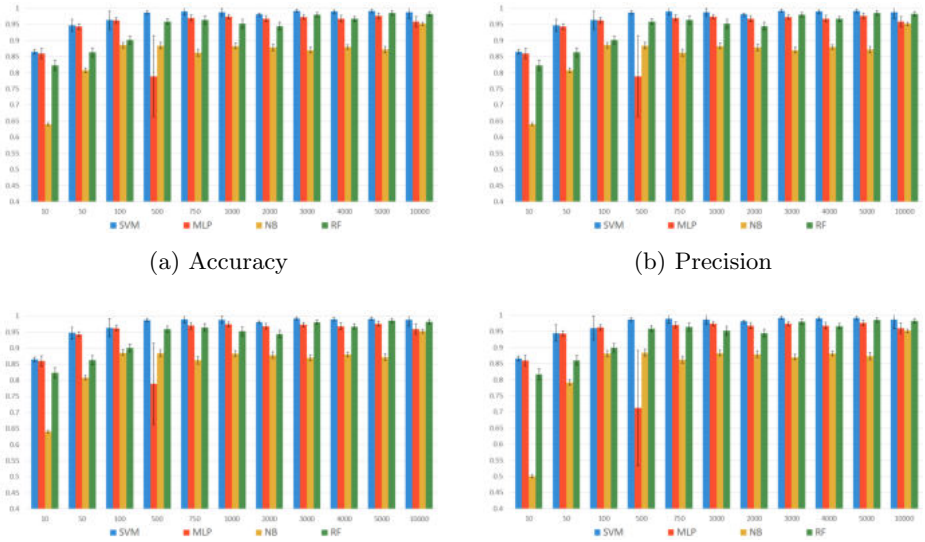


Fig. 1: Average and standard deviation of the execution times of the four algorithms studied for the different subsets of attributes used.

5 Conclusions and future work

In this work, execution time, accuracy, precision, recall and F1-measure of four Spark ML algorithms were measured by varying the number of attributes in the training dataset. The results obtained show that the algorithms that require the

ISBN 978 -987-633-574-4



(a) Accuracy (b) Precision

(c) Recall (d) F1-Score.

Fig. 2: Average and standard deviation of the accuracy, precision, recall and F1-measure metrics of the four algorithms studied for the different subsets of attributes used.

most computing time are SVM and MLP. However, in experiments with a low number of attributes, these two algorithms obtained the best models. RF turned out to be the algorithm that required the shortest execution time to achieve models with a high prediction rate.

In attribute selection algorithms, where a classification algorithm has to be run hundreds or thousands of times with different numbers of attributes, the dichotomy presented in this work opens the door to an interesting challenge, especially considering that, when selecting a subset of attributes with high predictive power, it is generally expected that the resulting subset is the smallest possible.

The results obtained show the need for tradeoff between an algorithm with low execution time while achieving good models with few attributes. On the one hand, SVM yields very good models, regardless of the number of attributes in the dataset, but it requires a lot of computation time as the number of attributes to be analyzed grows. On the other hand, Random Forest requires little execution time, but to obtain models that resemble those obtained by SVM, it needs thousands of attributes.

The dataset used in the experiments has a significant number of attributes ($>$ 15,000), but only a few samples (253). In the future, the same experiments will be carried out using datasets with more samples, and the performance of the classification algorithms in scenarios where the volume of data is greater will be studied. In these scenarios, measuring the volume of data transmitted by the

ISBN 978 -987-633-574-4

cluster during the execution of the algorithms is also of interest, since in the experiments carried out for this work, these times were negligible.

References

1. Hager Ahmed, Eman MG Younis, Abdeltawab Hendawi, and Abdelmgeid A Ali. Heart disease identification from patients' social posts, machine learning solution on spark. *Future Generation Computer Systems*, 111:714–722, 2020.
2. Wafaa S Albaldawi and Rafah M Almuttairi. Comparative study of classification algorithms to analyze and predict a twitter sentiment in apache spark. In *IOP Conference Series: Materials Science and Engineering*, volume 928, page 032045. IOP Publishing, 2020.
3. Salem Alelyani, Jiliang Tang, and Huan Liu. Feature selection for clustering: A review. *Data Clustering*, pages 29–60, 2018.
4. Zdravko Botev and Ad Ridder. *Variance Reduction*, pages 1–6. American Cancer Society, 2017.
5. Jie Cai, Jiawei Luo, Shulin Wang, and Sheng Yang. Feature selection in machine learning: A new perspective. *Neurocomputing*, 300:70–79, 2018.
6. Ashutosh Kumar Dubey, Abhishek Kumar, and Rashmi Agrawal. An efficient acopso-based framework for data classification and preprocessing in big data. *Evolutionary Intelligence*, 14(2):909–922, 2021.
7. Uma Gurav and Nandini Sidnal. Predict stock market behavior: Role of machine learning algorithms. In *Intelligent Computing and Information and Communication*, pages 383–394. Springer, 2018.
8. Eslam M Hassib, Ali I El-Desouky, Labib M Labib, and El-Sayed M El-kenawy. Woa+ brnn: An imbalanced big data classification framework using whale optimization and deep neural network. *soft computing*, 24(8):5573–5592, 2020.
9. Gerardo Hernández, Erik Zamora, Humberto Sossa, Germán Téllez, and Federico Furlán. Hybrid neural networks for big data classification. *Neurocomputing*, 390:327–340, 2020.
10. Gang Kou, Pei Yang, Yi Peng, Feng Xiao, Yang Chen, and Fawaz E Alsaadi. Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing*, 86:105836, 2020.
11. Konstantina Kourou, Themis P Exarchos, Konstantinos P Exarchos, Michalis V Karamouzis, and Dimitrios I Fotiadis. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal*, 13:8–17, 2015.
12. SK Lakshmanaprabu, K Shankar, M Ilayaraja, Abdul Wahid Nasir, V Vijayakumar, and Naveen Chilamkurti. Random forest for big data classification in the internet of things using optimal features. *International journal of machine learning and cybernetics*, 10(10):2609–2618, 2019.
13. Dorin Moldovan, Marcel Antal, Claudia Pop, Adrian Olosutean, Tudor Cioara, Ionut Anghel, and Ioan Salomie. Spark-based classification algorithms for daily living activities. In *Computer Science On-line Conference*, pages 69–78. Springer, 2018.
14. Elnaz Pashaei and Nizamettin Aydin. Binary black hole algorithm for feature selection and classification on biological data. *Applied Soft Computing*, 56, 03 2017.

15. Tomas Pranckevičius and Virginijus Marcinkevičius. Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing*, 5(2):221, 2017.
16. Franco Ronchetti, Facundo Quiroga, Genaro Camele, Waldo Hasperué, and Laura Lanzarini. Un estudio de la generalización en la clasificación de peatones. *Revista Cubana de Transformación Digital*, 2(1):33–45, 2021.
17. S Saravanan et al. Performance evaluation of classification algorithms in the design of apache spark based intrusion detection system. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 443–447. IEEE, 2020.
18. Saúl Solorio-Fernández, J Ariel Carrasco-Ochoa, and José Fco Martínez-Trinidad. A review of unsupervised feature selection methods. *Artificial Intelligence Review*, 53(2):907–948, 2020.
19. Jiang Xianya, Hai Mo, and Li Haifeng. Stock classification prediction based on spark. *Procedia Computer Science*, 162:243–250, 2019.
20. Wenchao Xing and Yilin Bei. Medical health big data classification based on knn classification algorithm. *IEEE Access*, 8:28808–28819, 2019.
21. Seyedfaraz Yasrobi, Jakayla Alston, Babak Yadranjiaghdam, and Nasseh Tabrizi. Performance analysis of sparks machine learning library. *Trans. MLDM*, 10(2):67–77, 2017.
22. Rizgar Zebari, Adnan Abdulazeez, Diyar Zeebaree, Dilovan Zebari, and Jwan Saeed. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(2):56–70, 2020.

ISBN 978 -987-633-574-4

Goodness of the GPU Permutation Index: Performance and Quality Results

Mariela Lopresti, Fabiana Piccoli, Nora Reyes

LIDIC. Universidad Nacional de San Luis,
Ejército de los Andes 950 - 5700 - San Luis - Argentina
{omlopres,mpiccoli,nreyes}@ unsl.edu.ar

Abstract. Similarity searching is a useful operation for many real applications that work on non-structured or multimedia databases. In these scenarios, it is significant to search similar objects to another object given as a query. There exist several indexes to avoid exhaustively review all database objects to answer a query. In many cases, even with the help of an index, it could not be enough to have reasonable response times, and it is necessary to consider approximate similarity searches. In this kind of similarity search, accuracy or determinism is traded for faster searches. A good representative for approximate similarity searches is the *Permutation Index*.
In this paper, we give an implementation of the *Permutation Index* on GPU to speed approximate similarity search on massive databases. Our implementation takes advantage of the GPU parallelism. Besides, we consider speeding up the answer time of several queries at the same time. We also evaluate our parallel index considering answer quality and time performance on the different GPUs. The search performance is promising, independently of their architecture, because of careful planning and the correct resources use.

1 Introduction

For a query in a multimedia database, it is meaningless to look for elements exactly equal to a given one as a query. Instead, we need to measure the similarity (or dissimilarity) between the query object and each database object. The similarity search problem can be formally defined through the metric space model. It is a paradigm that allows modeling all the similarity search problems. A metric space (X, d) is composed of a universe of valid objects X and a distance function defined among them, that determines the similarity (or dissimilarity) between two given objects and satisfies properties that make it a metric. Given a dataset of n objects, a query can be trivially answered by performing n distance evaluations, but a sequential scan does not scale for large problems. The reduction of the number of distance evaluations is meaningful to achieve better results. Therefore, in many cases, preprocessing the dataset is an important option to solve queries with as few distance computations as possible. An index helps to retrieve the objects from the database that are relevant to the query by making much less than n distance evaluations during searches [1]. One of these indices is the *Permutation Index* [2].

The *Permutation Index* is an approximate similarity search algorithms to solve *inexact similarity searching* [3]. In this kind of similarity search, accuracy or determinism is traded for faster searches [1, 4]. There are many applications where their metric-space modelizations already involve an approximation to reality. Hence, a second approximation at search time is usually acceptable.

For very large metric databases, it is not enough to preprocess the dataset to build an index. It is also necessary higher speed, in consequence, techniques of

ISBN 978 -987-633-574-4

high-performance computing (HPC)[5, 6] are considered. The Graphics Processing Units (GPU)[7] are a meaningful alternative to employ HPC in the dataset preprocess to obtain an index and to answer posed queries. The GPU is attractive in many application areas for its characteristics because of its parallel execution capabilities. They promise more than an order of magnitude speedup over conventional processors for some non-graphics computations.

In metric spaces, the indexing and query resolution are the most common operations. They have several aspects that accept optimizations through the application of HPC techniques. There are many parallel solutions for some metric space operations implemented to GPU. Querying by k-Nearest Neighbors (k-NN) has concentrated the greatest attention of researchers in this area, so there are many solutions that consider GPU. In [8–11] different proposal are made, all of them are improvements to brute force algorithm (sequential scan) to find the k-NN of a query object. In [9], Kruslis et al. propose an GPU solution to the Permutation Index. They focus in high dimensional DB and use Bitonic Sort. Their performance results are good.

The goal of this work is to analyze the trade-off between the quality of similarity queries answer and time performance, using a parallel permutation index implemented on GPU. In this analysis, we consider: different databases, two well known measures of answer quality in the information retrieval area: *recall* and *precision*, and some performance parameters to evaluate parallel implementations.

The paper is organized as follows: the two next sections describe all the previous concepts. Sections 4 and 5 sketch the characteristics of our proposal and its empirical performance. Finally, the conclusions and future works are exposed.

2 Metric Space Model

A metric space (X, d) is composed of a universe of valid objects X and a distance function $d : X \times X \rightarrow R^+$ defined among them. The distance function determines the similarity (or dissimilarity) between two given objects and satisfies several properties such as strict positiveness (except $d(x, x) = 0$, which must always hold), symmetry ($d(x, y) = d(y, x)$), and the triangle inequality ($d(x, z) \leq d(x, y) + d(y, z)$). The finite subset $U \subseteq X$ with size $n = |U|$, is called the *database* and represents the set of objects of the search space. The distance is assumed to be expensive to compute, hence it is customary to define the search complexity as the number of distance evaluations performed, disregarding other components. There are two main queries of interest [1, 4]: Range Searching and the k-NN. The goal of a range search (q, r) is to retrieve all the objects $x \in U$ within the radius r of the query q (i.e. $(q, r) = \{x \in U / d(q, x) \leq r\}$). In k-NN queries, the objective is to retrieve the set k-NN(q)$\subseteq U$ such that $\mid k\text{-NN}(q) \mid = k$ and $\forall x \in k\text{-NN}(q), v \in U \wedge v \notin k\text{-NN}(q), d(q, x) \leq d(q, v)$. These two queries are considered "exact" because both retrieve all the elements that satisfy the query criterium.

When an index is defined, it helps to retrieve the objects from U that are relevant to the query by making much less than n distance evaluations during searches. The saved information in the index can vary, some indices store a subset of distances between objects, others maintain just a range of distance values. In general, there is a tradeoff between the quantity of information maintained in the index and the query cost it achieves. As more information an index stores (more memory it uses), lower query cost it obtains. However, there are some indices that use memory better than others. Therefore in a database of n objects, the most information an index could store is the $n(n-1)/2$ distances among all

ISBN 978 -987-633-574-4

element pairs from the database. This is usually avoided because $O(n^2)$ space is unacceptable for realistic applications [12].

Proximity searching in metric spaces usually are solved in two stages: preprocessing and query time. During the preprocessing stage an index is built and it is used during query time to avoid some distance computations. Basically the state of the art in this area can be divided in two families [1]: *pivot-based algorithms* and *compact-partition-based algorithms*.

There is an alternative to "exact" similarity searching called *approximate similarity searching* [3], where accuracy or determinism is traded for faster searches [1, 4], and encompasses *approximate* and *probabilistic algorithms*. The goal of approximate similarity search is to reduce *significantly* search times by allowing some errors in the query output. In approximate algorithms one usually has a threshold ϵ as parameter, so that the retrieved elements are guaranteed to have a distance to the query q at most $(1 + \epsilon)$ times of what was asked for [13]. This relaxation gives faster algorithms as the threshold ϵ increases [13, 14]. On the other hand, probabilistic algorithms state that the answer is correct with high probability [15, 16]. That is, if a k-NN query of an element $q \in X$ is posed to the index, it answers with the k elements viewed as the k closest elements from U between only the elements that are actually compared with q. However, as we want to save as many distance calculations as we can, q will not be compared against many potentially relevant elements. If the exact answer of k-NN$(q) = \{x_1, x_2, \ldots, x_k\}$, it determines the radius $r_k = \max_{1 \leq i \leq k}\{d(x_i, q)\}$ needed to enclose these k closest elements to q.

2.1 Quality Measures of Approximate Search

An approximate answer of k-NN(q) could obtain some elements z whose $d(q, z) > r_k$. Besides, an approximate range query of (q, r) can answer a subset of the exact answer, because it is possible that the algorithm did not have reviewed all the relevant elements. However, all the answered elements will be at distance less or equal to r, so they belong to the exact answer to (q, r).

In most of information retrieval (IR) systems it is necessary to evaluate retrieval effectiveness [17]. Many measures of retrieval effectiveness have been proposed. The most commonly used are *recall* and *precision*, where *recall* is the ratio of relevant documents retrieved for a given query over the number of relevant documents for this query in the database; and *precision* is the ratio of the number of relevant retrieved documents over the total number of documents retrieved. Both recall and precision take on values between 0 and 1.

In general IR systems, only in small test collections, the denominator of both ratios is generally unknown and must be estimated by sampling or some other method. However, in our case we can obtain the exact answer for each query q, as the set of relevant elements for this query in U. By this way it is possible to evaluate both measures for an approximate similarity search index. For each query element q, the exact k-NN$(q) = Rel(q)$ is determined with some exact metric access method. The approximate-k-NN$(q) = Retr(q)$ is answered with an approximate similarity search index, let be the set $Retr(q) = \{y_1, y_2, \ldots, y_k\}$. It can be noticed that the approximate search will also return k elements, so $|Retr(q)| = |Rel(q)| = k$. Thus, we can determine the number of k elements obtained which are relevant to q by verifying if $d(q, y_i) \leq r_k$; that is $|Rel(q) \cap Retr(q)|$. In this case both measures are coincident:

$$recall = \frac{|Rel(q) \cap Retr(q)|}{|Rel(q)|} = \frac{|Rel(q) \cap Retr(q)|}{k}$$

and

ISBN 978 -987-633-574-4

$$precision = \frac{|Rel(q) \cap Retr(q)|}{|Retr(q)|} = \frac{|Rel(q) \cap Retr(q)|}{k},$$

and will allow us to evaluate the effectiveness of our proposal. In range queries the precision measure is always equal to 1. Thus, we decide to use recall in order to analyze the retrieval effectiveness of our proposal, both in k-NN and range queries.

2.2 GPGPU

Mapping general-purpose computation onto GPU implies to use the graphics hardware to solve any applications, not necessarily of graphic nature. This is called GPGPU (General-Purpose GPU), GPU computational power is used to solve general-purpose problems [18, 19, 7]. The parallel programming over GPUs has many differences from parallel programming in typical parallel computer, the most relevant are: the *number of processing units*, the *CPU-GPU memory structure*, and the *number of parallel threads*.

Every GPGPU program has many basic steps, first the input data transfers to the graphics card. Once the data are in place on the card, many threads can be started (with little overhead). Each thread works over its data and, at the end of the computation, the results should be copied back to the host main memory. Not all kind of problem can be solved in the GPU architecture, the most suitable problems are those that can be implemented with stream processing and using limited memory, i.e. applications with abundant parallelism. Each GPU-algorithm must be carefully analyzed and its data structures must be designed considering hierarchy of GPU memory, its architectures and limitations. A good GPU-algorithm has the next characteristics:

- As the data transfers between CPU and GPU could take significant amount of time, therefore, these have to be overlapped or reduced.
- The algorithm must adopt to the MIMD and SIMD paradings, and accept the SIMT execution model.
- A lot of workload needs to be spawned in order to utilize efficiently all available GPU cores.

The Compute Unified Device Architecture (CUDA) enables to use GPU as a highly parallel computer for non-graphics applications [20, 21]. CUDA provides an essential high-level development environment with standard C/C++ language. It defines the GPU architecture as a programmable graphic unit which acts as a coprocessor for CPU. The CUDA programming model has two main characteristics: the parallel work through concurrent threads and the memory hierarchy. The user supplies a single source program encompassing both host (CPU) and *kernel* (GPU) code. Each CUDA program consists of multiple phases that are executed on either CPU or GPU. All phases that exhibit little or no parallelism are implemented in CPU. Contrary, if the phases present much parallelism, they are coded as *kernel* functions in GPU. A *kernel* function defines the code to be executed by each thread launched in a parallel phase over GPU.

3 Sequential Permutation Index

Let $\mathcal{P}$ be a subset of the database U, $\mathcal{P} = \{p_1, p_2, \ldots, p_m\} \subseteq U$, that is called the permutants set. Every element x of the database sorts all the permutants according to the distances to them, thus forming a permutation of $\mathcal{P}$: $\Pi_x = \langle p_{i_1}, p_{i_2}, \ldots p_{i_m} \rangle$. More formally, for an element $x \in U$, its permutation Π_x of $\mathcal{P}$ satisfies $d(x, \Pi_x(i)) \leq d(x, \Pi_x(i+1))$, where the elements at the same distance

ISBN 978 -987-633-574-4

are taken in arbitrary, but consistent, order. We use $\Pi_x^{-1}(p_{i_j})$ for the *rank* of an element p_{i_j} in the permutation Π_x. If two elements are similar, they will have a similar permutation [2].

Basically, the permutation based algorithm is an example of probabilistic algorithm, it is used to predict proximity between elements, by using their permutations. The algorithm is very simple: In the offline preprocessing stage it is computed the permutation for each element in the database. All these permutations are stored and they form the index. When a query q arrives, its permutation Π_q is computed. Then, the elements in the database are sorted in increasing order of a similarity measurement between permutations, and next they are compared against the query q following this order, until some stopping criterion is achieved. The similarity between two permutations can be measured, for example, by *Kendall Tau*, *Spearman Rho*, or *Spearman Footrule* metrics [22]. All of them are metrics, because they satisfy the aforementioned properties. We use the Spearman Footrule metric because it is not expensive to compute and according to the authors in [2], and it has a good performance to predict proximity between elements. The Spearman Footrule distance is the *Manhattan distance* L_1, that belongs to the Minkowsky's distances family, between two permutations. Formally, Spearman Footrule metric F is defined as: $F(\Pi_x, \Pi_q) = \sum_{i=1}^{m} |\Pi_x^{-1}(p_i) - \Pi_q^{-1}(p_i)|$.

At query time we first compute the real distances $d(q, p_i)$ for every $p_i \in \mathcal{P}$, then we obtain the permutation Π_q, and next we sort the elements $x \in U$ into increasing order according to $F(\Pi_x, \Pi_q)$ (the sorting can be done incrementally, because only some of the first elements are actually needed). Then U is traversed in this sorted order, evaluating the distance $d(q, x)$ for each $x \in U$. For range queries, with radius r, each x that satisfies $d(q, x) \leq r$ is reported, and for k-NN queries the set of the k smallest distances so far, and the corresponding elements, are maintained. The database traversal is stopped at some point f, and the rest of the database elements are just ignored. This makes the algorithm probabilistic, as even if $F(\Pi_q, \Pi_x) < F(\Pi_q, \Pi_v)$ it does not guarantee that $d(q, x) < d(q, v)$, and the stopping criterion may halt the search prematurely. On the other hand, if the order induced by $F(\Pi_q, \Pi_x)$ is close to the order induced by the real distances $d(q, u)$, the algorithm performs very well. The efficiency and the quality of the answer obviously depend on f. In [2], the authors discuss a way to obtain good values for f for sequential processing.

4 GPU-Permutation Index

The GPU-CUDA system has two different steps: indexing and query resolution, they correspond whit two processes that have to be executed in sequence, first indexed process and next, query process. The Indexed process has two stages and the query process, four steps. The Figure 1 shows whole system.

Building a permutation index in GPU involves at least two steps. The first step (*Distance(O,P)*) calculates the distance among every object in database and the permutants. The second one (*Permutation Index(O)*) sets up the signatures of all objects in database, i.e. all object permutations. The process input is the database and the permutants. At the process end, the index is ready to be queried. The idea is to divide the work into threads blocks; each thread calculates the object permutation according to a global set of permutants.

In $Distances(O, P)$, the number of blocks will be defined according of the size of the database and the number of threads per block which depends of the quantity of resources required by each block. At the end, each threads block saves in the device memory its calculated distances. This stage requires a structure of

ISBN 978 -987-633-574-4

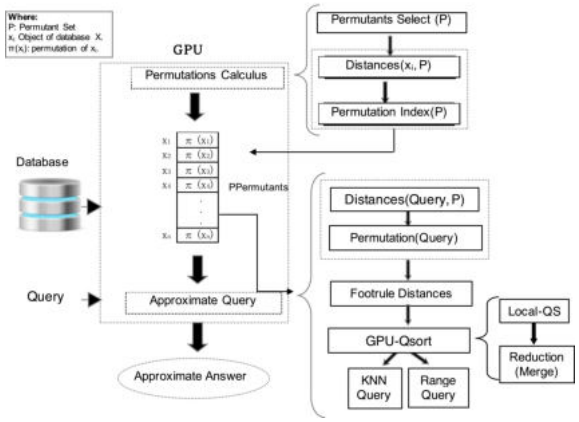


Fig. 1. Indexing and Querying in GPU-CUDA Permutation Index.

size $m \times n$ (m: permutants number, and n: database size), and an auxiliar structure in the shared memory of block (it stores the permutants, if the permutants size is greater than auxiliar structure size, the process is repeated). The second step ($Permutation\ Index(O)$) takes all calculated distances in the previous step and determines the permutations of each database object: its signature. To stablish the object permutation, each thread considers one database object and sorts the permutants according to their distance. The output of second step is the *Permutation Index*, which is saved in the device memory. Its size is $n \times m$.

The pemutation index allows to answer to all kinds of queries in approximated manner. Queries can be *"by range"* or *"*k-NN". This process implies four steps. In the first, the permutation of query object is computed. This task is carried out by so many threads as permutants exist. The next step is to contrast all permutations in the index with query permutation. Comparison is done through the $Footrule$ distance, one thread by each database object. In the third step, it sorts the calculated $Footrule$ distances. Finally, depending of query kind, the selected objects have to be evaluated. In this evaluation, the *Euclidean distance* between query object and each candidate element is calculated again. Only a database percentage is considered for this step, for example the 10% (it can be a parameter). If the query is by range, the elements in the answer will be those that their distances are less than reference range. If it is k-NN query, once each thread computes the *Euclidean distance*, all distances are sorted and the results are the first k elements of sorted list.

As sorting methodology, we implement the Quick-sort in the GPU, GPU-Qsort. The designed algorithm takes into account the highly parallel nature of GPUs. Its main characteristics are: iterative algorithm and heavy use of shared memory of each block, more details in[23].

By software and hardware characteristics, GPU allows us to think in to solve many approximated queries in parallel. The Figure 2 shows how the system is modified to solve many queries at the same time. In this Figure, you can observe that the Permutation Index is built once and then is used to answer all queries. In order to answer in parallel many approximate queries, GPU receives the queries set and it has to solve all of them. Each query, in parallel, applies the process explained in Figure 1. Therefore, the number of needed resources for this is equal to the amount of resources to compute one query multiplied by the

ISBN 978 -987-633-574-4

number of queries solved in parallel. This multiple-parallel computation involves a care management the blocks and their threads: blocks of different queries are accessed in parallel. Hence, it is important a good administration of threads. Each thread has to know which query it is solving and which database element is its responsibility. This is possible by establishing a relationship among *Thread Id*, *Block Id*, *Query Id*, and *Database Element*.

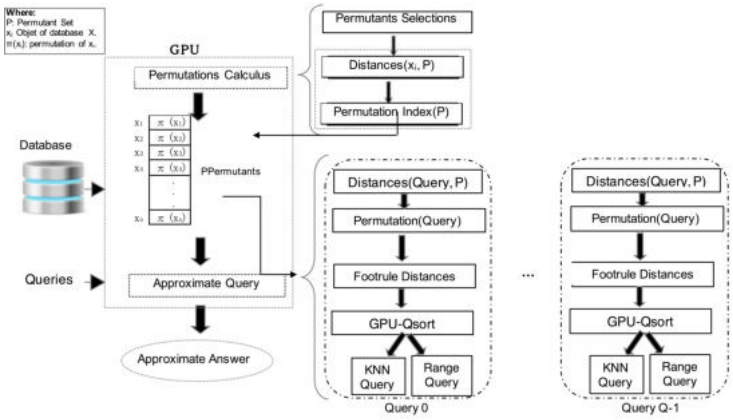


Fig. 2. Solving many queries in GPU-CUDA Permutation Index.

The number of queries to solve in parallel is determined according to the GPU resources, mainly its memory. If Q is the number of parallel queries, m the needed memory quantity per query and i the needed memory by the Permutation Index, $Q*m+i$ is the total required memory to solve Q queries in parallel. After Q parallel queries are solved, the results can be sent to CPU or they can be joined with other Q results and transfer them all together once via PCI-Express.

5 Experimental Results

Our experiments consider two metric databases selected from SISAP METRIC SPACE LIBRARY (www.sisap.org). The characteristics of each database are the following:

- English words(DBs): a set of English words. It uses the *Levenshtein* distance or *edit* distance.
- Colors histogram(DBh): a set of 112-vectors. It considers Euclidean distance.

In both cases, different DB size are considered, they are expressed in name: DBs or DBh + <DBsize> in kB.

The hardware scenario was:

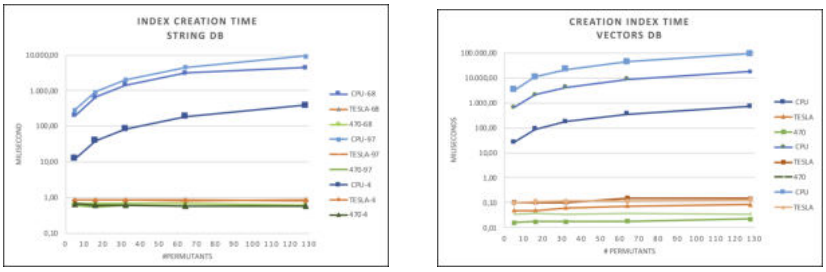
- CPU is an Intel(R) Xeon(R) CPU E5, 2603 v2 @1.80GHz x 8 and 15,6GB of memory.
- Two GPU are considered with the next characteristics (GPU Model, Memory, CUDA Cores, Clock Rate and Capability):
  - Tesla K20c, 4800 MB, 2496, 0.71 GHz, 3.5.

ISBN 978 -987-633-574-4

- GTX470, 1216 MB, 448, 1.22 GHz, 2.0.

The experiments consider for k-NN searches the values of k: 3 and 5; and for range the radii, for DBS: 1, 2, and 3, and DBh: 0,05, 0,08 and 0,13. For the parameter f of the Permutation Index, that indicates the fraction of DB revised during searches, we consider 10, 20, 30 and 50% of the DB size. The number of permutants used for the index are $5, 16, 32, 64$, and 128. In each case the results shown are the average over 1000 different queries.

In Figure 3, the Index Creation times for all the devices and for each BD are shown. We managed to increase the performance with respect to the CPU. Although only the times to build the index were taken into account in the time comparisons, the transfer time of the complete DB to the GPU was measured. In our case, both devices have the same PCI Express technology. For example the transfer time for BDs97 is 1.23 milliseconds. We can observe for the case of BDs, regarding the total creation time of the index, the load from the DB to the GPU implies 60% of the total process time in the Tesla K20c and 66% in the GTX 470.



(a) DBs (b) DBh

Fig. 3. Time of Index Creation for two DB.

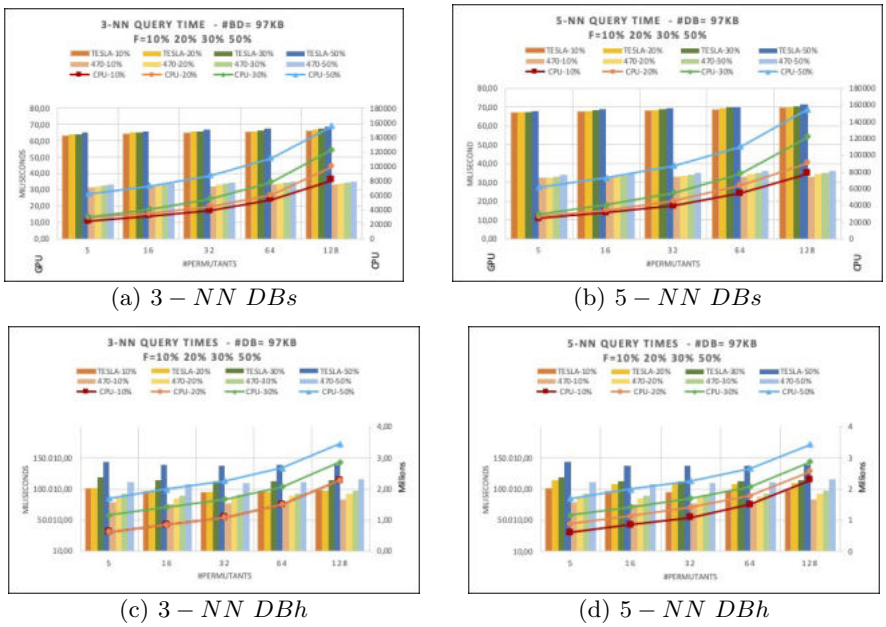
Figures 4 and 5 show the obtained time in k-NN and range queries respectively, for different parameters: permutants number, range, k, and DB percentage. In these results, 80 queries are solved in parallel. As it can be noticed *Range queries* show improvements respect to k-NN queries, but in both cases the achieved times are much less than CPU times. In all cases, it is clear the influence of DB size, but evenly we accomplish good performance. In all cases, the permutants number does not influence the time.

In Figure 6, we can see how the queries number to be solved in parallel influences the performance. Shorter times are achieved when queries number is greater. For BDh, the time to solve 1 (one) query vs 30 (thirty) queries decreases in the best of cases in the order of 1.8x (Tp1 / Tp30). In the case of the BDs, the gains obtained in solving multiple queries in parallel are greater: an improvement of 2.5x.

For the case of k-NN, the times are similar. This behavior is similar in both DB, i.e. the GPU resources have more work to do and, consequently, less idle time.

The trade-off between the answer quality and time performance of our parallel index with respect to the sequential index. For each k-NN or range query we have previously obtained the exact answer, that is $Rel()$, and we obtain the approximate answer $Retr()$. Figures 7 and 8 illustrate the average quality answer obtained for both kinds of queries, considering the Permutation Index respectively with 5,64 and 128 permutants, and different DB percentages. As it can

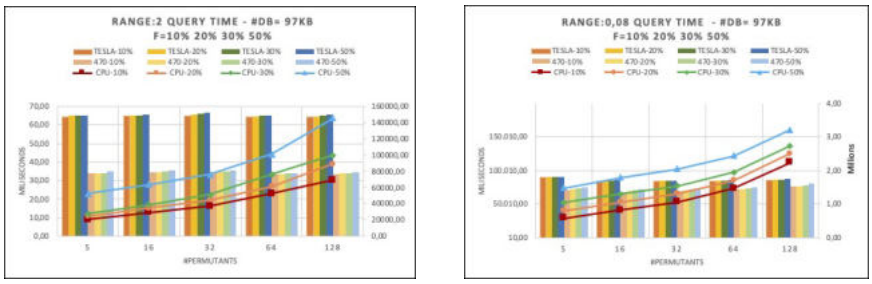
ISBN 978 -987-633-574-4



(a) $3-NN\ DBs$ (b) $5-NN\ DBs$

(c) $3-NN\ DBh$ (d) $5-NN\ DBh$

Fig. 4. k-NN Query Time for two DB.



(a) 2 Range DBs (b) 0,08 Range DBh

Fig. 5. Range Query Time for two DB.

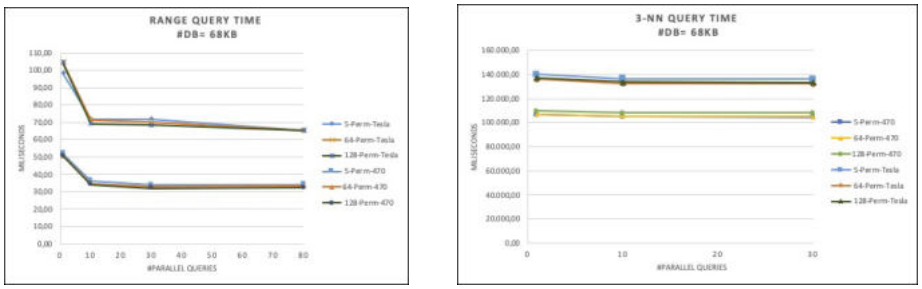
be noticed, the Permutation Index retrieves a good percentage of exact answer only reviewing a little fraction of the DB. For example, the 10% retrieves 40% and it needs to review the 30% to retrieve almost 80% of exact answer.

For lack of space, despite of we have tested another database sizes, we show only for the biggest database. In the other sizes have yielded similar results.

6 Conclusions

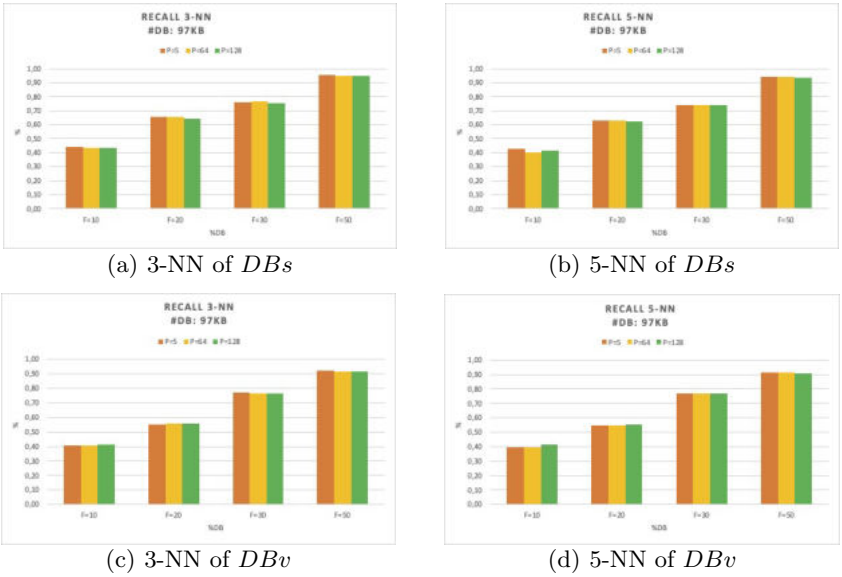
When we work with databases into large-scale systems such as Web Search Engines, it is not enough to speed up the answer time of only one query, but it is necessary to answer several queries at the same time. In this work, we present

ISBN 978 -987-633-574-4



(a) DBs (b) DBh

Fig. 6. Multi-Queries Time for two DB.



(a) 3-NN of DBs (b) 5-NN of DBs

(c) 3-NN of DBv (d) 5-NN of DBv

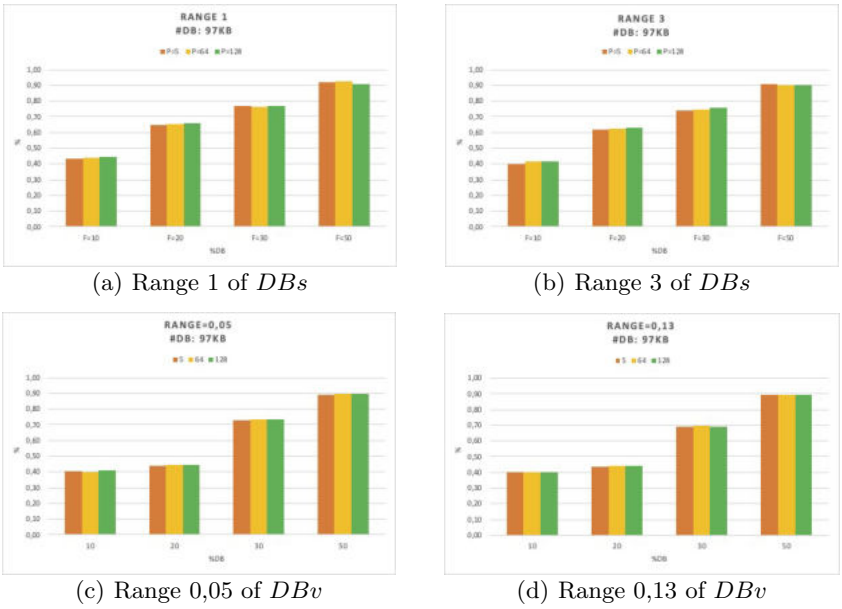
Fig. 7. Recall of approximate-k-NN queries for two DB.

a reliable solution to solve many queries in parallel, verifying the correctness of obtained results. This solution takes advantage of GPU and its high throughput: parallel processing for thousands of threads.

We check GPU-*Permutation Index* performance to the different GPUs. All accomplished performance results are very good, independently of the GPU architecture, because of careful planning and correct use of GPU resources. The index showed a good performance, allowing us to increase the fraction f of the database that will be examined to obtain better and accurate approximate results. An extensive validation process is carried out to guarantee the quality of the solution provided by the GPU.

In the future, we plan to make an exhaustive experimental evaluation, considering other types of databases and other solutions that apply GPUs to solve similarity searches in metric spaces; extend our proposal to other metric databases such as documents, DNA sequences, images, music, among others, and use other

ISBN 978 -987-633-574-4



(a) Range 1 of DBs (b) Range 3 of DBs

(c) Range 0,05 of DBv (d) Range 0,13 of DBv

Fig. 8. Recall of approximate-range queries for two DB.

distance functions. Another point to consider is to work with larger DBs, mainly when they are larger than the GPU memory. In this case, it is necessary to study strategies to partition the databases and/or use several GPUs. Besides, we plan to consider the Permutation's Signatures [24] to reduce the size of Permutation Index without removing any permutant.

References

1. E. Chávez, G. Navarro, R. Baeza-Yates, and J. Marroquín, "Searching in metric spaces," *ACM Comput. Surv.*, vol. 33, no. 3, pp. 273–321, 2001.
2. E. Chávez, K. Figueroa, and G. Navarro, "Proximity searching in high dimensional spaces with a proximity preserving order," in *Proc. 4th Mexican International Conference on Artificial Intelligence (MICAI)*, ser. LNAI 3789, 2005, pp. 405–414.
3. P. Ciaccia and M. Patella, "Approximate and probabilistic methods," *SIGSPATIAL Special*, vol. 2, no. 2, pp. 16–19, Jul. 2010. [Online]. Available: http://doi.acm.org/10.1145/1862413.1862418
4. P. Zezula, G. Amato, V. Dohnal, and M. Batko, *Similarity Search: The Metric Space Approach*, ser. Advances in Database Systems, vol.32. Springer, 2006.
5. P. Pacheco and M. Malensek, *An Introduction to Parallel Programming*. Elsevier Science, 2019. [Online]. Available: https://books.google.com.ar/books?id=uAfXnQAACAAJ
6. R. Robey and Y. Zamora, *Parallel and High Performance Computing*. Manning Publications, 2021. [Online]. Available: https://books.google.com.ar/books?id=jNstEAAAQBAJ
7. D. Kirk and W. Hwu, *Programming Massively Parallel Processors: A Hands-on Approach*. Elsevier Science, 2016. [Online]. Available: https://books.google.com.ar/books?id=wcS_DAAAQBAJ
8. R. Barrientos, F. Millaguir, J. L. Sánchez, and E. Arias, "Gpu-based exhaustive algorithms processing knn queries," *The Journal of Supercomputing*, vol. 73, pp. 4611–4634, 2017.

9. M. Kruliš, H. Osipyan, and S. Marchand-Maillet, "Employing gpu architectures for permutation-based indexing," *Multimedia Tools and Applications*, vol. 76, 05 2017.
10. S. Li and N. Amenta, "Brute-force k-nearest neighbors search on the gpu," in *Similarity Search and Applications*, G. Amato, R. Connor, F. Falchi, and C. Gennaro, Eds. Cham: Springer International Publishing, 2015, pp. 259–270.
11. P. Velentzas., M. Vassilakopoulos., and A. Corral., "In-memory k nearest neighbor gpu-based query processing," in *Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management - GISTAM,*, INSTICC. SciTePress, 2020, pp. 310–317.
12. K. Figueroa, E. Chávez, G. Navarro, and R. Paredes, "Speeding up spatial approximation search in metric spaces," *ACM Journal of Experimental Algorithmics*, vol. 14, p. article 3.6, 2009.
13. B. Bustos and G. Navarro, "Probabilistic proximity searching algorithms based on compact partitions," *Discrete Algorithms*, vol. 2, no. 1, pp. 115–134, Mar. 2004. [Online]. Available: http://dx.doi.org/10.1016/S1570-8667(03)00067-4
14. K. Tokoro, K. Yamaguchi, and S. Masuda, "Improvements of tlaesa nearest neighbour search algorithm and extension to approximation search," in *Proceedings of the 29th Australasian Computer Science Conference - Volume 48*, ser. ACSC '06. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2006, pp. 77–83. [Online]. Available: http://dl.acm.org/citation.cfm?id=1151699.1151709
15. A. Singh, H. Ferhatosmanoglu, and A. Tosun, "High dimensional reverse nearest neighbor queries," in *The twelfth international conference on Information and knowledge management*, ser. CIKM '03. New York, NY, USA: ACM, 2003, pp. 91–98. [Online]. Available: http://doi.acm.org/10.1145/956863.956882
16. F. Moreno-Seco, L. Micó, and J. Oncina, "A modification of the laesa algorithm for approximated k-nn classification," *Pattern Recognition Letters*, vol. 24, no. 1–3, pp. 47 – 53, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865502001873
17. R. A. Baeza-Yates and B. A. Ribeiro-Neto, *Modern Information Retrieval - the concepts and technology behind search, Second edition.* Pearson Education Ltd., Harlow, England, 2011.
18. J. Cheng and M. Grossman, *Professional Cuda C Programming.* CreateSpace Independent Publishing Platform, 2017. [Online]. Available: https://books.google.com.ar/books?id=BcjItAEACAAJ
19. J. Han and B. Sharma, *Learn CUDA Programming: A beginner's guide to GPU programming and parallel computing with CUDA 10.x and C/C++.* Packt Publishing, 2019. [Online]. Available: https://books.google.com.ar/books?id=dhWzDwAAQBAJ
20. D. B. Kirk and W. W. Hwu, *Programming Massively Parallel Processors, A Hands on Approach.* Elsevier, Morgan Kaufmann, 2010.
21. NVIDIA, "Nvidia cuda compute unified device architecture, programming guide," in *NVIDIA*, 2020.
22. R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," in *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, ser. SODA '03. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2003, pp. 28–36. [Online]. Available: http://dl.acm.org/citation.cfm?id=644108.644113
23. M. Lopresti, N. Miranda, F. Piccoli, and N. Reyes, "Permutation Index and GPU to Solve efficiently Many Queries," in *VI Latin American Symposium on High Performance Computing, HPCLatAm 2013*, 2013, pp. 101–112.
24. K. Figueroa and N. Reyes, "Permutation's signatures for proximity searching in metric spaces," in *Similarity Search and Applications*, G. Amato, C. Gennaro, V. Oria, and M. Radovanović, Eds. Cham: Springer International Publishing, 2019, pp. 151–159.

ISBN 978 -987-633-574-4

From Global to Local in the Sneakers Universe: A Data Science Approach

Luciano Perdomo and Leo Ordinez

Laboratorio de Investigación en Informática (LINVI), FI - UNPSJB
Bvd. Brown 3051, Puerto Madryn, Argentina
{lucianor.perdomo,leo.ordinez}@gmail.com

Abstract. In Argentina there was a great growth of e-commerce due to the COVID-19 pandemic. With the aim of helping local companies to understand the market and help them in decision making, data were obtained from online shoe sales sites and with them Machine Learning models were implemented to make price predictions in sneakers. It was concluded that higher-tier companies have greater competitive advantage over lower-tier companies. Nonetheless, the cost-effective methodology used would aid local companies scale up.

Keywords: E-commerce, Machine Learning, Linear Regression, Random Forest, LGBM Regressor

1 Introduction

According to CACE (Argentine Chamber of Electronic Commerce), during 2020, E-commerce turnover in Argentina grew 124% due to the COVID-19 pandemic, compared to the previous year, for a total of ARS 905.143 million, corresponding to more than 164 million purchase orders. The category that grew the most was clothing and sports articles, which in 2019 ranked 4th, and in 2018 3rd [2]. Based on this nation-wide tendency, which replicates also global tendencies towards e-commerce [3,4], an exploratory study was conducted to measure de impact of e-commerce on local stores. In particular, the sector chosen was sneakers (within shoes and clothing) and the territorial scale of the local context is the Patagonian zone in Argentina.

The aim of this research is to build knowledge around products sold through e-commerce channels, which can be leveraged by small local companies, that are joining global tendencies, for decision-making. In particular, through the use of cost-effective tools and information available on the websites of different competitors in distinct scales. This is, the analysis is multiscalar, which is not an impediment considering that e-commerce is inherently horizontal in terms of customers access.

Since the nature of this exploratory analysis is to obtain information publicly available without any intervention inside companies (*e.g.*, asking for sales information) nor action with customers (*e.g.*, surveying preferences), the main variable considered for the products is *price*. In order to predict sneaker prices

ISBN 978 -987-633-574-4

[8], linear regression will be used with price as the dependent variable and gender, brand and company as independent variables. Three experiments were designed. For each experiment different models of Machine Learning [1] are compared and the one with the best results is selected to be optimized and trained. Then, comparisons are made between the models selected above. To make the predictions, efficient models were selected in terms of execution time and resources, and effective in terms of the results.

The rest of the work is organized as follows: in Section 2 the methodology used is outlined; in Section 3 descriptive and inferential results are exposed; a discussion on those results is presented in Section 4; finally, conclusions are drawn in Section 5.

2 Materials and Methods

The standard methodology used in different domains and contexts [9,6] involves understanding the problem; selecting the analytical approach to use depending on the type of research to be carried out at that time; the definition of requirements, the collection and characterization of the data, in an iterative refinement process; the preparation of the data to be able to be worked under the proposed analytical approach, which involves another iterative sub-process of modeling; and the evaluation of the model, which implies its validation by domain experts. After passing the evaluation instance, the model is deployed in an environment available to be accessed. Finally, based on the knowledge obtained, the techniques developed and the products generated in the previous steps, the goal is to obtain learning that promotes better decision-making.

As previously mentioned, we are interested in analyzing Patagonian online sneakers stores in the context of bigger scales, such as nation-wide or globally. Depending on the scope of the company that owns the e-commerce site, it was decided to categorize them as local, national or regional and global. For this, the cities where the stores are physically located and the number of branch offices were considered, where it applies. This is, in the first place, we consider companies which have a physical store; and secondly, in some cases the number of branch offices was not taken as a limit for categorizing but an indicator, and the reach of their marketing strategy was considered (*i.e.*, advertising in international sports events). Seven sites were selected out of seventeen. Globally, Stockcenter, a subsidiary of NetShoes, was chosen. At the National level, Dash and Solodeportes were selected. At a regional level, Sporting. Finally, at the local level, Ferreira (Bahía Blanca and South West of Buenos Aires Province), Quonam (Chubut, Patagonia) and Newsport (Córdoba) were considered. Each company has different types of shoe offerings (Men, Women, Unisex, Children, Boy/Girl). For simplicity and homogeneity of data, it was decided to analyze offerings by categories “Women” and “Men”.

The scraping tools used were parseHub (for Quonam and Stockcenter) and the Python library Beautiful Soup for the rest of the sites. The following data

ISBN 978 -987-633-574-4

were obtained *brand*, *model*, *list price* (price without discount), *net price* (price with discount) and *sex*.

The dataset was cleaned and structured as follows: *brand*, represents the brand of the sneaker; *footwear*, represents the type of shoe; *sex*, women or men; *original_model*, text from the original dataset, that is, without parsing; *model*, parsed text; *net_price*, discounted price applied; *list_price*, price without discount applied; *item discount*, percentage of discount applied; *company*, name of the company where the data was extracted.

As said before, the tools used were ParseHub, Python3, and the following Python packages: Jupyter-Notebook, Pandas, BeautifulSoup, ScikitLearn, Seaborn, plotly-express, XGBoost, LightGBM, yellowbrick, hyperopt.

3 Results

The scraping was carried out on March 19, 2021. Then the data set was cleaned up and structured. The exploratory data analysis was performed. Then the predictive analysis was performed, in which three experiments were carried out to predict sneakers prices.

3.1 Exploratory Data Analysis

In the first place, we considered the variables list price, brands, sex and company. A comparison among them is presented in Fig. 1.

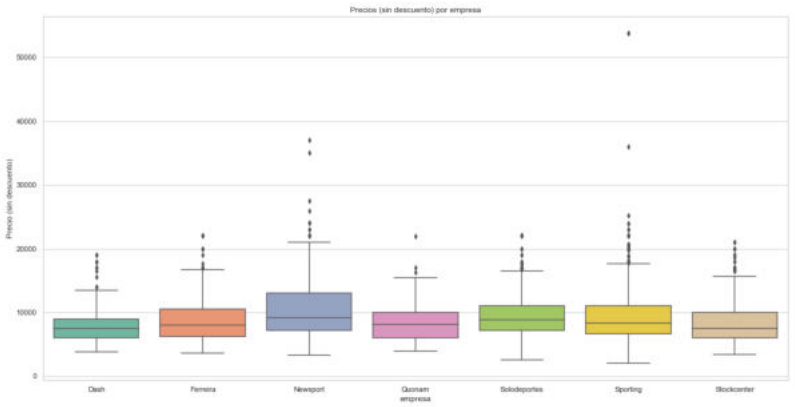


Fig. 1. List prices per company.

Fig. 2 shows the amount of sneakers that each company offers by sex.

ISBN 978 -987-633-574-4

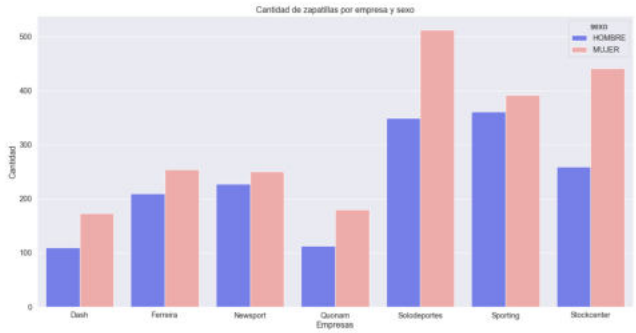


Fig. 2. Amount of sneakers by company and sex.

The amount of sneakers with prices ranging from $2.000 to $30.000 is presented in Fig. 3.

The distribution of prices within the companies is presented in Fig. 4.

Finally, a comparison of characteristics among each company is performed by a radar chart and presented in Fig. 5. The characteristics are as follows:

- Variables: Maximum price, quantity of men's sneakers, quantity of women's sneakers, number of brands, brand dispersion (represented as "HHI Marcas").
- For the dispersion of marks, the Blau Index[10] was used, which quantifies the probability that two individuals taken at random from a population are in different categories of one variable.
- The data was scaled to be in an approximate range of 100 to 1.000, for visualization purposes.
  - The max prices were divided by 100.
  - The Blau index was multiplied by 1,000.
  - The number of brands multiplied by 10.
- With all these data, a Radar Chart was made for each company. Then an overlay radar chart was made to compare all companies.

3.2 Predictive Analysis

In the first place, outliers were removed from the data set, leaving a maximum price of $25.000 and a minimum of $4.000. In addition, brands that have less than 40 items were removed, resulting in 92.09% of the data set. With this, 75% of the data was used for training. Training and test data subsets were saved with the Python library Pickle.

In all cases, it is used as a dependent variable the list price and as independent variables, in turn, brand, company and sex. The following experiments were conducted:

ISBN 978 -987-633-574-4

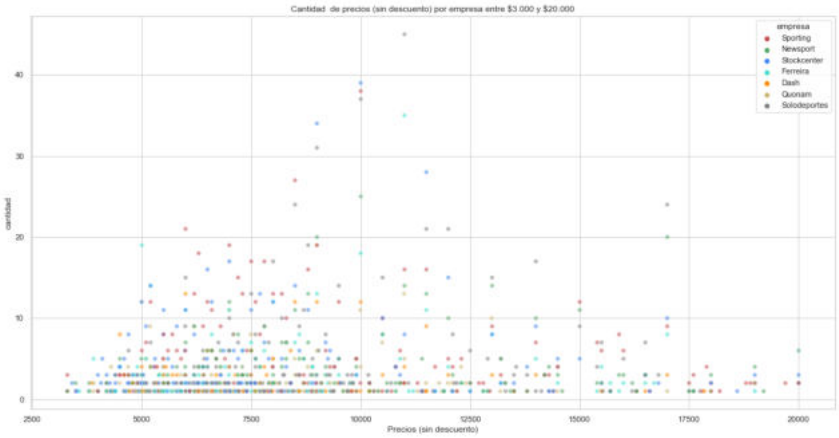


Fig. 3. Amount of prices per company.

1. Comparison of linear models: Linear Regression, Ridge Regression and SGD Regression.
2. Comparison of: Random Forest, XGBoost and Decision Tree Regressor.
3. Comparison of: SVM Regressor, Random Forest and Light GBM Regressor.

Finally, a comparison of the experiments was carried out.

Experiment 1 At first, the Recursive Feature Elimination (RFE) model was used to obtain the number of optimal features, but the idea was discarded, and it was decided to train the model with 3 features, then 2 and finally 1, and then make comparisons. The models were found to better fit a polynomial function.

Ridge Regression It is similar to linear regression, but uses L2 regularization. Hyperparameters can be adjusted to find the correct alpha value, which is the parameter with which you can make the model perform overfitting or underfitting.

The alpha parameter is searched with the yellowbrick library and a value of 1.6907141034735782 was obtained. It was also found that a polynomial function of degree 11 fits the model better, to create a Polynomial Ridge. Iterating between the three features, the best r2 that was obtained was 33.32% with 2 features (brand, sex).

SGD (Stochastic Gradient Descent) Regression It is a linear model that uses L2 regularization and minimize empirical loss with SGD (loss gradient is estimated for each sample and the model is updated with the learning rate). It is better suited to linear models than Ridge Regression and Linear Regression.

ISBN 978 -987-633-574-4

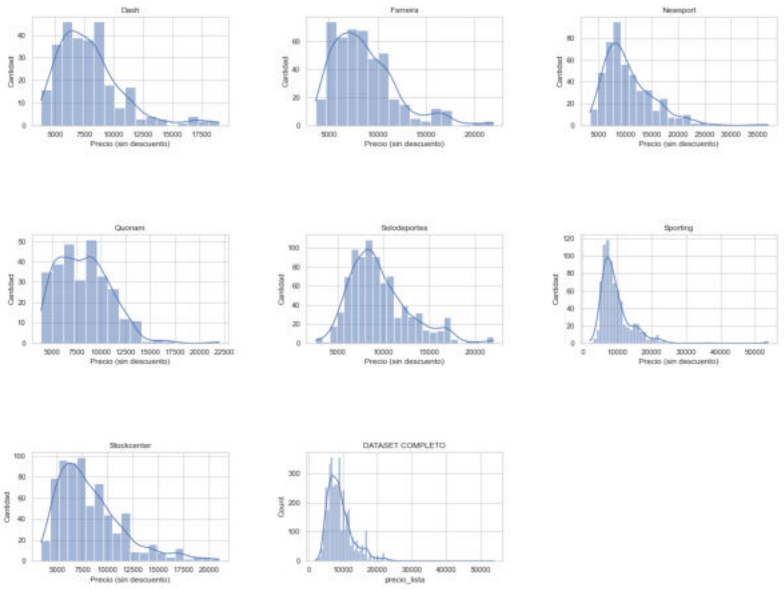


Fig. 4. Distribution of prices.

With the default model and data, a r2 of 2% was obtained. Finally, it was iterated with the 3 features, then with 2 and last with 1, and the polynomial function with 14 degrees. The best r2 result was 6.28% with 2 features.

Lineal Regression The model was trained with a polynomial function of degree 10, giving the following results:

- 1 feature (brand): r2 = 49.47%
- 2 features (brand, sex): r2 = 49.66%
- 3 features (brand, sex, company): r2 = 50.46%

Based on this, it was decided to train each company with a different model, taking into account that the results of linear regression are better than the other two models (Ridge and SGD);

One model was made per company, with polynomial degree 10 and using brand and sex as features.

Following, companies and best r2 (with one or two features) are shown:

- Dash: 1 feature r2=59.11% MAE=0.0600 MSE=0.0082
- Ferreira: 1 feature r2=24.41% MAE=0.0973 MSE=0.0228
- Newsport: 2 features r2=37.1% MAE=0.1331 MSE=0.0317

ISBN 978 -987-633-574-4

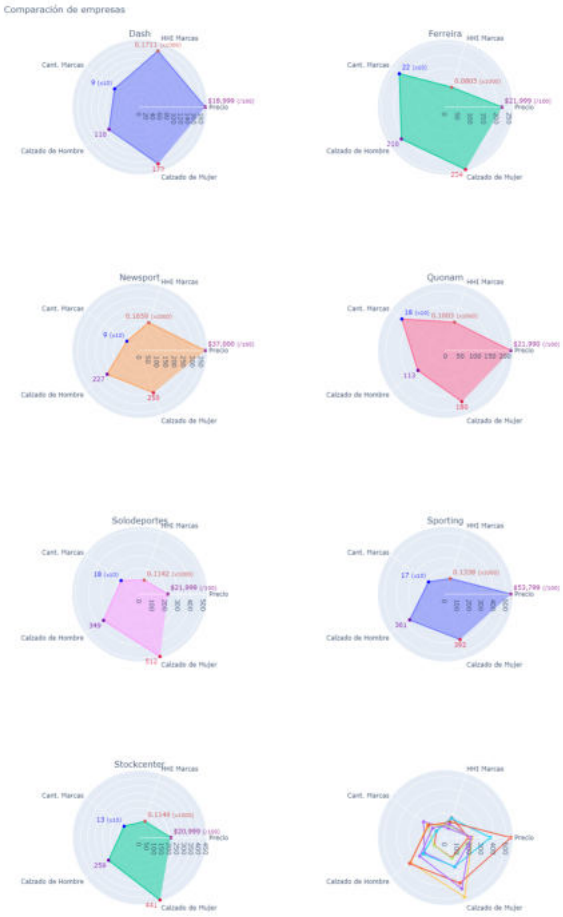


Fig. 5. Radar Chart with all companies.

- Quonam: 1 feature r2=53.66% MAE=0.0634 MSE=0.0068
- Solodeportes: 1 feature r2=33.03% MAE=0.0937 MSE=0.0140
- Sporting: 1 feature r2=44.32% MAE=0.1006 MSE=0.0187
- Stockcenter: 1 feature r2=68.35% MAE=0.0635 MSE=0.0069

Experiment 2 This experiment is based on [7]. Here, Random Forest, XGBoost and Decision Tree Regressor were chosen. Random Forest because each tree draws a different sample, avoiding overfitting and improving the accuracy of predictions. XGBoost (Extreme Gradient Boosting) which, as a Gradient Boosting algorithm, generalize Boosting models as I/O to get better models (they are trained sequentially). Finally Decision Tree Regressor is used as a comparison against the previous two, besides being simple and effective.

ISBN 978 -987-633-574-4

The models were compared, each one with its default parameters. The results were:

- Decision Tree Regressor r2 = 50.20%
- XGBoost r2 = 50.5204% (chosen model)
- Random Forest r2 = 50.5157% (secondary model)

XGBoost Regressor First we tried using the polynomial function with 3 and 6 degrees, but in no case was an r2 of 43.7% exceeded, a value lower than the r2 of 50.52% of the previous comparison, so the hyperopt library was used for parameter optimization, along with three sets of different parameters to test the model and performing tests using one, two and three features: (brand), (brand, sex) and (brand, sex, company).

The best result of all the tests was r2 of 50.32%, with three features, no polynomial function. Since the optimization did not work better than the default parameters, it was decided to try Random Forest, which in the initial comparison yielded similar values.

Random Forest Regression Since there was little difference, it was decided to use it in the experiment. With the default parameters, an R2 = 50.49% was obtained. The search for hyperparameters was carried out with hyperopt. Four different parameters were used; without polynomial function, and with polynomial function of 3 and 6 degrees; with one, two and three features; and parameter max_evals=100 (hyperopt). The best r2 was 51,237%; with 3 features, polynomial function of degree 6. Because it gives better results than XGBoost, Random Forest is used for the iteration of each company, with the parameters that hyperopt showed in the winning test.

Companies and best r2 (with one or two features):

- Dash: 1 feature r2=58.85% MAE=0.0603 MSE=0.0083
- Ferreira: 1 feature r2=25.7% MAE=0.0948 MSE=0.0224
- Newsport: 2 features r2=37.74% MAE=0.1301 MSE=0.0314
- Quonam: 1 feature r2=53.85% MAE=0.0629 MSE=0.0068
- Solodeportes: 1 feature r2=34.56% MAE=0.092 MSE=0.01368
- Sporting: 1 feature r2=44.09% MAE=0.1011 MSE=0.0188
- Stockcenter: 1 feature r2=68.33% MAE=0.0632 MSE=0.0069

Experiment 3 It is based on [5]. Here, SVM, Random Forest and LightGBM Regressor were chosen. SVM is more accurate than Linear Regression and by default it uses a linear RGB kernel. Random Forest, for the same reasons as the previous experiment it is used as an indicator within this experiment, due to its use in the previous experiment. LightGBM Regressor, is a Gradient Boosting model, similar to XGBoost, uses algorithms based on decision trees.

The models were compared with the default parameters:

- SVM r2 = 0.1638
- Random Forest r2 = 0.5046 (secondary model)
- LGBM r2= 0.5084 (chosen model)

ISBN 978 -987-633-574-4

LightGBM Regressor For parameter optimization, first, we tried to obtain r2 with the polynomial function of degree 3 (r2=50.74%) and degree 5 (r2=50.67%). The previous comparison (without polynomial function) gives a better result than XGBoost in the previous experiment. Hyperopt was used to find the best hyperparameters, three sets of different parameters, with one, two and three features, and the variable max_evals=100 were used. The best result was with 3 features and without a polynomial function. Because the model yields a promising r2 value, it is used to make predictions for each company and then compare results.

Companies and best r2 (with one or two features) with default parameters:

- Dash: 1 feature r2=59.17% MAE=0.0606 MSE=0.0082
- Ferreira: 1 feature r2=22.29% MAE=0.1001 MSE=0.0234
- Newsport: 2 features r2=34.31% MAE=0.1357 MSE=0.0332
- Quonam: 1 feature r2=15.18% MAE=0.0899 MSE(0.0125)
- Solodeportes: 1f r2=32.64% MAE=0.0944 MSE=0.0140
- Sporting: 1 feature r2=42.11% MAE=0.1021 MSE=0.0194
- Stockcenter: 1 feature r2=64.03% MAE=0.0659 MSE=0.0078

Companies and best r2 (with three features) using optimized parameters:

- Dash: 2 features r2=59.24% MAE=0.0606 MSE=0.0082
- Ferreira: 1 feature r2=22.75% MAE=0.0997 MSE=0.0233
- Newsport: 2 features r2=34.98% MAE=0.1349 MSE=0.0328
- Quonam: 1 features r2=15.28% MAE=0.0899 MSE=0.0125
- Solodeportes: 1 feature r2=33.34% MAE=0.0936 MSE=0.0139
- Sporting: 2 features r2=42.12% MAE=0.1021 MSE=0.0194
- Stockcenter: 1 feature r2=68.35% MAE=0.0635 MSE=0.0069

4 Discussion

It can be seen in the comparison of the companies, that the determination coefficient in the Stockcenter predictions (global category), is the most accurate of all.

Dash and Sporting companies have National or Regional category. Dash is in second place and Sporting in the last experiment takes third place and fourth place in the first two.

The companies at the local level are Ferreira, Quonam and Newsport. Quonam is third in the first two experiments and last in the third one. The companies Newsport, Solodeportes (regional) and Ferreira maintain their order in all the experiments, being fifth, sixth and seventh in the first two and gaining a position in the last.

Except for Quonam company, in the first two experiments, it is true that the higher order companies have better price predictions. This company may have been benefited from cleaning out the outliers in the dataset when training the models.

ISBN 978 -987-633-574-4

On the other hand, the number of brands offered by each company and the number of sneakers per brand, allow companies to compete in different market segments. However, as shown in the comparison of prices of the exploratory analysis, there are no big differences in the dispersion of them. This may be because certain brands impose to be narrowed to certain price ranges.

5 Conclusions

In this work, a data science approach was performed over a local market sector in the context of bigger scale competitors. Although the case study was specific, such as online sale of sneakers, the methodology used was proven to be cost-effective and adaptable to other situations. With that, an analysis of a small particular company can be performed.

Including sales data to the analysis would allow the definition of better marketing strategies. Thus allowing local companies to start competing with national companies; and the national ones with the global ones.

On the other hand, the work in this paper would allow the sneaker buyer to make a more sound decision when buying, not lead by advertisements and publicity.

As future work, it is possible to scrape data weekly from each of the e-commerce sites to create a data warehouse and also collecting more information about the sneakers (such as color, sizes, etc.). Insights can be obtained, for example on the brand/price relationship or between brands.

References

1. B. Boehmke and B. Greenwell. *Hands-on machine learning with R.* Chapman and Hall/CRC, 2019.
2. CACE. Estudio anual de comercio electrónico 2020.
3. W. E. Forum. Covid-19 has reshaped last-mile logistics, with e-commerce deliveries rising 25
4. M. Keenan. Global ecommerce explained: Stats and trends to watch in 2021, 05 2021.
5. A. Kumar. Price prediction using machine learning regression — a case study.
6. I. Martinez, E. Viles, and I. G. Olaizola. Data science methodologies: Current challenges and future approaches. *Big Data Research*, 24:100183, 2021.
7. L. Norman. Predicting stockx sneaker prices with machine learning.
8. D. Raditya, N. E. P, F. A. S, and N. Hanafiah. Predicting sneaker resale prices using machine learning. *Procedia Computer Science*, 179:533–540, 2021. 5th International Conference on Computer Science and Computational Intelligence 2020.
9. J. B. Rollins. Foundational methodology for data science. IBM Analytics, 2015.
10. A. Solanas, R. Selvam, J. Navarro, and D. Leiva. Some common indices of group diversity: Upper boundaries. *Psychological reports*, 111:777–96, 12 2012.

ISBN 978 -987-633-574-4

Un Análisis Experimental de Sistemas de Gestión de Bases de Datos para Dispositivos Móviles

Fernando Tesone, Pablo Thomas, Luciano Marrero, Verena Olsowy, Patricia Pesado

Instituto de Investigación en Informática LIDI,
Facultad de Informática, Universidad Nacional de La Plata
La Plata, Argentina
{ftesone,pthomas,lmarrero,volsowy,ppesado}@lidi.info.unlp.edu.ar

Resumen Con el crecimiento en el alcance y uso de internet, de los smartphones, y de las redes sociales, se está produciendo un aumento exponencial en el volumen de datos administrados, pudiendo ser éstos estructurados, semiestructurados, o sin estructura. En este contexto surgen las bases de datos NoSQL, que facilitan el almacenamiento de datos semiestructurados o sin estructura.
Por otra parte, las mejoras en las prestaciones de hardware de los dispositivos móviles conducen a que éstos administren cada vez más información, y que surjan nuevos sistemas de gestión de bases de datos que se instalan en dichos dispositivos. Este trabajo tiene por objetivo realizar un relevamiento de los sistemas de gestión de bases de datos para dispositivos móviles, y realizar un análisis experimental de los sistemas más representativos de los modelos de bases de datos más utilizados.

Palabras clave: Bases de Datos para Dispositivos Móviles, DBMS Relacional, DBMS NoSQL

1. Introducción

Los sistemas de gestión de bases de datos (DBMS, por su sigla en inglés correspondiente a *Database Management System*) jugaron un rol fundamental en el desarrollo de software desde su surgimiento en la década de 1960, ya que proveían una forma eficiente de generar aplicaciones complejas, al eliminar la necesidad de programar la persistencia y el acceso a los datos [1,2].

En 1970 Edgar Codd desarrolla el modelo de base de datos relacional, que a partir de entonces, y hasta la actualidad, se volvió el modelo dominante [1,3].

Se presentan en 2007 el *Apple iPhone* y en 2008 el sistema operativo *Android*, hechos que cambiarían radicalmente la industria de los *smartphones*, ya que incrementaron la popularidad del uso de los dispositivos móviles, llegando a ser usados por el 80 % de la población en algunos países [4,5,6]. Este crecimiento en el uso traería acoplado una diversificación de plataformas. Para maximizar la presencia en el mercado, las *apps* deben estar disponibles en múltiples plataformas o sistemas operativos, por lo que los desarrolladores de software deben

ISBN 978 -987-633-574-4

optar por realizar desarrollos nativos, específicos de cada plataforma, o desarrollos multiplataforma [7,8].

Con el crecimiento en el alcance y uso de internet y de los dispositivos móviles, sumado a la aparición de las redes sociales, se está produciendo un crecimiento exponencial en el volúmen de datos administrados [9], pudiendo ser éstos estructurados, semi–estructurados o sin estructura. Ante esta situación de crecimiento en el volúmen de información, surgen las bases de datos no relacionales o NoSQL (Not-only SQL) como alternativa a las bases de datos relacionales, que facilitan el almacenamiento masivo de datos semi–estructurados o no estructurados.

Con la aparición de estas tecnologías, los desarrolladores de software deben analizar cuáles DBMSs son adecuados para las necesidades del problema a resolver.

El objetivo de este trabajo es (1) realizar un relevamiento de los DBMSs existentes, tanto relacionales como no relacionales, para dispositivos móviles —es decir, que pueden ser embebidos en aplicaciones para dispositivos móviles—, (2) seleccionar DBMSs que se consideren representativos del conjunto relevado, a partir de la definición de una serie de criterios, y (3) realizar un experimento que permita análizar, para cada DBMS seleccionado, características específicas, ventajas y desventajas, desde el punto de vista de la experiencia del ingeniero de software.

Este trabajo se organiza del siguiente modo: en la Sección 2 se discuten los trabajos relacionados; en la Sección 3 se presenta un relevamiento de sistemas de gestión de bases de datos para dispositivos móviles; en la Sección 4 se seleccionan DBMSs representativos, y se realiza una experimentación que permite analizar características específicas, y ventajas y desventajas de cada DBMS seleccionado. Posteriormente, en la Sección 5 se analizan los resultados obtenidos a partir de la experimentación realizada. Finalmente, en la Sección 6 se presentan las conclusiones y se definen posibles líneas de investigación como trabajo futuro.

2. Trabajos Relacionados

En esta sección se describen los trabajos relacionados encontrados, vinculados al tema presentado.

En [10] se expone un listado de características que los sistemas de gestión de bases de datos móviles y embebibles deben tener: ser distribuídos junto con la aplicación; minimizar el uso de memoria principal y secundaria; permitir incluir sólo los componentes del DBMS necesarios; soportar el almacenamiento en memoria principal; ser portables; ejecutarse en dispositivos móviles, y sincronizar datos con DBMSs de backend.

En [11] se realiza un relevamiento de diferentes opciones de almacenamiento en dispositivos móviles, siendo éstas: HTML5, a partir de la utilización del framework WebKit, posibilitando el uso de la API *localStorage*; SQLite, una librería que encapsula funcionalidad SQL y almacena la información en un archivo local; almacenamiento en la nube, utilizando servicios como *Apple iCloud*, *Google Drive*, *Dropbox*, *Amazon S3*; almacenamiento específico del dispositivo, como son *Shared Preferences* en Android y *Core Data* en iOS, sumados a las opciones anteriores, presentes en ambos sistemas.

ISBN 978 -987-633-574-4

En [12] se realiza una descripción de la arquitectura de la plataforma Android y se analiza, entre otras cuestiones, la arquitectura y la forma en que las aplicaciones se ejecutan. Finalmente, se hace una introducción al uso de SQLite en Android.

En [13] se analizan ventajas y desventajas del uso de computación en la nube de forma integral con las aplicaciones móviles, mencionando como ventajas el almacenamiento, los respaldos (*backups*), y la redundancia de datos, entre otras.

En los trabajos previamente descriptos, si bien se analizan temas relacionados al almacenamiento de datos en dispositivos móviles, no se establece un análisis de los distintos sistemas de gestión de bases para dispositivos móviles que permita al ingeniero de software seleccionar el DBMS más adecuado a utilizar para resolver un problema determinado, motivo por el cual se pretende cubrir este aspecto con la presentación de este artículo.

3. Bases de Datos para Dispositivos Móviles

Se presenta un relevamiento de DBMSs relacionales (RDBMSs, por su sigla en inglés correspondiente a *Relational Database Management System*) y DBMSs NoSQL que pueden utilizarse embebidos en aplicaciones móviles. La búsqueda de DBMSs a analizar se realizó a través de los buscadores Google y Google Scholar, utilizando los términos “mobile database”, “mobile dbms”, entre otros. También se realizó una búsqueda en Google con el término “mobile site:db-engines.com/en/system”, ya que el sitio DB-Engines utiliza ese *path* para las páginas en las que se describen características de cada DBMS relevado por el sitio.

DB-Engines se trata de un proyecto que busca recopilar y presentar información sobre DBMSs, creado y mantenido por solidIT, una compañía austríaca especializada en el desarrollo de software, consultoría y formación para aplicaciones centradas en datos. El sitio elabora un ranking mensual a partir de la puntuación que obtenga cada DBMS relevado según su popularidad, definida por diferentes parámetros [3].

3.1. DBMSs Relacionales

Los DBMSs relacionales o RDBMSs existentes para dispositivos móviles, ordenados de acuerdo al ranking elaborado por DB-Engines [3] son:

1. SQLite
2. Interbase
3. SAP SQL Anywhere
4. SQLBase

SQLite es una librería que implementa un motor de base de datos autocontenido (embebido). Tiene licencia Public Domain, y puede utilizarse tanto en el desarrollo nativo de aplicaciones, ya sea en Android o en iOS, como también en el desarrollo de aplicaciones multiplataforma en los enfoques híbrido, interpretado o de compilación cruzada [14].

Interbase, un RDBMS embebido, con licencia comercial, y conforme al estándar SQL. En cuanto a su uso en el desarrollo de aplicaciones móviles, se puede utilizar en el desarrollo nativo en Android y iOS [15].

ISBN 978 -987-633-574-4

SAP SQL Anywhere integra un paquete de DBMSs relacionales y tecnologías de sincronización para servidores, en entornos de escritorio y móviles [16]. Se encuentra disponible para utilizar en el desarrollo de aplicaciones móviles nativas, tanto en Android como en iOS.

SQLBase es un RDBMS desarrollado por la empresa Opentext. Tiene licencia de uso comercial. Se encuentra disponible para el desarrollo de aplicaciones móviles nativas, en Android y iOS [17].

3.2. DBMSs NoSQL

El término NoSQL se utiliza para denominar a bases de datos de distintos modelos, diferentes al modelo relacional. Dentro de los distintos modelos de bases de datos NoSQL existentes, el modelo documental representa al más utilizado en la actualidad [3]. De los DBMSs NoSQL existentes para dispositivos móviles, la mayoría corresponde al modelo documental.

Los DBMSs NoSQL para dispositivos móviles existentes, ordenados de acuerdo al ranking DB-Engines son:

1. Couchbase Lite (Documental)
2. Firebase Realtime Database (Documental)
3. Realm (Documental)
4. Google Cloud Firestore (Documental)
5. Oracle Berkeley DB (Clave-valor)
6. PouchDB (Documental)
7. LiteDB (Documental)
8. ObjectBox (Orientada a objetos)
9. Sparksee (Grafos)

DBMSs Documentales

Couchbase Lite es un DBMS embebido, que utiliza JSON como formato de los documentos. Al formar parte del paquete Couchbase, es posible utilizar el sistema *Sync Gateway* para sincronizar datos con bases de datos remotas [18]. Tiene licencia dual, y es posible utilizarlo para el desarrollo de aplicaciones móviles nativas Android y iOS, y multiplataforma en los enfoques híbrido, interpretado, y de compilación cruzada [3].

Firebase Realtime Database es un DBMS alojado en la nube, que almacena los datos en un único JSON, y cuenta con sincronización de datos en tiempo real con todos los clientes conectados, manteniéndose disponibles aún sin conexión. Se encuentra disponible para su uso en el desarrollo de aplicaciones nativas en Android y en iOS, y en el desarrollo de aplicaciones móviles multiplataforma en enfoques híbridos, interpretados, y de compilación cruzada [19].

Realm es un DBMS embebido que utiliza un modelo de datos orientado a objetos. Es posible utilizarlo de forma autónoma, o de forma sincronizada con una base de datos *backend* MongoDB. Se encuentra disponible para el desarrollo de aplicaciones nativas Android y iOS, y el desarrollo de aplicaciones móviles multiplataforma en enfoques interpretado y de compilación cruzada [3, 20].

Google Cloud Firestore es un DBMS alojado en la nube, que almacena los datos en documentos JSON y cuenta con sincronización de datos en tiempo

ISBN 978 -987-633-574-4

real con los clientes, manteniéndo los datos disponibles aún sin conexión. Es posible utilizarlo en el desarrollo de aplicaciones móviles nativas Android y iOS, y en el desarrollo de aplicaciones móviles multiplataforma con enfoques híbrido, interpretado, y de compilación cruzada [21].

PouchDB es una librería javascript que implementa un DBMS inspirado en CouchDB, y que utiliza su protocolo de sincronización. Es distribuido bajo licencia Apache 2.0 y se encuentra disponible para el desarrollo de aplicaciones móviles multiplataforma bajo los enfoques híbrido e interpretado [22].

LiteDB es una librería distribuida como un único archivo DLL bajo licencia MIT. Utiliza documentos BSON para almacenar la información. Se encuentra disponible para el desarrollo de aplicaciones móviles multiplataforma en Xamarin (compilación cruzada) [23].

DBMSs de otros tipos

Oracle Berkeley DB es una familia de bases de datos de clave-valor embebidas. Tiene licencia open source y se encuentra disponible para el desarrollo de aplicaciones nativas Android y iOS [3,24].

ObjectBox es un DBMS orientado a objetos para dispositivos móviles e IoT. Tiene licencia Apache 2.0, y se encuentra disponible para el desarrollo de aplicaciones móviles nativas en Android y iOS, y multiplataforma en Flutter. Cuenta con un servicio de sincronización y almacenamiento en la nube [3,25].

Sparksee es un DBMS de grafos. Se distribuye bajo licencia comercial y cuenta con licencias gratuitas para educación e investigación. Se puede utilizar en el desarrollo de aplicaciones nativas en Android y iOS [3,26].

4. Experimentación

Para la realización de la experimentación se seleccionan DBMSs bajo las siguientes condiciones, definidas para este trabajo: poder ser utilizado sin una licencia comercial; poder ser utilizado de forma autónoma, completamente offline; contar con herramientas para realizar sincronización con bases de datos remotas; poder ser utilizado en aplicaciones móviles desarrolladas de forma nativa en las principales plataformas (Android y iOS); poder ser utilizado en aplicaciones móviles desarrolladas con diferentes enfoques multiplataforma en frameworks de desarrollo conocidos (React Native, NativeScript, Ionic Framework, Xamarin, Flutter); que esté en las primeras tres posiciones de acuerdo al ranking por modelo de DBMS elaborado por DB-Engines.

A partir de las características que presenta cada DBMS, y de los criterios enunciados previamente, se seleccionan para la realización del análisis experimental comparativo SQLite, Couchbase Lite y Realm.

Para llevar a cabo el análisis se desarrollaron parcialmente tres aplicaciones móviles en la plataforma Android, cada una utilizando los DBMSs seleccionados. La aplicación consiste en una agenda de contactos que cumple con los siguientes requerimientos:

Requerimientos funcionales:

R1 La aplicación debe permitir crear un nuevo contacto, con los siguientes datos: Apellido (requerido); Nombre (requerido); Fecha de nacimiento (opcional);

ISBN 978 -987-633-574-4

Emails (cero o más); Teléfonos (cero o más; para cada Teléfono se almacena: Número —requerido— y Tipo —requerido, se debe seleccionar entre las siguientes opciones: Móvil, Casa, Trabajo, Otro—)

R2 La aplicación debe listar los contactos almacenados ordenados por apellido y nombre de forma ascendente.

R3 La aplicación debe permitir filtrar los contactos almacenados, a partir de un único término de búsqueda, listando los contactos que coincidan parcial o totalmente con el término de búsqueda en algunos de sus campos. Los contactos filtrados se mostrarán ordenados por Apellido y Nombre ascendentemente.

Requerimientos no funcionales:

R4 Toda la información debe almacenarse de forma local en la base de datos.

Para el desarrollo de la aplicación se define un único diagrama correspondiente al modelo conceptual de base de datos, utilizando el Modelo Entidad-Relación, que luego se deriva a los modelos lógico y/o físico correspondientes según cada modelo de base de datos y DBMS.

El análisis se realizará desde el punto de vista de la experiencia del ingeniero de software en el desarrollo de cada aplicación, considerando la complejidad de implementación para la utilización del DBMS correspondiente, en cuanto a factores como instalación/configuración, implementación de componentes requeridos —por ejemplo, estructuras de datos, clases—, definición/ejecución de consultas, entre otros.

Teniendo en cuenta el objetivo del análisis no se considera relevante que el esquema de datos definido se adapte mejor a un modelo de base de datos específico, ya que el interés está focalizado en analizar la implementación de las diferentes características específicas de cada modelo de base de datos.

Los requerimientos y esquemas de datos definidos tienen por objetivo la implementación de las características específicas de los modelos de bases de datos seleccionados: *relaciones* en el modelo relacional; *documentos embebidos*, *arrays de tipos escalares*, y *arrays de documentos embebidos* en el modelo NoSQL documental.

4.1. SQLite

Para la experimentación utilizando SQLite como DBMS, se deriva el diagrama correspondiente al modelo conceptual al modelo físico (Figura 1)

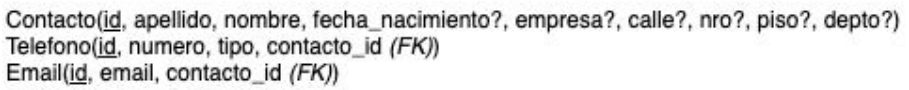


Figura 1: Diagrama correspondiente al modelo físico para SQLite

Para la utilización de SQLite en el desarrollo de aplicaciones móviles, Android provee dos formas de gestionar bases de datos en dicho DBMS. La primera

ISBN 978 -987-633-574-4

de ellas (recomendada por la documentación oficial) es utilizando Room, una biblioteca de persistencia que funciona a modo de capa de abstracción de SQLite; la segunda es utilizando la API de SQLite directamente [27]. La forma elegida para la experimentación es la primera.

Instalación/Configuración La instalación de Room se realiza agregando las dependencias necesarias en el archivo *gradle*.

Definición del esquema de datos El uso de Room se lleva a cabo definiendo clases e interfaces de objetos a los que se les deben definir determinadas anotaciones.

La biblioteca cuenta con tres componentes principales: *entidades*, que son clases que representan a las entidades del modelo, y que representan las tablas del modelo relacional. Estas clases deben anotarse con `@Entity`.

El segundo componente principal de Room son los *DAO*s, definidos a partir de interfaces con la anotación `@Dao`, que son utilizados para realizar consultas sobre la tabla que representa cada entidad, permitiendo obtener entidades, modificarlas, crearlas y eliminarlas.

El tercer componente es la *base de datos*, que sirve como punto de acceso principal para la conexión subyacente a la base de datos relacional. Se debe definir una clase abstracta que extienda de la clase `RoomDatabase` y que esté anotada con `@Database`.

En las relaciones entre las diferentes entidades deben anotarse las restricciones de clave foránea, indicando para cada una la entidad y propiedad/es que referencia, y la propiedad sobre la que aplica.

Debido a la imposibilidad de referenciar objetos en Room, las propiedades cuyo tipo no correspondan a un tipo de dato escalar (tipos númericos, *strings*, y booleano) deben ser convertidas para poder ser almacenadas en la base de datos. Para ello deben definirse dos métodos de conversión por cada tipo de dato que desee persistirse.

Inserción de datos Para satisfacer el requerimiento funcional *R1* deben insertarse las tuplas correspondientes al contacto a crear, y a los teléfonos y emails del contacto. Para ello se deben definir métodos en las interfaces correspondientes a los *DAOs* anotados con `@Insert`.

Recuperación de datos La aplicación desarrollada en el marco de la experimentación debe listar todos los contactos almacenados (*R2*), y los contactos que coincidan parcial o totalmente con un término de búsqueda (*R3*). Para consultas de recuperación de tuplas deben definirse métodos en la interfaz correspondiente al *DAO* anotadas con `@Query`, cuyo valor de la anotación es la consulta SQL. Para filtrar a partir de un término de búsqueda es necesario parametrizar éste, lo que se logra definiendo un parámetro en el método, cuyo valor es posible referenciar en la consulta prefijando con dos puntos el nombre del parámetro.

En los casos que sea necesario recuperar información de múltiples tablas interrelacionadas, se debe implementar una clase con propiedades cuyo tipo corresponda a las entidades involucradas.

Código fuente El código fuente de la experimentación se encuentra en `https://github.com/ftesone/tesina-room/tree/clei-2021`.

ISBN 978 -987-633-574-4

4.2. Couchbase Lite

Una característica particular de las bases de datos documentales es que ofrecen la posibilidad de almacenar datos sin estructura o semiestructurados. Asimismo, es posible definir parcial o totalmente un esquema para los documentos.

Para la implementación no se define un esquema para la base de datos, pero sí se estructuran los documentos según el modelo físico de la Figura 2.

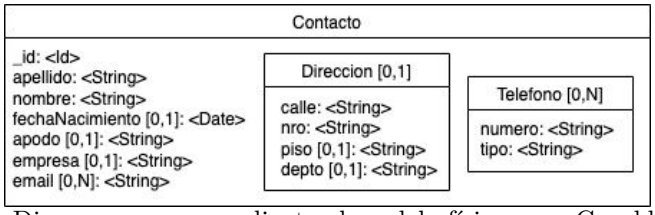


Figura 2: Diagrama correspondiente al modelo físico para Couchbase Lite

Instalación/Configuración La instalación de Couchbase Lite se realiza agregando las dependencias necesarias en el archivo *gradle*.

Definición del esquema de datos La administración de datos se realiza utilizando clases ya definidas por Couchbase Lite, entre las que se encuentran `MutableDocument`, `MutableArray`, `MutableDictionary`, entre otras. Objetos de estas clases se utilizan para definir la estructura de cada documento a almacenar en la base de datos; cada documento puede tener una estructura única.

Si bien es posible definir una estructura de clases para realizar una implementación orientada a objetos, debe también implementarse la hidratación de los objetos.

Inserción de datos Para realizar altas de documentos se deben crear instancias de la clase `MutableDocument` y definir los campos del documento (los campos opcionales que no tienen un valor no se definen).

Para embeber documentos se utiliza la clase `MutableDictionary`, ya sean éstos correspondientes a un único documento o a un array de documentos.

En el caso de los arrays, ya sea de documentos embebidos o de tipos de datos escalares, para la definición de éstos se debe utilizar la clase `MutableArray`. Por lo tanto, un array de documentos embebidos se define como un `MutableArray` cuyos valores son instancias de `MutableDictionary`.

Recuperación de datos Para realizar consultas de recuperación de información se debe utilizar un objeto de la clase `QueryBuilder`, al que deben enviarse mensajes para definir expresiones que permitan establecer condiciones sobre propiedades de los documentos, criterios de orden, entre otros. Los tipos de datos devueltos por la ejecución de consultas son variantes inmutables de los utilizados para inserciones (`Dictionary`, `Array`; los documentos se encapsulan en objetos de clase `Result`).

Código fuente El código fuente de la experimentación se encuentra en `https://github.com/ftesone/tesina-couchbase/tree/clei-2021`.

ISBN 978 -987-633-574-4

4.3. Realm

Para la implementación de Realm se utiliza el mismo modelo físico definido para Couchbase Lite, presentado en la Figura 2, ya que, si bien hay diferencias entre los dos DBMS, las estructuras definidas en el modelo también aplican para Realm.

Instalación/Configuración La instalación de Realm se realiza agregando las dependencias necesarias en el archivo *gradle*.

Definición del esquema de datos La utilización de Realm se lleva a cabo definiendo clases que representen los documentos del modelo, que deben definirse como subclases de `RealmObject` o que implementen la interfaz `RealmModel`, ya sea que se traten de documentos embebidos o no.

Para los documentos *top-level* se debe definir una propiedad con tipo `ObjectId` y anotarse con `@PrimaryKey`. Las clases correspodientes a documentos que se utilicen embebidos en otros documentos, ya sea directamente o como un array de documentos, deben anotarse con `@RealmClass(embedded = true)`.

En los casos que se definan documentos con arrays de valores de tipos de datos escalares o arrays de documentos, los tipos de las propiedades deben definirse como `RealmList<T>`, donde `T` es el tipo de dato escalar o la clase correspondiente al documento embebido.

Inserción de datos Las operaciones de escritura en la base de datos en Realm se realiza a través de transacciones, las cuales se definen invocando al método `executeTransaction` de la instancia de la base de datos, en donde se pasa como parámetro una expresión lambda en la que se define el comportamiento. En el cuerpo de la expresión debe obtenerse una instancia de la clase correspondiente al documento que se insertará y deben definirse los valores de las propiedades a persistir.

Recuperación de datos Para realizar consultas para recuperar documentos almacenados se debe obtener una instancia de `RealmQuery<T>` —donde `T` corresponde a la clase de documento que se desea recuperar—, invocando al método `where` de la instancia de la base de datos, pasando como parámetro el nombre de la clase `T`.

La clase `RealmQuery` permite invocar distintos métodos para filtrar y ordenar, entre otras, y obtener los documentos encapsulados en una instancia de `RealmResults<T>`.

Código fuente El código fuente de la experimentación se encuentra en `https://github.com/ftesone/android-realm/tree/clei-2021`.

5. Análisis de Resultados

5.1. SQLite

En la experimentación utilizando SQLite como DBMS para persistir la información se decidió utilizar la biblioteca Room, que representa una capa de abstracción sobre SQLite. En la implementación realizada se destacan como ventajas: (1) la estructura de clases que se requiere definir lleva a trabajar de una forma ordenada y sistemática; (2) la definición del esquema de la BD a partir de las clases que conforman las entidades del modelo, por lo que no es necesario

ISBN 978 -987-633-574-4

definirlo con sentencias SQL; (3) la flexibilidad que se provee para definir consultas para obtener información, ya que sólo se debe definir la consulta como valor de una anotación, incluídas las consultas parametrizadas; (4) la facilidad con la que se definen conversiones de tipos que no se pueden almacenar directamente en la base de datos; (5) la facilidad con la que se insertan tuplas.

Asimismo, se enumeran algunas desventajas en la utilización de Room: (1) la necesidad de utilizar estructuras auxiliares para obtener tuplas resultado de productos (*joins*) entre tablas; (2) la necesidad de definir un conversor para un tipo de dato ampliamente utilizado como es el tipo de dato *Date*.

5.2. Couchbase Lite

En la implementación realizada con Couchbase Lite, considerando la experimentación realizada, pueden observarse como ventajas: (1) la flexibilidad en la definición de la estructura de documentos, ya que ésta puede ser definida de forma completamente dinámica; (2) la utilización de estructuras definidas por Couchbase Lite para la persistencia y recuperación de información.

Pueden considerarse como desventajas: (1) no es posible agrupar los documentos en colecciones; (2) la implementación de consultas complejas requiere definir gran cantidad de código, lo que dificulta su comprensión; (3) la implementación de consultas que referencien valores en arrays es compleja debido a la necesidad de definir variables que hagan referencia a cada valor del array; (4) la necesidad de implementar la hidratación de objetos en caso de desarrollar la aplicación utilizando el paradigma de programación orientada a objetos;

5.3. Realm

La implementación realizada con Realm permite enumerar las siguientes ventajas: (1) la definición de la estructura de los documentos a través de la definición de clases; (2) la posibilidad de agrupar documentos en colecciones; (3) la implementación de consultas resulta concisa y clara, especialmente porque no es necesario definir variables para referenciar a campos en documentos embebidos o en arrays de documentos embebidos.

Asimismo, se encontró como desventaja la imposibilidad de referenciar a valores correspondientes a arrays de tipos escalares en las consultas.

El Cuadro 1 resume los aspectos evaluados en SQLite, Couchbase Lite y Realm. Cada aspecto fue calificado en una escala de cinco valores: *muy bajo*, *bajo*, *medio*, *alto*, *muy alto*. Los aspectos evaluados son:

- Complejidad: expresa el nivel de dificultad para realizar la tarea;
- Flexibilidad: aplica sólo a la definición del esquema de datos. Se refiere a cuán flexible es el esquema para adaptarse a cambios;
- Legibilidad del código: expresa el nivel de dificultad para comprender el código fuente;
- Integración con POO (Programación Orientada a Objetos): nivel de interacción de los tipos de datos del DBMS con los objetos definidos en la aplicación;
- Soporte de tipos de datos no escalares: nivel de dificultad en la persistencia de tipos de datos no escalares, es decir, tipos de datos distintos a tipos numéricos, *strings*, y booleanos.

ISBN 978 -987-633-574-4

La mejor calificación posible es *muy alto* para todos los aspectos analizados, a excepción de Complejidad, cuya mejor calificación es *muy baja*.

Tarea	**Aspectos**	**SQLite**	**Couchbase Lite**	**Realm**
Instalación/Configuración	Complejidad	Muy baja	Muy baja	Muy baja
Definición del esquema de datos	Complejidad	Baja	Muy baja	Muy baja
	Flexibilidad	Baja	Muy alta	Media
	Legibilidad del código	Alta	Muy baja	Muy alta
	Integración con POO	Alta	Muy baja	Alta
	Soporte de tipos de datos no escalares	Muy baja	Baja	Baja
Inserción de datos	Complejidad	Baja	Media	Baja
	Legibilidad de código	Muy alta	Alta	Alta
	Integración con POO	Muy alta	Baja	Alta
Recuperación de datos	Complejidad	Baja	Muy alta	Muy baja
	Legibilidad de código	Muy alta	Muy baja	Muy alta
	Integración con POO	Alta	Muy baja	Muy alta

Cuadro 1: Aspectos evaluados en los DBMSs

6. Conclusiones y Trabajo Futuro

Este trabajo intenta abordar la problemática que representa la elección de un DBMS adecuado que pueda ser embebido en una aplicación móvil, acorde al problema a resolver.

El aumento exponencial en el volúmen de datos administrados, incluyendo datos estructurados, semi-estructurados, y no estructurados, provocó el surgimiento de nuevos modelos de bases de datos, denominados bases de datos NoSQL, para facilitar el almacenamiento masivo de datos semi-estructurados y no estructurados.

Por otra parte, las mejoras en las prestaciones de hardware de los dispositivos móviles conducen a que éstos administren cada vez más información, y que surjan nuevos sistemas de gestión de bases de datos que se instalan en dichos dispositivos.

A raíz de esto, este trabajo analiza distintos aspectos referidos a la instalación y/o configuración, implementación de componentes requeridos, definición y/o ejecución de consultas de modificación y recuperación de datos, para asistir al ingeniero de software en la selección de un DBMS adecuado para ser embebido en aplicaciones móviles acorde al problema a resolver.

A partir de la experimentación y análisis realizados puede considerarse que el problema a resolver resulta determinante en la elección de un DBMS. SQLite y Realm serían los más adecuados dado que presentan una baja complejidad, alta legibilidad de código y alta integración con estructuras del paradigma de programación orientada a objetos, en las tareas correspondientes a definición del esquema de datos, inserción de datos, y recuperación de datos. Sin embargo, presentan una baja flexibilidad en la definición del esquema de datos, aspecto en el que Couchbase Lite sería más adecuado.

ISBN 978 -987-633-574-4

En situaciones en las que no se requiera una gran flexibilidad en el esquema de datos, SQLite y Realm serían los más adecuados. Por el contrario en situaciones en las cuales se requiera una gran flexibilidad en el esquema de datos, Couchbase Lite representaría una mejor opción. En la experimentación planteada, que no requiere un esquema de datos flexible, SQLite y Realm resultan más adecuados, considerándose el primero como la mejor opción, debido a que el problema planteado se ajusta mejor al modelo relacional.

Finalmente, se proponen varias líneas de investigación como posible trabajo futuro:

1. extender la experimentación realizada agregando el requerimiento no funcional de sincronización de la base de datos de la aplicación móvil con una base de datos *backend*;
2. analizar el impacto que produce la utilización de SQLite, Couchbase Lite o Realm en requerimientos no funcionales que resultan determinantes en el éxito de la aplicación, como el uso de memoria principal y de memoria secundaria, el rendimiento en la ejecución de consultas, y el consumo de energía;
3. extender el análisis realizado utilizando otros DBMSs, resultando particularmente interesante aquellos DBMSs que se encuentren disponibles para su uso tanto en Android como en iOS, y en frameworks conocidos para el desarrollo de aplicaciones móviles multiplataforma.

Referencias

1. K. L. Berg, T. Seymour, and R. Goel, "History of databases," *International Journal of Management & Information Systems (IJMIS)*, vol. 17, no. 1, pp. 29–36, 2013.
2. B. Grad and T. J. Bergin, "Guest editors' introduction: History of database management systems," *IEEE Annals of the History of Computing*, vol. 31, no. 4, pp. 3–5, 2009.
3. "Db-engines ranking - populariry ranking of database management systems." https://db-engines.com/en/ranking. Accedido por última vez: 17/02/2021.
4. M. Campbell-Kelly and D. D. Garcia-Swartz, *From mainframes to smartphones: a history of the international computer industry*, vol. 1. Harvard University Press, 2015.
5. "List of countries by smartphone penetration - wikipedia." https://en.wikipedia.org/wiki/List_of_countries_by_smartphone_penetration#2013_rankings. Accedido por última vez: 17/02/2021.
6. "• cell phone sales worldwide 2007-2020 | statista." https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/. Accedido por última vez: 17/02/2021.
7. L. Delia, N. Galdamez, P. Thomas, L. Corbalan, and P. Pesado, "Multi-platform mobile application development analysis," in *2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*, pp. 181–186, IEEE, 2015.
8. S. Xanthopoulos and S. Xinogalos, "A comparative analysis of cross-platform development approaches for mobile applications," in *Proceedings of the 6th Balkan Conference in Informatics*, pp. 213–220, 2013.

ISBN 978 -987-633-574-4

9. L. Marrero, V. Olsowy, P. J. Thomas, L. N. Delía, F. Tesone, J. Fernández Sosa, and P. M. Pesado, "Un estudio comparativo de bases de datos relacionales y bases de datos nosql," in *XXV Congreso Argentino de Ciencias de la Computación (CACIC)(Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019)*, 2019.
10. A. Nori, "Mobile and embedded databases," in *Proceedings of the 2007 ACM SIGMOD International Conference on Management of data*, pp. 1175–1177, 2007.
11. Q. H. Mahmoud, S. Zanin, and T. Ngo, "Integrating mobile storage into database systems courses," in *Proceedings of the 13th annual conference on Information technology education*, pp. 165–170, 2012.
12. S. Lee, "Creating and using databases for android applications," *International Journal of Database Theory and Application*, vol. 5, no. 2, 2012.
13. A. Alzahrani, N. Alalwan, and M. Sarrab, "Mobile cloud computing: advantage, disadvantage and open challenge," in *Proceedings of the 7th Euro American Conference on Telematics and Information Systems*, pp. 1–4, 2014.
14. "About sqlite." https://www.sqlite.org/about.html. Accedido por última vez: 21/02/2021.
15. "Interbase - embarcadero website." https://www.embarcadero.com/es/products/interbase. Accedido por última vez: 03/03/2021.
16. "Sap sql anywhere | rdbms for iot & data-intensive apps | technical information." https://www.sap.com/products/sql-anywhere/technical-information.html. Accedido por última vez: 03/03/2021.
17. "Opentext gupta sqlbase." https://www.opentext.com/products-and-solutions/products/specialty-technologies/opentext-gupta-development-tools-databases/opentext-gupta-sqlbase. Accedido por última vez: 03/03/2021.
18. "Lite | couchbase." https://www.couchbase.com/products/lite. Accedido por última vez: 21/02/2021.
19. "Firebase realtime database | firebase realtime database." https://firebase.google.com/docs/database. Accedido por última vez: 21/02/2021.
20. "Home | realm.io." https://realm.io/. Accedido por última vez: 22/02/2021.
21. "Cloud firestore | firebase." https://firebase.google.com/docs/firestore/. Accedido por última vez: 22/02/2021.
22. "Pouchdb, the javascript database that syncs!." https://pouchdb.com/. Accedido por última vez: 21/02/2021.
23. "Litedb :: A .net embedded nosql database." http://www.litedb.org/. Accedido por última vez: 03/03/2021.
24. "Oracle berkeley db." https://www.oracle.com/database/technologies/related/berkeleydb.html. Accedido por última vez: 03/03/2021.
25. "Mobile database | android database | ios database | flutter database." https://objectbox.io/mobile-database/. Accedido por última vez: 03/03/2021.
26. "Sparsity-technologies: Sparksee high-performance graph database." http://sparsity-technologies.com/#sparksee. Accedido por última vez: 03/03/2021.
27. "Descripción general del almacenamiento de archivos y datos." https://developer.android.com/training/data-storage. Accedido por última vez: 21/02/2021.

CACIC 2021

WORKSHOP INGENIERIA DE SOFTWARE

COORDINADORES

Patricia Pesado (UNLP)
Elsa Estevez (UNS)
Alejandra Cechich (UNCOMA)
Horacio Kuna (UNaM)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978 -987-633-574-4

An expressive and enriched specification language to synthesize behavior in BIG DATA systems

Fernando Asteasuain[1,2] and Luciana Rodriguez Caldeira[2]

[1] Universidad Nacional de Avellaneda, Argentina
fasteasuain@undav.edu.ar
[2] Universidad Abierta Interamericana - Centro de Altos Estudios CAETI, Argentina
luciana.rodriguezcaldeira@alumnos.uai.edu.ar

Abstract. In this work we extend our behavioral specification and controller synthesis framework FVS to deal with BIG DATA requirements. For one side, we enriched FVS expressive power by exhibiting how our language can handle fluents and partial specifications. For the other side, we combined FVS with a parallel model checker in order to automatically obtain a controller given the behavior specification. In this way, FVS can be presented as an attractive tool to formally verify and synthesize behavior for BIG DATA systems. Our approach is compared to other well known parallel tool analyzing a complex big data system.

Keywords: Formal Verification, BIG DATA, Parallel Model Checkers

1 Introduction

Nowadays BIG DATA systems are surprisingly present in every day life. The Software Engineering community has been trying to adapt, to extend or to create tools, techniques and methods to deal with the new conditions and requirements that these systems expose [9, 24, 15, 30, 21, 31, 25]. For example, data warehouse approaches have emerged since traditional ways to structure data such as relational data bases are no that useful in this domain.

According to some approaches, one of the software engineering areas that more urgently need attention and contributions is formal verification [24, 17]. The challenges to be addressed are inherent to BIG DATA systems: handling and reasoning about tons of unstructured, informal and heterogeneous data and information. To deal with this kind of context flexible and expressive formalism are needed to express, validate and reason about the expected behavior of the systems [25, 24, 17]. In addition, more efficient mechanisms are needed since performance is a crucial feature to achieve when dealing with BIG DATA systems. In this sense, most of the formal verification approaches to big data try to obtain better execution times by providing parallel version of the algorithms involved. However, the flexibility and expressive power of the formalisms has been somehow neglected.

ISBN 978 -987-633-574-4

One of the most applied techniques in traditional formal verification is controller synthesis [18, 10]. In these approaches, a controller is automatically build upon the expected behavior of the system and the environment it interacts. Usually, the controller takes the form of an automaton which decides which actions to take based on the received information (mostly provided by external sensors). The controller is built using game theory concepts, obtaining a winning strategy that takes the system to an accepting state no matter which actions the environment chooses [10]. Regarding expressive power, fluents [20] and partial specifications are powerful formalisms to deal with unstructured behavior. A fluent allows to model behavior that take place between intervals or moments, introducing a new layer of abstraction in the specification. Starting and ending actions of these intervals are defined, and then properties can be stated using the mentioned intervals. Partial specifications introduce the possibility of specifying optional behavior, a feature that is highly appreciated when dealing with requirements in early stages since conditions and behavior itself are not clearly determined. MTS (Modal Transition Systems) [26] is one of the most widely known formalisms. Since controller synthesis deal with exponential algorithms, some parallel and distributed extensions have emerged. One of them is the parallel version of the MTSA (Modal Transition System Analyzer) model checker [12], which is available at: https://mtsa.dc.uba.ar/. MTSA allows to efficiently obtain controllers in different domains [28, 29]. However, it has not been explored in the BIG DATA domain.

In previous work we took an initial step to adapt our behavioral specification language FVS (FeatherWeight Visual Scenarios) to deal with BIG DATA requirements [2]. Specifically, we have parallelized the way our specification is build, introducing a parallel algorithm to translate FVS graphical scenarios into Büchi automata. In this work we continued this path by combining FVS specifications with the MTSA parallel model checker. *In this way, we end up with a flexible and extremely rich expressive power formalism to specify behavior and perform controller synthesis in BIG DATA systems in a efficient way.* In order to interact with the MTSA tool we enriched FVS in two orthogonal aspects, illustrating how FVS can express fluents and partial specifications, and providing a combination to a parallel model checker. As a case of study we analyzed a big data system provided in the literate [9] and validate our approach by comparing execution times with another parallel technique [9].

The rest of this work is structured as follows. Section 2 briefly presents FVS and explains how a controller can be obtained. Sections 3 and 4 show how FVS can denote fluents and partial specifications. Section 5 exhibits the case of study and the interaction with the MTSA model checker while Section 6 discusses the obtained results. Finally, Sections 7 and 8 analyze some related and future work and the conclusions of this research.

ISBN 978 -987-633-574-4

2 Feather weight Visual Scenarios

In this section we will informally describe the standing features of FVS. The reader is referred to [1] for a formal characterization of the language. FVS is a graphical language based on scenarios. Scenarios are partial order of events, consisting of points, which are labeled with a logic formula expressing the possible events occurring at that point, and arrows connecting them. An arrow between two points indicates precedence. For instance, in figure 1-(a) A-event precedes B-event. In figure 1-b the scenario captures the very next B-event following an A-event, and not any other B-event. Events labeling an arrow are interpreted as forbidden events between both points. In figure 1-c A-event precedes B-event such that C-event does not occur between them. Finally, FVS features aliasing between points. Scenario in 1-d indicates that a point labeled with A is also labeled with $A \wedge B$. It is worth noticing that A-event is repeated on the labeling of the second point just because of FVS formal syntaxis.

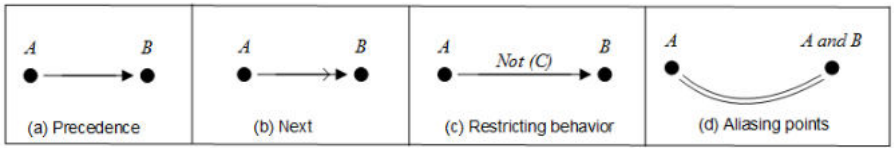


Fig. 1. Basic Elements in FVS

We now introduce the concept of FVS rules, a core concept in the language. Roughly speaking, a rule is divided into two parts: a scenario playing the role of an antecedent and at least one scenario playing the role of a consequent. The intuition is that whenever a trace "matches" a given antecedent scenario, then it must also match at least one of the consequents. In other words, rules take the form of an implication: an antecedent scenario and one or more consequent scenarios. Graphically, the antecedent is shown in black, and consequents in grey. Since a rule can feature more than one consequent, elements which do not belong to the antecedent scenario are numbered to identify the consequent they belong to. An example is shown in figure 2. The rule describes requirements for a valid writing pipe operation. For each write event, then it must be the case that either the pipe did not reach its maximum capacity since it was ready to perform (Consequent 1) or the pipe did reach its capacity, but another component performed a read over the pipe (making the pipe available again) afterwards and the pipe capacity did not reach again its maximum (Consequent 2).

2.1 FVS and Ghosts Events

FVS can denote high level behavior. This is due to the introduction of abstraction, which is incorporated in our notation by introducing a new type of events.

ISBN 978 -987-633-574-4

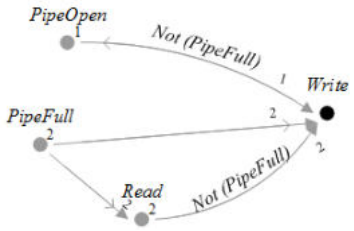


Fig. 2. An FVS rule example

By using these events the user can abstract behavior and reason about events that are not present in the system traces, but actually represent a higher level of abstraction. We call these special events as "ghost" events, in contrast with "actual" events, the set of events present in the system's specification. In order to verify that a rule containing ghosts events satisfies a certain trace of the system (which only contains actual events) there is a internal process based on morphisms that discards ghost events based on a classic process of existential elimination [1].

2.2 Behavioral Synthesis in FVS

FVS specifications can be used to automatically obtain a controller employing a classical behavioral synthesis procedure. We now briefly explains how this is achieved while the complete description is available in [3]. Using the tableau algorithm detailed in [1] FVS scenarios are translated into Büchi automata. Then, if the obtained automata is deterministic, then we obtain a controller using a technique [27] based on the specification patterns [19] and the GR(1) subset of LTL. If the automaton is non deterministic, we can obtain a controller anyway. Employing an advanced tool for manipulating diverse kinds of automata named GOAL [32] we translate these automata into Deterministic Rabin automata. Since synthesis algorithms are also incorporated into the GOAL tool using Rabin automata as input, a controller can be obtained. In this work, we add a new way to obtain a controller, combining FVS with the MTSA model checker, as shown in the remaining sections.

3 Fluents and FVS

Fluents [20] constitute a powerful variant of LTL. A fluent allows to model behavior that take place between intervals or moments, introducing a new layer of abstraction in the specification. Starting and ending actions of these intervals are defined, and then properties can be stated using the mentioned intervals. In [20] a simple example is given to illustrate how fluents works. Suppose we are validating a new decentralized system for organizing television control software. The property to verify is the following: *If the TV tuner is tuning, then the screen*

ISBN 978 -987-633-574-4

must be blanked, and the available events are: *blank* (blanks the screen), *unblank* (displays the new channel signal), *tune* (the tuner starts tuning into the new channel) and *endtune* (the tuner finishes). The tunning interval is defined as beginning with the *tune* event and ending with the *endtune* event. Similarly, the blanking interval starts with the *blank* event and finishes with the *unblank* event. Once these intervals are defined then the property to be checked can be simply formulated as $\Box$ (Tunning $\Rightarrow$ *Blanked*).

FVS can specify fluents in a very simple and direct way employing ghosts events. Fluents starting and ending delimiters are modeled with FVS rules, fluents predicates are modeled with ghost events, and intervals behavior are simply FVS rules using those ghosts events. Rules in Figure 3 specify the TV control system property mentioned before, using two ghosts events, namely *Tunning* and *Blanked*. The rule in the top of the picture defines the *Tunning* event, where the rule in the middle does the same for the *Blanked* event. Finally, the rule in the bottom models the desired property.

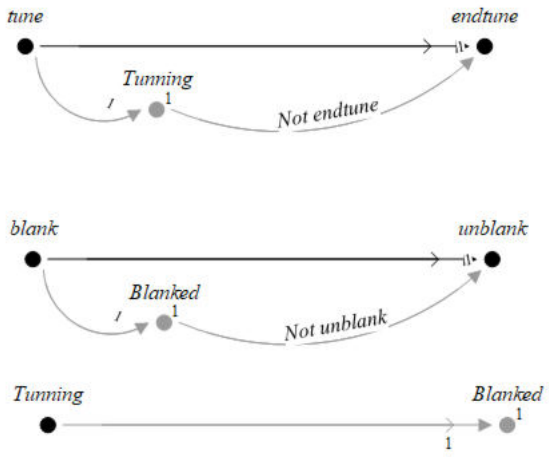


Fig. 3. Fluents in FVS

4 Partial Specifications and FVS

Partial Specifications are a crucial tool to model and shape early behavior of computer systems. They aim to capture early interactions between the elements involved in a stage where the requirements are not yet thoroughly defined. Transitions are divided into *required* and *maybe* categories, where the latter introduce partial or optional behavior. In successive versions of the system *maybe* transitions are either discarded or turned into required behavior. This process

ISBN 978 -987-633-574-4

in known as *refinement*. Perhaps the most known formalism addressing Partial Specifications is Modal Transitions Systems (MTS) [26].

In FVS partial specifications are inherently included since it features optional behavior by employing multiples consequents in its rules. The semantic of the system is given by those traces satisfying all the rules, and a rule with two or more consequents is satisfied if at leat one of them is found, or all of them, if that is the case. So, FVS provides the refinement operation by its traces semantics definition. A rule with multiple consequents can be replaced in next versions with a rule with less consequents (the optional behavior is discarded), or combining two consequents into one (making mandatory an optional behavior).

As an example, suppose a new requirement arises in the TV control system previously described. A new publicity system might be added to the main system. In few words, when the screen is blanked two events could happen: either a publicity is shown or the blanking process ends normally. At this stage, the publicity feature is handled as an optional behavior. In FVS, this is achieved by adding a new consequent, as shown in figure Figure 4.

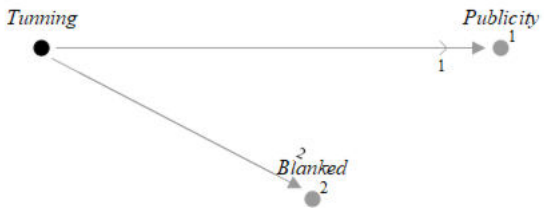


Fig. 4. Describing partial behavior in FVS

5 Case Study: Dekker Algorithm

The case of study we analyzed is introduced on [9] based on the benchmarks in [23]. This model represents a variant of Dekker's mutual exclusion algorithm. As described in [23], the main functioning of this algorithm is the following. Each process has three states, *p0*, *p1*, and *p3*. *p0* is initial. From there, the process executes try and raises its flag, reaching *p1*. In *p1*, if at least one of the other process has a high flag, it withdraws its intent and goes back to *p0*. In *p1*, it enters the critical section if all other process flag is zero. From *p3,* the process can only exit the critical section. The rules in Figure 5 show some of the FVS specification fulfilling the algorithm's requirements. The rules considered actions for one process. The complete specification is obtained by composing all the rules for every process involved. We employed several ghosts events like *Flag*, *EnterCritical* and *ExitCritical* and rules with several consequents to handle partial specifications.

ISBN 978 -987-633-574-4

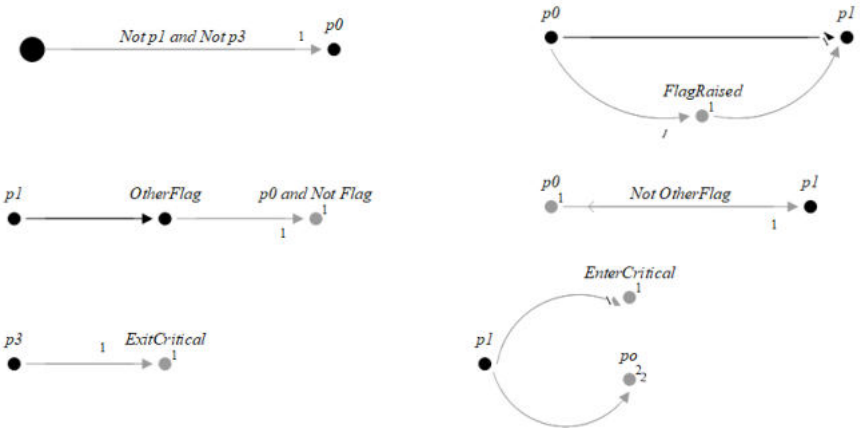


Fig. 5. FVS rules describing the behavior of the Dekker's algorithm

We modeled the complete behavior of the algorithm, and then we obtained a controller using FVS rules as input in the MTSA model checker. Part of the controller is shown in Figure 6.

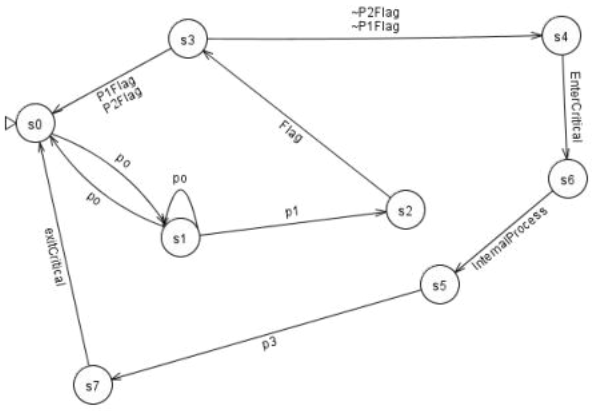


Fig. 6. Part of the behavior of the Dekker's Algorithm Controller

ISBN 978 -987-633-574-4

6 Experimental Results

In this section we describe the results we obtained aiming to measure FVS distributed model checking and controller synthesis performance regarding execution time. The system under analyses consisted of the Dekker algorithm detailed in Section 5. We compared our execution times against the technique in [9], which proposed this case of study. Although they verified the behavior of the algorithm and we obtained a controller instead, the involved tasks and objectives are similar enough to produce valuable results from their performance execution time comparison. We ran our experiments in a Bangho Inspiron5458, with a Dual Core i5-5200U and 8GB RAM memory.

As in [9], we conducted the experiment running the algorithm 10, 15 and 20 processes. Table 1 subsumes the obtained results, where the column Map-Reduce CTL stands for the technique described in [9] and times is denoted in seconds. It can be noted that although our execution times are worst the difference is not a critical value, and it reduces as the complexity of the problem increase. Thus, it can be preliminary observed that FVS provides great flexibility and expressive power to synthesize behavior in complex cases without neglecting performance.

Table 1. Dekker Algorithm Execution Time

Example	Map-Reduce CTL	Parallel FVS
10-Dekker	50 sec	87 sec
15-Dekker	825 sec	965 sec
20-Dekker	11134 sec	11529 sec

7 Related and Future Work

There are several approaches implementing different versions of parallel model checking algorithms for both linear and branching [22, 11, 13, 6, 7, 14] . It would be interesting to compare the FVS-MTSA duo explored in this work with some of the mentioned tools. Similarly, other approaches aim to speed-up the model checking by performing parallel verification of very small units pieces of behavior [16, 5] . For example, these units are called *swarms* in [16].

In [9, 15]] a interesting framework for distributed CTL (computation tree logic) model checker is presented. They introduce a novel architecture employing HADOOP MAPREDUCE as its computational engine. They provide a very solid empiric evaluation with several case of studies employing Amazon Elastic MapReduce [15] and the GRID5000 cloud infrastructure [4]. For generating and building distributed state space exploration they rely on a framework called Mardigras [8]. This framework introduces a general scheme to verify systems, allowing behavior to be specified using logics, Petri Nets and other formalisms. It would be interesting to explore if FVS can be added in this list.

ISBN 978 -987-633-574-4

Regarding future work, we would like to deepen our empirical evaluation by introducing more case of studies and also a space comparison besides the execution time. We believe a comparison taking into account, for example, the number of states and transitions of the automata involved in the verifying process can enrich the results analyzed in this work. Similarly, from the theoretical view we would like to provide formal proofs regarding the equivalence with the fluents and partial specification mechanisms.

8 Conclusions

In this work we present a powerful, flexible and highly expressive specification language to denote behavior and perform controller systems in BIG DATA systems. In particular, we show how FVS is able to express behavior in terms of fluents and partial specifications. In order to deal with BIG DATA performance requirements we combine FVS specifications with the parallel model checker MTSA. In this way, a controller can be found using FVS specifications as input. We compared our executions times with other well known parallel approach analyzing a compelling case of study. By looking at the preliminary results obtained so far we can conclude that FVS exhibits great flexibility and expressive power without a significative loss in performance.

References

1. F. Asteasuain and V. Braberman. Declaratively building behavior by means of scenario clauses. *Requirements Engineering*, 22(2):239–274, 2017.
2. F. Asteasuain and L. R. Caldeira. A parallel tableau algorithm for big data verification. In *CACIC. ISBN 978-987-4417-90-9, pp 360-369*, 2018.
3. F. Asteasuain, F. Calonge, and M. Dubinsky. Exploring specification pattern based behavioral synthesis with scenario clauses. In *CACIC*, 2018.
4. D. Balouek, A. C. Amarie, G. Charrier, F. Desprez, E. Jeannot, E. Jeanvoine, A. Lèbre, D. Margery, N. Niclausse, L. Nussbaum, et al. Adding virtualization capabilities to the grid5000 testbed. In *CLOSER*, pages 3–20. Springer, 2012.
5. J. Barnat, P. Bauch, L. Brim, and M. Češka. Employing multiple cuda devices to accelerate ltl model checking. In *2010 IEEE 16th International Conference on Parallel and Distributed Systems*, pages 259–266. IEEE, 2010.
6. J. Barnat, L. Brim, M. Češka, and P. Ročkai. Divine: Parallel distributed model checker. In *2010 ninth PDMC*, pages 4–7. IEEE, 2010.
7. A. Bell and B. R. Haverkort. Sequential and distributed model checking of petri nets. *STTT journal*, 7(1):43–60, 2005.
8. C. Bellettini, M. Camilli, L. Capra, and M. Monga. Mardigras: Simplified building of reachability graphs on large clusters. In *RP workshop*, pages 83–95, 2013.
9. C. Bellettini, M. Camilli, L. Capra, and M. Monga. Distributed ctl model checking using mapreduce: theory and practice. *CCPE*, 28(11):3025–3041, 2016.
10. R. Bloem, B. Jobstmann, N. Piterman, A. Pnueli, and Y. Sa'Ar. Synthesis of reactive (1) designs. 2011.
11. M. C. Boukala and L. Petrucci. Distributed model-checking and counterexample search for ctl logic. *IJSR 3*, 3(1-2):44–59, 2012.

ISBN 978 -987-633-574-4

12. M. V. Brassesco. Síntesis concurrente de controladores para juegos definidos con objetivos de generalized reactivity(1). *Tesis de Licenciatura.*, http://dc.sigedep.exactas.uba.ar/media/academic/grade/thesis/tesis_18.pdf UBA FCEyN Dpto Computacion 2017.
13. L. Brim, I. Černá, P. Moravec, and J. Šimša. Accepting predecessors are better than back edges in distributed ltl model-checking. In *FMCAD*, pages 352–366, 2004.
14. L. Brim, K. Yorav, and J. Žídková. Assumption-based distribution of ctl model checking. *STTT*, 7(1):61–73, 2005.
15. M. Camilli. Formal verification problems in a big data world: towards a mighty synergy. In *ICSE*, pages 638–641, 2014.
16. R. DeFrancisco, S. Cho, M. Ferdman, and S. A. Smolka. Swarm model checking on the gpu. *STTT*, 22(5):583–599, 2020.
17. J. Ding, D. Zhang, and X.-H. Hu. A framework for ensuring the quality of a big data service. In *2016 SCC*, pages 82–89. IEEE, 2016.
18. N. DIppolito, V. Braberman, N. Piterman, and S. Uchitel. Synthesising non-anomalous event-based controllers for liveness goals. *ACM Tran*, 22(9), 2013.
19. M. Dwyer, M. Avrunin, and M. Corbett. Patterns in property specifications for finite-state verification. In *ICSE*, pages 411–420, 1999.
20. D. Giannakopoulou and J. Magee. Fluent model checking for event-based systems. In *European software engineering conference*, pages 257–266, 2003.
21. O. Hummel, H. Eichelberger, A. Giloj, D. Werle, and K. Schmid. A collection of software engineering challenges for big data system development. In *SEAA*, pages 362–369. IEEE, 2018.
22. O. Inverso and C. Trubiani. Parallel and distributed bounded model checking of multi-threaded programs. In *PPoPP*, pages 202–216, 2020.
23. F. Kordon, A. Linard, M. Becutti, D. Buchs, L. Fronc, L. M. Hillah, F. Hulin-Hubard, F. Legond-Aubry, N. Lohmann, A. Marechal, et al. Web report on the model checking contest@ petri net 2013. 2013.
24. V. D. Kumar and P. Alencar. Software engineering for big data projects: Domains, methodologies and gaps. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 2886–2895. IEEE, 2016.
25. R. Laigner, M. Kalinowski, S. Lifschitz, R. S. Monteiro, and D. de Oliveira. A systematic mapping of software engineering approaches to develop big data systems. In *SEAA*, pages 446–453. IEEE, 2018.
26. K. G. Larsen and B. Thomsen. A modal process logic. *LICS, pages 203210, IEEE.*
27. S. Maoz and J. O. Ringert. Synthesizing a lego forklift controller in gr (1): A case study. *arXiv preprint arXiv:1602.01172*, 2016.
28. L. Nahabedian, V. Braberman, N. D'Ippolito, S. Honiden, J. Kramer, K. Tei, and S. Uchitel. Dynamic update of discrete event controllers. *IEEE Transactions on Software Engineering*, 46(11):1220–1240, 2018.
29. L. Nahabedian, V. Braberman, N. Dippolito, J. Kramer, and S. Uchitel. Dynamic reconfiguration of business processes. In *International Conference on Business Process Management*, pages 35–51. Springer, 2019.
30. C. E. Otero and A. Peter. Research directions for engineering big data analytics software. *IEEE Intelligent Systems*, 30(1):13–19, 2014.
31. P. A. Sri and M. Anusha. Big data-survey. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, 4(1):74–80, 2016.
32. Y.-K. Tsay, Y.-F. Chen, M.-H. Tsai, K.-N. Wu, and W.-C. Chan. Goal: A graphical tool for manipulating büchi automata and temporal formulae. In *TACAS*, pages 466–471. Springer, 2007.

ISBN 978 -987-633-574-4

Identificación de Variedad Contextual en Modelado de Sistemas Big Data

Líam Osycka[1] *, Agustina Buccella[1], and Alejandra Cechich[1]

GIISCO Research Group
Departamento de Ingeniería de Sistemas - Facultad de Informática
Universidad Nacional del Comahue
Neuquen, Argentina
liam.osycka, agustina.buccella, alejandra.cechich@fi.uncoma.edu.ar

Resumen La propiedad de los sistemas Big Data con respecto a diversidad de los datos se denomina Variedad y su análisis permite identificar distintos tipos; por ejemplo, la variedad estructural denota la variedad en formatos y tipos de datos, clasificándolos como estructurados, semi-estructurados y no estructurados. En particular, el agregado de información de contexto (o dominio) permite análisis más complejos en la variedad, llevando a una nueva fase de investigación en su modelado que incluye la posibilidad de reuso. En este artículo, presentamos una propuesta para modelar sistemas Big Data para/con reuso teniendo en cuenta variaciones en el contexto que surgen del análisis de datos existentes para un problema dado. La propuesta incluye un caso de estudio a modo de prueba de conceptos.

Keywords: Modelado de Sistemas Big Data, Reusabilidad, Variedad, Líneas de Productos Software

1. Introducción

La propiedad de los sistemas Big Data (SBD) [2] con respecto a diversidad de los datos se denomina Variedad; y en [1] se clasifica en una taxonomía que divide el análisis de variedad en cuatro casos de diversidad: estructural, de las fuentes, de contenido y de procesamiento. Por ejemplo, la *diversidad estructural* denota la variedad en formatos y tipos de datos, clasificándolos como estructurados, semi-estructurados y no estructurados; la *diversidad de las fuentes* se clasifica en tres grupos - datos generados por humanos, generados por máquinas o mediados por procesos; la *diversidad de contenido* aborda diferentes tipos de soporte; y la *diversidad de procesamiento* enfoca en las distintas necesidades de procesamiento algorítmico.

La variedad en los datos también ha sido considerada desde el punto de vista de incorporación de semántica al proceso de modelado de arquitecturas en SBDs,

* Este trabajo esta parcialmente soportado por el Proyecto Desarrollo de Software basado en Reuso Parte II

ISBN 978 -987-633-574-4

e incluso ha sido relacionada con diversas propiedades como interoperabilidad, seguridad, reusabilidad, etc. En SBDs, la reusabilidad ha sido abordada también desde diversos ángulos. Por ejemplo, en [9] se discuten conceptos de reusabilidad en el contexto de analítica de datos distinguiendo entre uso y reuso del dato. En otro sentido, incorporando la detección de aspectos comunes y variables a modo de familia de sistemas, en [7] se propone una arquitectura de referencia acotada por medio de casos de uso. De esos casos, se identifican requerimientos relevantes al SBD, incluyendo categorías, como tipos de datos, transformaciones, visualizaciones, etc. Luego, la arquitectura se organiza como una colección de módulos que descomponen la solución en funciones o capacidades para un conjunto de aspectos. En este contexto, y respondiendo a la pregunta de investigación:

RQ: *¿Cómo puede identificarse la variedad de la información de dominio de manera de incorporar reusabilidad en el desarrollo de SBDs?*

este artículo extiende la arquitectura para la construcción de SBDs presentada en [5], incorporando variedad a modo de líneas de productos [10]. A diferencia de la propuesta en [7], que descompone la arquitectura en módulos asociados a intereses guiados por soluciones existentes, nuestra propuesta toma como partida una estructura de etapas asociadas al desarrollo de SBDs, instanciada en artefactos software producidos durante esas etapas, e incorpora el modelado de variedad de contexto de manera similar a líneas de productos. La propuesta se ejemplifica mediante un caso de estudio en el dominio hidrológico a modo de prueba de conceptos.

El artículo se organiza de la siguiente manera. En la sección 2 se introduce nuestro enfoque en el sentido bottom-up y luego, la sección 3 presenta el caso de estudio. Finalmente, se abordan conclusiones y trabajos futuros.

2. Enfoque bottom-up para identificar características de contexto variantes

A partir de la pregunta de investigación que hemos definido en la introducción (RQ), en la Figura 1 mostramos la visión global del enfoque bottom-up de nuestra propuesta, es decir, la identificación de variedad a partir de los datos.

En principio, al centrarnos en SBDs, el primer elemento a considerar es el proceso de desarrollo, donde las etapas básicas pueden resumirse en [6] (círculos centrales de la Figura): (1) *Adquisición de datos*, que consiste en extraer los datos desde las fuentes, agregando un proceso de carga y filtrado para que los datos sean adecuados a su posterior procesamiento; (2) *Transformación y Mejora*, que consiste en estructurar el formato de los datos, realizar la limpieza de los mismos y eventualmente, también su integración; (3) *Análisis*, que contiene las funcionalidades que permiten derivar conocimiento a partir de los datos, enfocando en análisis descriptivo, predictivo y/o prescriptivo; y (4) *Visualización*, que es el punto de acceso a los resultados del proceso.

A su vez, en la Figura podemos observar que nuestro enfoque parte de un proceso bottom-up. Esto significa que las características variantes serán identificadas a partir de un proceso de análisis de datos, donde las variedades serán

ISBN 978 -987-633-574-4

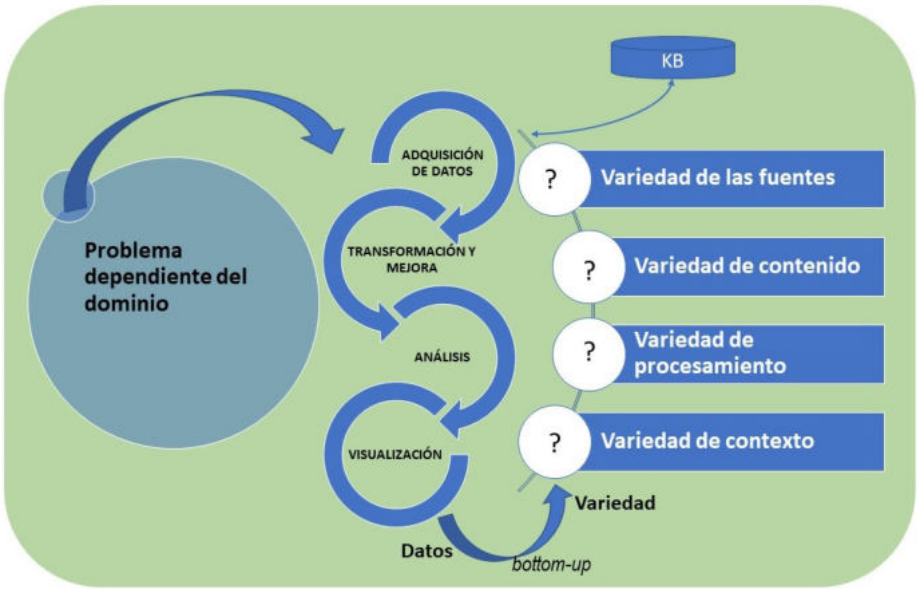


Figura 1: Visión global del enfoque bottom-up: desde los datos al dominio

inferencias a ser corroboradas por expertos de dominio. En contraposición, un enfoque top-down iniciaría con análisis de información de dominio, a ser corroborada por evidencia en las bases de datos. En principio, nuestra propuesta engloba ambos enfoques; sin embargo, en este artículo sólo presentaremos el sentido bottom-up para una mejor comprensión.

Así, partimos desde la *definición de un problema dependiente del dominio* e intentamos detectar características variantes dentro de cada una de las etapas del proceso de análisis de datos. Por ejemplo, la *variedad de las fuentes*, dentro de la etapa de *adquisición de datos* intentará detectar diferencias en las fuentes en cuanto a los cambios de su estructura, datos, formas de adquirirlas, etc. La *variedad de contenido* detectará cambios en las formas en que los datos deben ser transformados y procesados de acuerdo a los objetivos planteados, sobre todo considerando cambios o evoluciones de las fuentes. La *variedad de procesamiento* permitirá detectar variantes en cuanto a las técnicas de análisis posibles de utilizar y por último la *variedad de contexto* permitirá detectar variaciones del dominio que condicionen o cambien los resultados de los análisis.

Nuestro enfoque propone documentar las variedades encontradas en un dominio determinado de forma tal de almacenarlas en una base de conocimiento (KB) para que puedan ser reusadas en las mismas situaciones pero en contextos (dominios o casos) diferentes [5].

En trabajos previos, hemos presentado una propuesta de diseño de Líneas de Productos Software (LPSs) dirigida por funcionalidades, donde cada fun-

ISBN 978 -987-633-574-4

cionalidad se documenta a través de una hoja de datos funcional (*datasheet*), representando el conjunto de servicios comunes[1] y variantes [3,4].

Para el caso de reusabilidad en SBDs, la Figura 2 muestra la hoja de datos funcional definida para reusar modelos de análisis y detectar *variedades de contexto*. En la primera funcionalidad, *Buscar modelo a utilizar*, vemos las acciones necesarias para definir el objetivo del proceso de análisis y recuperar la técnica que se desea aplicar. Allí podemos observar el modelo de variabilidad asociado que posee el punto de variación *técnicas de análisis* con una variabilidad alternativa, es decir, se puede instanciar sólo una de las variantes definidas, ya sea una *red neuronal*, un análisis usando *k-means*, etc. Una vez seleccionada la técnica, en la siguiente datasheet podemos ver la funcionalidad que se despliega por haber elegido el modelo de red neuronal. Aquí debemos buscar los modelos existentes en la KB y determinar si podemos reusarlos, o si existen variedades de contexto que requieren la creación de nuevos modelos. Así, en la Figura vemos asociados dos modelos de variabilidad con puntos de variación opcionales (para el caso de encontrar similitudes entre el modelo a aplicar y los existentes) y puntos de variación alternativos para almacenar un nuevo modelo (cuando el mismo debe ser creado[2]) o reusar alguno existente.

3. Aplicación del enfoque bottom-up al caso de estudio

La calidad del agua es medida por los cambios en los parámetros químicos, ecológicos y espaciales, de los cuales además de estudiar sus valores, hay que ver sus interdependencias. Entre esos parámetros se encuentran la *concentración de pH* (una medida usada para testear acidez), el *Oxigeno Disuelto*, la *Temperatura del Agua*, etc. [8]. La Figura 3 muestra el enfoque bottom-up definido en la sección anterior, pero instanciado a nuestro caso de estudio. A la izquierda de la Figura puede verse un problema de dominio dado, en el cual se debe detectar variedad contextual (*Causas de variación de la temperatura en dos localizaciones de un curso de agua*). En nuestro caso de estudio y a modo de prueba de conceptos, estableceremos estabilidad (no variación) en las fuentes, contenido y procesamiento, intentando identificar variaciones contextuales en el dominio de estudio. La variedad de contexto a identificar consiste en relacionar las inferencias realizadas a partir de los datos con información del dominio (ubicaciones geográficas del curso de agua en **L1** y **L2** que sean caracterizadas en términos de variables comunes y variantes).

Adquisición de los datos

Para el caso de estudio, seleccionamos un dataset que contiene muestras de cuerpos de agua en King County, Washington, Estados Unidos. La cantidad de

[1] Los servicios comunes son aquellos que son parte de todos los productos derivados de la LPS

[2] La restricción «require» entre las variantes determina que si se debe crear un nuevo modelo, debe a su vez guardarse en la KB con la documentación correspondiente

ISBN 978 -987-633-574-4

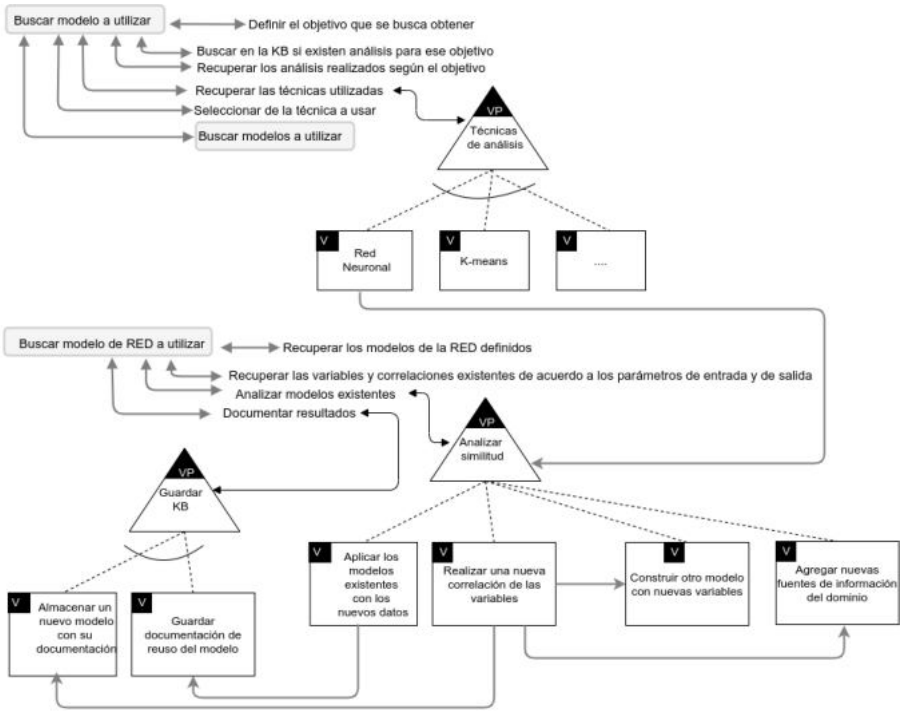


Figura 2: Datasheets para reusar modelos de análisis y detectar *variedades de contexto*

tuplas (1.589.362) y 25 columnas de variables lo hicieron adecuado para nuestra prueba de conceptos. Entre esas variables, se registran la identificación de la muestra, su fecha de recolección, tipo de sitio donde se recolectó (ej. ríos, lagos), punto donde se extraen las muestras (*locator*), el tipo de parámetro correspondiente a la muestra, etc. En particular, esta última variable identifica si el parámetro corresponde a pH, temperatura, turbidez, total de fósforo, coliformes fecales, oxígeno disuelto, conductividad, amoníaco de nitrógeno, densidad o clorofila.

Este dataset es entrada para el análisis de dos localizaciones - **L1** y **L2** (Figura 3), con lo que se mantiene la igualdad de la fuentes y por lo tanto, no hay variedad que se incorpore en esta etapa.

Transformación y mejora

Después de analizar los tipos de datos suministrados y de seleccionar los relevantes al problema abordado, procedimos a la transformación en columnas, agregando aquellas correspondientes a información geográfica de cada muestra en cada localización y seleccionando dos de ellos (**L1** y **L2**) para nuestro análisis. Para identificar relaciones entre los parámetros y la variable *temperatura*, foco

ISBN 978 -987-633-574-4

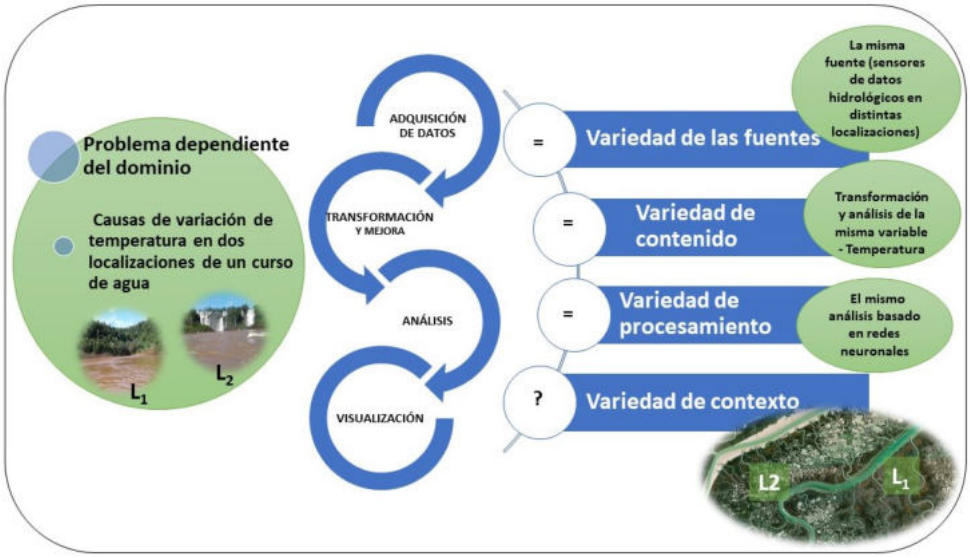


Figura 3: Enfoque bottom-up aplicado al caso de estudio

del problema, realizamos un análisis de correlación de Pearson determinando el grado de intensidad y dirección de las relaciones lineales entre cada par de variables. Este análisis se aplicó primero en **L1**, donde pudimos observar que las variables más relacionadas con el objetivo eran *'Nitrógeno Total'*, *'Alcalinidad Total'*, *'Oxígeno Disuelto del Suelo'*, *'Oxígeno Disuelto'* y *'Conductividad del Suelo'*. En la Figura 4, del lado izquierdo, podemos observar gráficamente estas correlaciones. Al tomar sólo variables significativas para la temperatura en **L1** y graficándo nuevamente las correlaciones, pero ahora con los datos en **L2**, observamos que existen diferencias en la intensidad de las relaciones. Esto llevó a que repliquemos el análisis para obtener las variables más significativas en esta segunda localización.

En la Tabla 1 se observan las relaciones de cada variable con temperatura para **L1** y **L2**. Puede verse que algunas variables, como el *'pH del suelo'*, tienen una fuerte relación con temperatura en **L2** mientras que en **L1** no sucede lo mismo (*'Oxígeno Disuelto'* muestra mayor impacto). Resumiendo en las dos columnas intermedias de la Tabla 1, podemos ver la diferencia que hay entre los valores de ambas localizaciones para cada parámetro y el orden de los mismos de acuerdo a su incidencia.

Análisis

En la Figura 5 mostramos la instanciación del datasheet para este caso. De acuerdo al problema enunciado en la Figura 3 y considerando inexistencia de antecedentes en la base de conocimientos (KB), para este caso se decidió la utili-

ISBN 978 -987-633-574-4

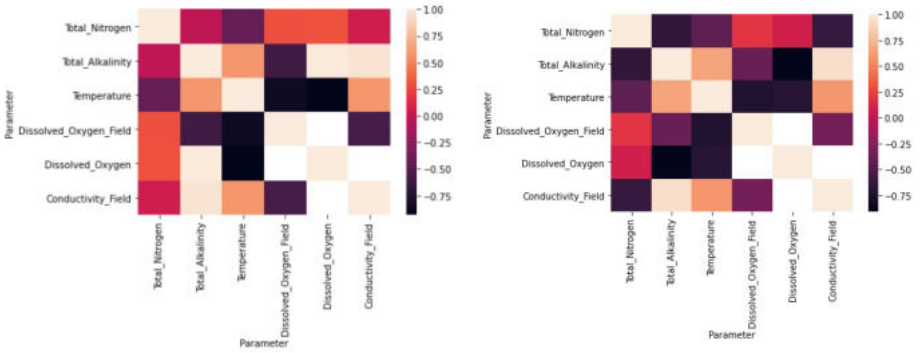


Figura 4: Correlación gráfica de variables seleccionadas en L1 (izquierda) y L2 (derecha)

Parámetro	L1	L2	Parámetro
Oxígeno Disuelto	-0.928462	-0.703626	pH, del Suelo
Oxígeno Disuelto, del Suelo	-0.857255	-0.74100	Silica
Alcalinidad Total	0.591276	0.654430	Nitrito + Nitrato Nitrógeno
Conductividad, del Suelo	0.582982	0.594806	pH
Nitrógeno Total	-0.420579	-0.439275	Conductividad
Ortofosfato de Fósforo	0.320313	0.110690	Nitrógeno Amoniacal
Nitrógeno Amoniacal	0.257200	0.019802	Fósforo Total
Nitrito + Nitrato Nitrógeno	-0.246479	-0.622934	Oxígeno Disuelto
Fósforo Total	0.242684	-0.012616	Ortofosfato de Fósforo
Conductividad	0.208746	0.509993	Sólidos Suspendidos Totales
Sólidos Suspendidos Totales	-0.202858	-0.038395	Coliformes Fecales
Coliformes Fecales	0.047256	0.205362	Turbidez
Turbidez	-0.171750	-0.029406	E. coli
E. coli	0.111787	0.253594	Oxígeno Disuelto, del Suelo
Silica	0.073081	0.527815	Alcalinidad Total
pH	-0.050635	0.424880	Enterococo
Enterococo	0.042639	0.093202	Nitrógeno Total
pH, del Suelo	0.011134	0.557562	Conductividad, del Suelo

Cuadro 1: Correlación de parámetros con temperatura en L1 y L2

ISBN 978 -987-633-574-4

zación de redes neuronales con el objetivo de estimar el valor de la temperatura, en base a otros parámetros relacionados, para **L1** y **L2**.

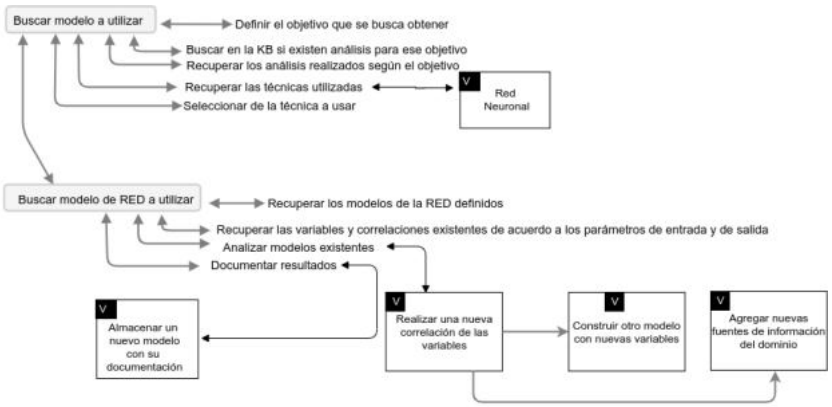


Figura 5: Modelo de Variabilidad instanciado para reflejar nuestro caso de estudio

Así, el primer modelo M_1 fue entrenado en **L1** y se procedió a comprobar la variación entre la temperatura real y estimada, calculando la diferencia entre las mismas. Para M_1 en L1 el valor promedio de las diferencias, calculadas para una serie de variables aleatorias, fue de **2.56841430**. Luego, el modelo M_1 se reutilizó en **L2** sin modificación en la configuración y sin realizar ningún ajuste de contexto (Figura 5 "Recuperar los modelos de la RED definidos"). El resultado de esta reutilización arrojó una diferencia promedio de **3.43957511** (Figura 6 (a)).

Como pudimos observar en la Tabla 1, los valores de las correlaciones de la temperatura en **L1** y **L2** muestran variaciones entre ellos (Figura 5 "Recuperar las variables y correlaciones ..."). Teniendo en cuenta esta variedad, se definió un nuevo modelo, M_2, con la misma arquitectura de M_1 (sin variedad de procesamiento) pero cambiando las variables de entrada. Mientras que en M_1 se utilizaron las variables *'Nitrógeno Total', 'Alcalinidad Total', 'Oxígeno Disuelto, del Suelo', 'Oxígeno Disuelto' y 'Conductividad del Suelo'*; para M_2 se utilizaron *'Alcalinidad Total'*, *'Oxígeno Disuelto, del Suelo'*, *'Oxígeno Disuelto'*, *'Nitrito + Nitrato de Nitrógeno'*, *'Conductividad del Suelo'*, *'pH del Suelo'* y *'Silica'*. Este nuevo modelo se ajusta a las características de contexto de **L2**, por lo que debería tener mejor desempeño en la predicción con respecto al valor obtenido al reusar M_1 (Figura 5 "Analizar los modelos existentes"). Efectivamente, en la Figura 6 (b) puede observarse la ejecución de M_2, mostrando una diferencia en promedio **2.303304169** que es menor que la obtenida con la ejecución anterior.

Ahora, el nuevo modelo documentado y almacenado en la KB podrá ser en un futuro elegido en otro contexto con características similares (Figura 5 "Documentar resultados").

ISBN 978 -987-633-574-4

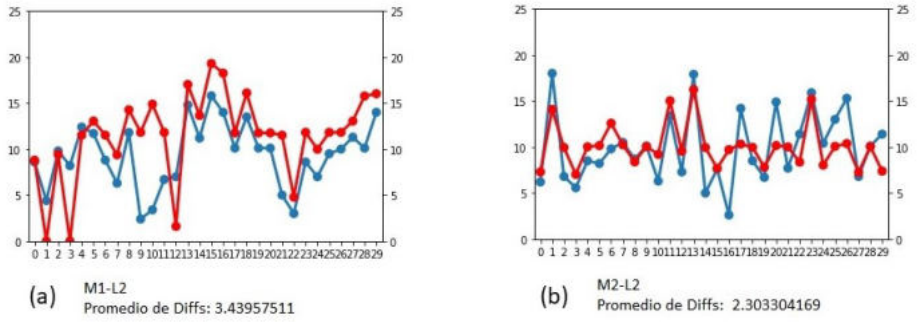


Figura 6: Diferencias entre valores reales y estimados de M_1 en L2 (a) y M_2 en L2 (b)

Visualización

La Figura 6 permite visualizar las ejecuciones realizadas durante el análisis. Sin embargo, existen posibilidades adicionales en el ejemplo presentado. La Figura 7 muestra las localizaciones de **L1** y **L2** en el espacio geográfico. Como parte del análisis, es interesante notar que, aunque ambas pertenecen al mismo curso de agua, su entorno es bastante diferente. Mientras que **L1** se encuentra en una zona relativamente urbanizada, **L2** es un área boscosa con poca intervención humana. Esas características, que se obtienen a simple vista, podrían complementar las condiciones de contexto que den explicación a las diferentes variaciones. Por ejemplo, en **L2** el *'pH del Suelo'* es mucho mayor que en **L1** - probablemente debido a la zona forestada y al tipo de vegetación. Este tipo de inferencias, debidamente contrastadas por expertos de dominio, podrían enriquecer la caracterización de cada localización y almacenarse para futuras identificaciones de áreas geográficas donde el mismo problema sea relevante.

4. Conclusiones y Trabajo Futuro

En este artículo, hemos presentado una propuesta para incorporar reusabilidad en el modelado de sistemas big data, identificando la manera en que la variedad de contexto puede impactar en actividades típicas como la transformación y el análisis de los datos.

Como hemos visto, la participación activa de expertos de dominio es fundamental para la definición del problema y para el análisis de los resultados. En ese sentido, actualmente, estamos definiendo casos de estudio con el acompañamiento de expertos del INTA (Instituto Nacional de Tecnología Agropecuaria) para extender la propuesta con su enfoque top-down y validar los resultados de ambos enfoques en SBDs para el análisis de la napa freática, en función de la variedad de fuentes acuíferas de diversas zonas geográficas.

ISBN 978 -987-633-574-4

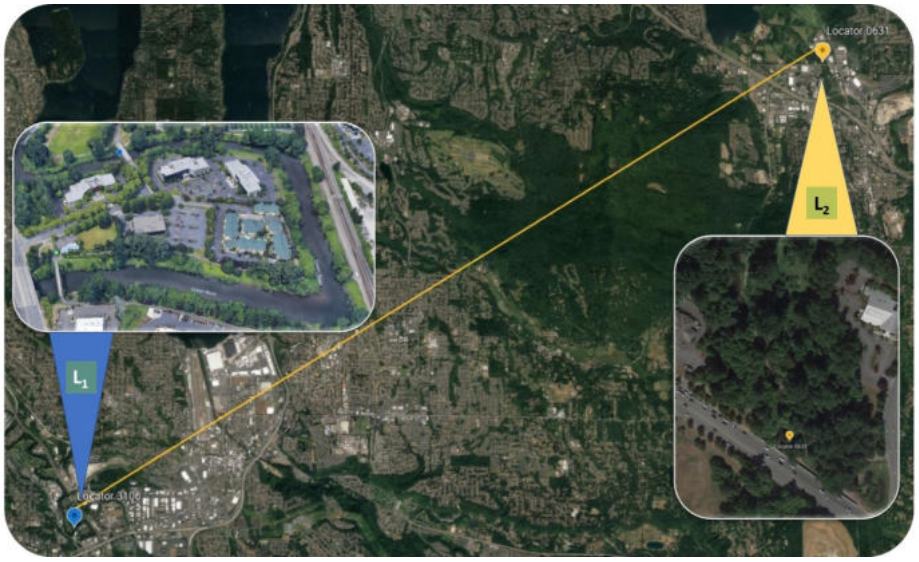


Figura 7: Localizaciones de L1 y L2 en el espacio geográfico

Referencias

1. Abawajy, J.: Comprehensive analysis of big data variety landscape. International Journal of Parallel, Emergent and Distributed Systems 30(1), 5–14 (2015)
2. Bahga, A., Madisetti, V.: Big Data Science & Analytics: A Hands-On Approach. VPT, 1st edn. (2016)
3. Buccella, A., Cechich, A., Arias, M., Pol'la, M., Doldan, S., Morsan, E.: Towards systematic software reuse of gis: Insights from a case study. Computers & Geosciences 54(0), 9 – 20 (2013)
4. Buccella, A., Cechich, A., Pol'la, M., Arias, M., Doldan, S., Morsan, E.: Marine ecology service reuse through taxonomy-oriented SPL development. Computers & Geosciences 73(0), 108 – 121 (2014)
5. Buccella, A., Luzuriaga, J., Cechich, A., Osycka, L., Paterno, F., Pol'la, M., Cruz, M., Martinez, R., Mazalu, R., Moyano, M.: Reusabilidad en el contexto de desarrollo de sistemas para big data. In: Actas del XXIII Workshop de Investigadores en Ciencias de la Computación, Chilecito, La Rioja. pp. 525–529 (2021)
6. Davoudian, A., L., M.: Big data systems: A software engineering perspective. ACM Computing Surveys 53(5) (2020)
7. Klein, J.: Reference architectures for big data systems, carnegie mellon university's software engineering institute blog. http://insights.sei.cmu.edu/blog/reference-architectures-for-big-data-systems/ (Accessed June 9, 2021) (2017)
8. Loucks, D.P., van Beek, E.: Water Resource Systems Planning and Management: An Introduction to Methods, Models, and Applications. Springer (2017)
9. Pasquetto, I., Randles, B., Borgman, C.: On the reuse of scientific data. Data Science Journal 16(8) (201720)
10. Pohl, K., Böckle, G., Linden, F.J.v.d.: Software Product Line Engineering: Foundations, Principles and Techniques. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2005)

ISBN 978 -987-633-574-4

Aplicación de contratos inteligentes y blockchain como apoyo en la implementación de sistemas de gestión basados en ISO 9000.

Kristian Petkoff Bankoff. Ariel Pasini, Marcos Boracchia, Patricia Pesado

Instituto de Investigación en Informática LIDI (III-LIDI)*

Facultad de Informática – Universidad Nacional de La Plata 50 y 120 La Plata Buenos Aires
*Centro Asociado Comisión de Investigaciones Científicas de la Pcia. de Bs. As. (CIC)

{kpb,apasini,marcosb,ppesado}@lidi.info.unlp.edu.ar

Abstract. A los términos blockchain y contratos inteligentes generalmente se los relaciona con temas financieros, pero existen otro tipo de aplicaciones que pueden aprovechar los beneficios de esta tecnología. Los sistemas de gestión de la calidad (SGC) requieren de la validación de autenticidad de documentos y evidencias de los procesos que podrían ser beneficiados con el uso de estas tecnologías, Se presenta una propuesta de aplicar contratos inteligentes para el control documental de un SGC y el resguardo de las evidencias generadas por el proceso del mismo.

Keywords: Blockchain, contratos inteligentes, Sistemas de gestión de la calidad, ISO 9001

1 Introducción

Con el auge que tienen en la actualidad los términos blockchain y contratos inteligentes muchas organizaciones buscan conocer estas tecnologías y evaluar la viabilidad de su uso como soporte para sus operaciones. Habitualmente se encuentran implementaciones relacionadas con las finanzas, la logística, trazabilidad de alimentos y medicamentos, soluciones de identificación/autorizaciones personales. Los sistemas de gestión (de la calidad, ambiental, de salud y seguridad en el trabajo, entre otros) por otra parte plantean una serie de procedimientos estandarizados para que las organizaciones mejoren sus procesos y optimicen sus operaciones con la premisa de documentar toda su información, trazabilidad de los procesos, seguimiento de objetivos y desvíos. La familia de normas ISO (International Organization for Standardization) relacionadas con los sistemas de gestión se ha ido adaptando para mantener una estructura estándar, con lo que hoy se encuentran similitudes generales entre las últimas versiones de las familias ISO 9000 (gestión de la calidad), ISO 14000 (gestión ambiental), ISO 22000 (seguridad alimentaria), ISO 27000 (seguridad

de la información), ISO 45000 (seguridad y salud en el trabajo), entre otras. Las organizaciones podrían optimizar el proceso de auditoría y de control de la documentación utilizando contratos inteligentes en una blockchain de la que participen las autoridades certificantes, de manera de mitigar las no conformidades por errores humanos en la gestión de documentos, con una solución adaptable a cualquiera de las normas citadas y las que tengan similitudes en general con ISO 9000.

En la sección 2 se desarrollan los conceptos generales de los contratos inteligentes y de blockchain. A continuación, características generales de los sistemas de gestión. Luego, se plantean puntos de contacto entre las necesidades de control y auditoría de los documentos y las soluciones que aportan estas tecnologías para plantear algunos casos de uso y analizar su viabilidad. Por último, se presentan las conclusiones.

2 Conceptos generales

2.1 Contratos inteligentes

Los contratos inteligentes son aplicaciones descentralizadas que ejecutan un algoritmo ante un evento con el propósito de hacer cumplir un acuerdo entre pares. Proponen digitalizar la celebración y ejecución de contratos legales con la premisa de que el objeto de contrato no es llevado a cabo si no se cumplen previamente las condiciones acordadas. El algoritmo garantiza que el objeto del contrato no se cumple si no están dadas las precondiciones y una vez registrada una transacción no debe permitirse su anulación.

Para lograr esta característica de inmutabilidad en las transacciones es habitual hoy en día el uso de blockchain como base de datos subyacente para la implementación de los contratos inteligentes. Esto es así ya que blockchain provee una estructura donde almacenar las transacciones (y los propios contratos, eventualmente) que es descentralizada, incremental e inmutable [1].

2.2 Blockchain

Blockchain es una estructura de datos en la que la información es agregada en bloques enlazados de manera que todo nuevo dato hace referencia al predecesor y además por medio de un algoritmo criptográfico se asegura la inmutabilidad de la cadena ya que un cambio en un bloque antecesor debería producir un cambio en cascada a todos los sucesores para mantener la validez de la estructura.

Esta tecnología surge para proveer la base sobre la que se almacena toda la información de las transacciones de Bitcoin, que además opera sobre una red de nodos descentralizados, en la que cada uno mantiene una copia completa de la cadena

y participa del proceso de validación de los nuevos bloques que se agregan a la misma [2].

La aplicación Bitcoin es un contrato inteligente y está respaldada por su estructura de datos subyacente Blockchain que posee características de integridad, inmutabilidad, anonimato, descentralización, y no confianza entre los nodos.

La no confianza entre los nodos que forman parte de la red (de la mano de la descentralización) implica que la disputa sobre la validez de las transacciones o bloques a agregar a la base de datos se debe resolver mediante un protocolo de consenso sin arbitraje de una entidad específica.

El mantenimiento de los datos y sobre todo el alta de nuevos bloques en la cadena generan algunas de las principales desventajas de su uso: el costo de las operaciones, la escalabilidad y vulnerabilidades propias del consenso. Agregar nuevos datos a una blockchain requiere que los nodos reciban y validen la operación (ejecutando las operaciones que el contrato inteligente determina), generen el nuevo bloque de datos y luego validen al bloque mediante algún algoritmo de consenso para finalmente sumarlo a la cadena. El aumento del costo se evidencia en todas las acciones adicionales que se realizan con respecto a una aplicación tradicional que solamente validaría los datos de entrada y registraría los resultados; además, existe una sobrecarga en la transferencia de información a través de la red que es otro costo extra al cómputo. Por otra parte, existen problemas en las distintas blockchains relacionados con la escalabilidad, ya que a medida que se agregan nodos, la operatoria se ralentiza significativamente en virtud de la necesidad de intercambiar información entre todos los pares. El algoritmo de consenso, por último, caracteriza significativamente la performance y la confiabilidad de la cadena de bloques; por una parte se necesita garantizar que un bloque es agregado solo si es validado por los pares, pero además este proceso debe realizarse con una velocidad que resulte aceptable y sin que un usuario malicioso pudiese apoderarse del consenso acaparando una cantidad de nodos (vulnerabilidad del 51%, situación en la cual un usuario podría forzar la aceptación de cadenas que contengan operaciones que le beneficien a pesar de que no sean válidas) [3,4].

2.3 Adopción de blockchain en organizaciones

Bitcoin surge como una alternativa para realizar operaciones financieras entre personas sin estar sujeto a las regulaciones de las entidades bancarias y los gobiernos. Sin embargo, el surgimiento de diversos proyectos despertó el interés de las propias entidades financieras que comenzaron a implementar blockchains para algunas de sus aplicaciones; el número de proyectos que se respaldan en esta tecnología continúa creciendo y distintos tipos de negocio se adaptan al uso de contratos inteligentes [4]. También existen experiencias de aplicación de blockchain para gobierno electrónico [5].

ISBN 978 -987-633-574-4

Con el uso de blockchain en organizaciones tanto privadas como públicas surge una caracterización de éstas como públicas, privadas, federadas, e incluso Blockchain-as-a-Service en proveedores de servicios en la nube. Una blockchain que no es pública permite circunscribir la red a nodos previamente autorizados a formar parte e incluso establecer quiénes pueden invocar funciones de los contratos inteligentes como contraposición a una cadena pública donde cualquier usuario puede invocar un contrato o formar parte de la red como nodo. Los esquemas privados, federados o de consorcios pueden suponer un consenso forzado por una autoridad central, pero estas implementaciones se fundamentan en el aprovechamiento de las otras características de blockchain [4], exceptuando así la no confianza entre nodos.

3 Sistemas de gestión de la calidad

Un sistema de gestión es un conjunto de documentos, formularios y registros que describen los procesos de una organización, los objetivos, los indicadores del cumplimiento de éstos y los procedimientos para registrar las mediciones. La Organización Internacional de Estandarización (ISO) es define distintos marcos según la naturaleza de la organización o del ámbito específico al que se desea aplicar un sistema de gestión, enumerando una serie de documentos y procedimientos mínimos (y genéricos) a cumplimentar para alcanzar una certificación en la implementación del sistema.

La generalización de los documentos y procedimientos puede llevarse a un nivel mayor y extraer un factor común entre distintos tipos de sistemas de gestión; en todos ellos, por citar solo un ejemplo, se debe definir una política de calidad que orienta los objetivos, se deben registrar los cambios en cada documento del sistema generando números de versión incrementales y se debe guardar registro de la ejecución de los procedimientos verificando el uso de los medios que el mismo procedimiento (o el documento que corresponda) especifica.

Por otra parte, lo anterior debe cumplirse sin descuidar que en un sistema de gestión debe registrarse la detección y el tratamiento de los desvíos con respecto a los objetivos, ya que esto es en definitiva lo que coadyuva a alcanzar la mejora continua. Durante una auditoría, una de las principales actividades es, entonces, verificar que existan registros de esas mediciones, que los desvíos estén documentados y tratados, y además que se utilicen los documentos y formularios correctos para cumplir con estas tareas.

Si bien en principio la implementación de estos procesos de control genera un esfuerzo adicional, y sin perder de vista que se plantea un punto de contacto entre los sistemas de gestión y el uso de contratos inteligentes, también es cierto que la mejora en la calidad de las operaciones de la organización genera una optimización de sus costos [6].

ISBN 978 -987-633-574-4

3.1 Similitud entre sistemas de gestión de la calidad

Como se mencionó anteriormente, la propia ISO ha promovido en los últimos años cambios en sus estándares relacionados con la gestión de procesos de las organizaciones para extraer el factor común, estandarizando la propia estructura de sus normas. De esta manera aparecen conceptos comunes presentes en todas ellas y además una serie de capítulos que son iguales o equivalentes entre sí.

Si bien en versiones anteriores de normas como ISO 9001:2008 se exigía una estructura documental predeterminada (para los procesos de la organización, por ejemplo, en el llamado Manual de la calidad), este requisito también se ha modificado para dar paso a una mayor flexibilidad (concepto de Información documentada). Esto implica que ya no sea obligatorio, aunque sí recomendable, definir formalmente los procedimientos de la organización, sino que el requisito es contar con registros de los procesos y del tratamiento de las no conformidades.

Esta mayor flexibilidad supone un desafío para el desarrollo de aplicaciones de software genéricas que se adapten a cualquier contexto organizacional y estructura documental.

En líneas generales, en ISO 9001, ISO 14001, ISO 45001 y otras se describe la terminología y el objetivo de cada estándar en sus primeros tres capítulos, dando paso luego a los requisitos específicos del sistema (de gestión de la calidad, de gestión ambiental, de gestión de seguridad y salud en el trabajo respectivamente en los tres ejemplos citados). Dadas las similitudes, las organizaciones pueden incluso proponerse la certificación de varias normas realizando cierto esfuerzo de adaptación y reutilización [7]. En la tabla1 se muestra la estructura de estas normas en sus últimas versiones.

Tabla 1. Similitud en la estructura de las normas.

Capítulo	**ISO 9001:2015, ISO 14001:2015, ISO 45001:2018**
4	Contexto de la organización
5	Liderazgo y compromiso
6	Planificación
7	Soporte
8	Operación
9	Evaluación de desempeño
10	Mejora

En cada uno de estos capítulos se describe qué tipo de información se debe documentar; las normas no establecen ningún formato ni los detalles específicos, sino que sirven como una guía para que la organización pueda adaptar el sistema de gestión a su contexto particular. Por la naturaleza de cada clase de documento, la cantidad de partes interesadas en el aspecto que trata, quiénes son los responsables, cómo afecta (o describe) a los procesos operativos y de gestión del sistema se puede

elaborar un esquema de categorías según la cantidad de documentos y la frecuencia de actualización de versiones que suelen tener. Estas características son recopiladas en la tabla 2, donde se refleja la diferencia de volumen documental típica entre los capítulos y también la frecuencia de actualización que suelen tener los documentos que los componen, que finalmente repercute en los procesos de gestión y control de versiones del sistema documental.

Tabla 2. Visión comparativa de la cantidad de documentos y frecuencia de actualización entre los distintos capítulos.

Capítulo	**Cantidad típica de documentos**	**Frecuencia de actualización**
Contexto de la organización	Baja.	Muy baja.
Liderazgo y compromiso	Baja.	Muy baja pero con alta demanda de distribución y control de versión.
Planificación	Baja.	Baja.
Soporte	Moderada.	Moderada.
Operación	Alta.	Alta, con alta demanda de distribución y control de versión.
Evaluación de desempeño	Baja.	Baja.
Mejora	Baja.	Baja.

3.2 Dificultades comunes en el control de documentos y auditoría externa.

En el apartado anterior surge una problemática a la hora de implementar un sistema de gestión conforme con algunos estándares de ISO y superar una auditoría para su certificación o recertificación: el control documental en sí mismo, que es común a todos los tipos de sistemas y organizaciones, y que no forma parte de las incumbencias específicas de los procesos principales que se someten al control. No es extraño hallar durante un proceso de auditoría situaciones como el uso de un documento en una versión anterior a la vigente o bien documentos que a pesar de estar vigentes carecen de la marca de control documental, lo que no constituye un error en el proceso operativo principal o de soporte, pero sí formal de los procesos de control del sistema de gestión implementado [8].

La cantidad de documentos que conforman el sistema de gestión puede ser muy grande, y como se desprende de la tabla 2, en algunos casos su frecuencia de actualización puede ser relativamente alta y ser de uso de una gran cantidad de actores; esto es especialmente cierto para los documentos relacionados con los

registros de los procedimientos primarios, con el tratamiento de las no conformidades y en general con todo el capítulo de soporte y el de operación.

El uso de un determinado formulario para registrar acciones o mediciones por parte de diferentes personas requiere que se garantice el acceso a la última versión vigente y también la validación de esta situación a la hora de incorporar y almacenar el registro; por otra parte, durante una auditoría externa también se debe validar que la documentación puesta a disposición de las partes interesadas se corresponde con la versión vigente.

La flexibilidad a la hora de especificar el sistema de gestión, por otra parte, constituye una dificultad para implementar o adaptar software empaquetado sobre todo si se debe controlar el uso de versiones vigentes para adjuntar registros: estos pueden tener un formato arbitrario según las necesidades de la organización. La solución más popular es la de utilizar aplicaciones de ofimática y mantener los registros y documentos almacenados en archivos; se presenta como una alternativa económica ya que no requiere la adquisición de software especializado y la capacitación se puede afrontar con menores recursos. En contraposición, esta opción requiere un esfuerzo mayor para controlar las versiones y está más expuesta a errores humanos.

En otra dimensión del problema, durante el proceso de auditoría externa para acceder a la certificación (o bien renovar una certificación ya obtenida) un auditor debe revisar los detalles de implementación del sistema de gestión comenzando por interiorizarse en el contexto de la organización; debe comprender procesos dentro del alcance del sistema, verificar que la documentación se encuentre completa, que exista evidencia de los registros, tratamiento de las no conformidades, posibles no conformidades no tratadas, que se realicen acciones de mejora, más todos los procesos de soporte y de enfoque al cliente. Recordando como se mencionó anteriormente que la flexibilidad del sistema permite a cada organización estructurar la documentación y mantenerla en el formato que mejor se adecue a su realidad y posibilidades, la tarea del auditor sin herramientas que den soporte a su tarea es especialmente ardua.

4 Registro y validación de documentos apoyándose en contratos inteligentes.

El uso de contratos inteligentes como apoyo para algunos de los procesos de la gestión de los documentos y evidencia de registros permite aprovechar algunas ventajas de esta tecnología para mitigar los riesgos propios de la implementación y certificación de los sistemas de gestión bajo normas basadas en ISO 9000 y similares.

Se deben considerar las ventajas y desventajas de esta opción como así también tener en cuenta ciertos detalles al realizar el diseño del software. Con respecto a esto último, cabe destacar que el almacenamiento de información en la blockchain es

extremadamente costoso y por lo tanto no es conveniente para el almacenamiento de archivos en general; para este caso de uso existen implementaciones específicas de almacenamiento distribuido basado en blockchain como por ejemplo se plantea en [9] y se prefiere utilizar los contratos inteligentes para registrar referencias a los archivos [10].

Por otro lado, el auge de la tecnología blockchain y de los contratos inteligentes está sostenido por una comunidad activa de desarrolladores y por una corriente de investigaciones que promueven la implementación de soluciones que pueden adaptarse a los requisitos de gestión de los documentos de las normas ISO 9000 y similares, como el registro de autor y verificador de los documentos [10].

A continuación, se proponen cuatro procedimientos que buscan asistir a los usuarios a la hora de utilizar formularios controlados para agregar registros al sistema de gestión. Como se menciona en el apartado 3, una falla que se encuentra con cierta frecuencia es el uso de versiones incorrectas de los formularios o la distribución de versiones no vigentes cuya trazabilidad es dificultosa según cómo la organización decida estructurar su sistema documental.

4.1 Registro de nuevo documento / nueva versión

Este procedimiento recibe como entrada un código de documento según el esquema que la organización adopte para su estructura documental, el número de versión del documento, la fecha de publicación de la versión y el archivo en formato PDF.

La aplicación cliente calcula el hash del archivo utilizando el algoritmo keccak256 e invoca una función del contrato inteligente para validar que la versión se haya incrementado con respecto a la anterior y que el hash sea diferente del de la versión vigente (se contempla como positivo el caso de ser un nuevo documento en su primera versión registrada).

En la figura 1 se muestra el diagrama de secuencia del registro de un nuevo documento; éste puede tratarse de cualquiera de los requeridos/sugeridos por la norma y en particular un formulario utilizado para agregar registros o evidencia de registros al sistema de gestión. Este diseño no hace ninguna asunción sobre el tipo de almacenamiento a utilizar; la aplicación cliente es responsable de gestionar este aspecto y basta con que el archivo sea recuperable a través de alguna referencia (URI) almacenada en el propio contrato inteligente.

El documento registrado es almacenado tal cual, sin modificaciones, y se computa el hash del mismo para posteriores validaciones. El código QR generado e insertado en el documento contiene información del registro para permitir a la aplicación determinar que se trata del documento correcto en el diagrama de la figura 3.

Es importante remarcar que la verificación propuesta no persigue objetivos de seguridad o integridad de los archivos, sino que busca prevenir errores involuntarios

ISBN 978 -987-633-574-4

de los participantes del sistema de gestión y asistir a la validación de las versiones de los documentos durante la auditoría.

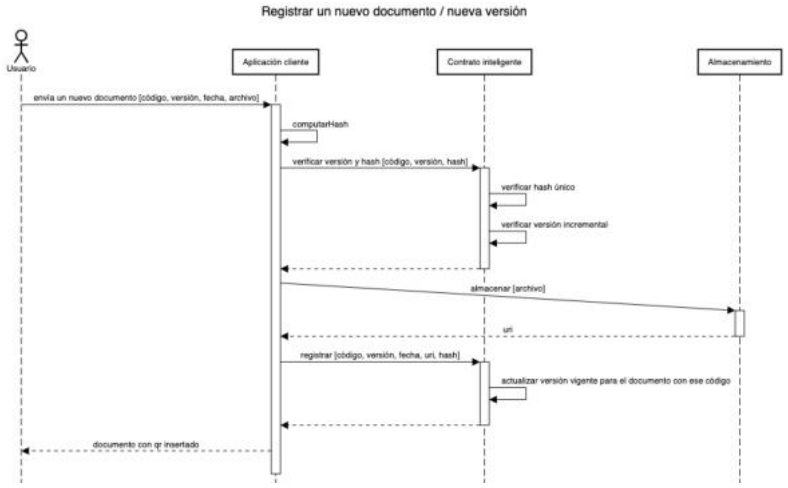


Figura 1. Diagrama de secuencia de registro de un documento

4.2 Descarga de copia controlada

Como se mencionaba en el procedimiento anterior, los archivos de los documentos son almacenados en un servicio externo a la aplicación sin incluir ninguna modificación, lo que permite que al computar su hash se pueda validar la integridad.

El procedimiento de descarga de documento recibe como entrada un código de documento, obtiene del contrato inteligente la URI donde se persiste el archivo más el hash registrado y la aplicación cliente valida que el archivo obtenido sea el originalmente registrado computando y comparando su hash contra el almacenado en la blockchain, tras lo que inserta el código QR y texto típico con el código de documento y número de versión y produce la salida al usuario. Esta marca no necesariamente constituye la que requiere el propio sistema de gestión, aunque es compatible con el mismo.

ISBN 978 -987-633-574-4

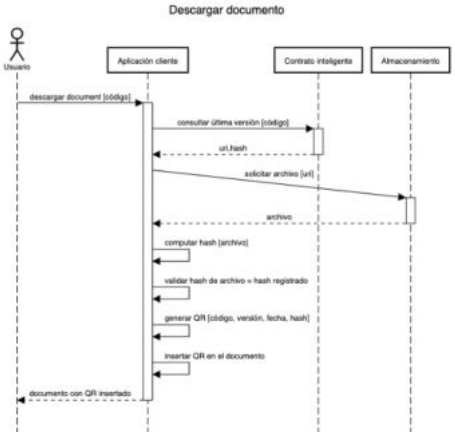


Figura 2. Diagrama de secuencia de descarga de un documento registrado.

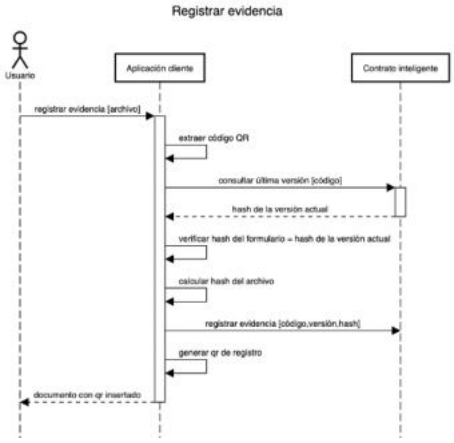


Figura 3. Diagrama de secuencia de validación de registro con formulario controlado.

4.3 Registro de evidencia

Este procedimiento recibe como entrada un archivo representando un formulario completo por el participante del sistema de gestión. Este formulario debe cumplir como precondición haber sido descargado mediante el procedimiento de descarga de copia controlada o bien durante el registro de nuevo documento / versión. A modo de sugerencia, se puede utilizar la funcionalidad de formularios en PDF o bien cualquier otro mecanismo que permita incluso capturar la imagen del formulario con los datos y donde el código QR sea visible.

La aplicación cliente extrae el código QR del archivo, que contiene la información que permite identificar al documento y el hash del mismo, obtiene los datos registrados para ese documento en el contrato inteligente y valida que se trate de la versión vigente del formulario en cuestión.

En el contexto de un sistema de gestión genérico, un registro es un formulario cuyos campos de entrada han sido completados con datos; la solución que se presenta no almacena los registros propiamente dichos, sino que valida el formulario utilizado y genera como salida el mismo documento con un código de verificación estampado.

4.4 Verificación de vigencia documento

Este procedimiento recibe como entrada los datos de un código QR ya capturado, que contiene el código de documento y el hash del archivo generado al registrarlo, obtiene los datos del documento vigente con dicho código en el contrato inteligente y compara el hash registrado con el capturado en el código QR. Se toma como base la

verificación utilizada en el procedimiento anterior, pero esta vez extendiéndolo a cualquier tipo de documento. La verificación permite fácilmente conocer el estado de validez de cualquier copia sin necesidad de verificar manualmente los listados de documentos del sistema de gestión.

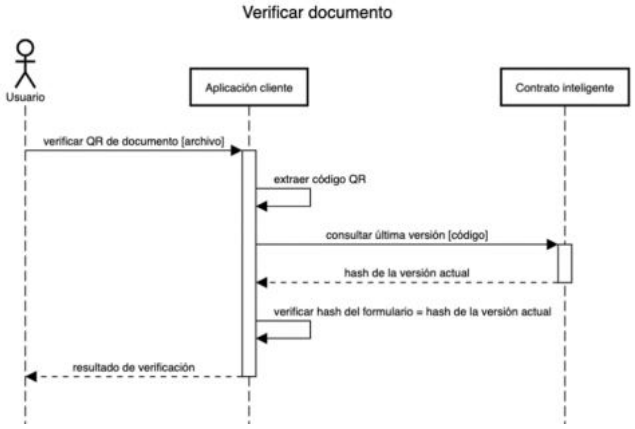


Figura 4. Diagrama de secuencia para la validación de vigencia de un documento.

5 Conclusiones

Luego de analizar casos de uso de blockchain y contratos inteligentes en organizaciones privadas, características generales de la tecnología y por otra parte experiencias en la implementación y certificación de sistemas de gestión en organizaciones pequeñas, se han planteado conceptualmente soluciones posibles a una de las problemáticas que se enfrentan y que muchas veces son detectadas como no conformidades durante una auditoría. Esto abre camino a la expansión de la solución para incluir los datos de los registros en la propia blockchain, lo que permitiría que un contrato inteligente pudiese detectar e incluso registrar las etapas de tratamiento de las no conformidades, que es un aspecto muy sensible en la implementación de los sistemas de gestión.

A diferencia de una aplicación tradicional, el uso de contratos inteligentes permite que las organizaciones que implementan sistemas de gestión formen parte de una red de nodos junto con las que certifican los sistemas, como parte de un ecosistema de “auditoría continua”; por otra parte, la flexibilidad a la hora de escribir los algoritmos permite que, utilizando una infraestructura predefinida, cada organización pudiera agregar y personalizar los contratos según sus necesidades.

Estas ventajas deben ser siempre contrastadas con los inconvenientes propios de blockchain para asegurar que se alcanzan los requerimientos del sistema: el problema de la escalabilidad, la ineficiencia en las operaciones, el costo de las transacciones, el impacto ambiental. Sin embargo, en estos últimos dos aspectos no hay que dejar de

lado que la adopción de sistemas de gestión tiende a la disminución de los costos operativos porque ayuda a anticipar los errores y mejorar los procesos, a la vez que puede asistir a la digitalización de los documentos de soporte disminuyendo el uso de papel.

6 Referencias

[1] Hu, Y., Liyanage, M., Manzoor, A., Thilakarathna, K., Jourjon, G., Seneviratne, A. (2019). Blockchain-based Smart Contracts -Applications and Challenges. arXiv:1810.04699v2 [cs.CY]

[2] Nakamoto, S. (2008). "Bitcoin: A peer-to-peer electronic cash system."

[3] Sayeed, S., Marco-Hisbert, H. (2019). Assessing Blockchain Consensus and Security Mechanisms against the 51% Attack. DOI:10.3390/app9091788

[4] Ben H., Elyes, Brousmiche, K. L., Levard, H., Thea, E.. (2017). Blockchain for Enterprise: Overview, Opportunities and Challenges. International Conference on Wireless and Mobile Communications, ICWMC.

[5] Preisegger, J.S., Muñoz, R., Pasini, A., Pesado, P. (2019). Blockchain y gobierno digital. II Track de Gobierno Digital y Ciudades Inteligentes. oai:sedici.unlp.edu.ar:10915/91367

[6] Silvera, J. A., Figueroa, D. A., Gil, G. D., Sánchez, E., Orosco, C. I. (2013). Ingeniería de software aplicada a un sistema de gestión de calidad en centros educativos. XV Workshop de Investigadores en Ciencias de la Computación, pp 517-520.

[7] Garbarini, R., Cigliuti, P., Burstyn, A., Pollo-Cattaneo, F. (2013). Implementación de un sistema de gestión de calidad y servicios en laboratorio universitario de Ingeniería en Sistemas de Información. VIII Congreso de Tecnología en Educación y Educación en Tecnología.

[8] Dias, R., Arrojo, C. D., Nastta, H. A., Herlein, M. E., Álvarez Martini, C. A., Scaramutti, J. C., Danessa, F. (2019). Implementación de un sistema de la calidad según norma ISO/IEC 17025 en un laboratorio de ensayos eléctricos del ámbito universitario. V Jornadas de Investigación, Transferencia y Extensión de la Facultad de Ingeniería (UNLP), pp 419-426. ISBN 978-950-34-1749-2

[9] Ali, S., Wang, G., White, B., Cotrell, R. L.. (2018). 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering.

[10] Nizamuddin, N., Salah, K., Ajmal Azad, M., Arshad, J., Rehman, M. H. (2019). Decentralized Document Version Control using Ethereum Blockchain and IPFS. Computers & Electrical Engineering 76. DOI:10.1016/j.compeleceng.2019.03.014

ISBN 978 -987-633-574-4

Expanding the scope of a testing framework for Industry 4.0

Martín L. Larrea[1,2,3] and Dana K. Urribarri[1,2,3]

[1] Department of Computer Science and Engineering, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina
[2] Computer Graphics and Visualization R&D Laboratory, Universidad Nacional del Sur (UNS) - CIC Prov. Buenos Aires, Bahía Blanca, Argentina
[3] Institute for Computer Science and Engineering, Universidad Nacional del Sur (UNS) - CONICET, Bahía Blanca, Argentina
{mll, dku}@cs.uns.edu.ar

Abstract. Software has become a transversal and foundational element of the new industrial revolution, with a growing presence in every stage of production processes. Software inclusion has grown not only quantitative but also qualitative; this increase turns critical the impact of software on society, the environment, and individuals. This new era requires new methodologies, techniques, and tools to verify and validate the foundations of Industry 4.0. This paper aims to expand the capacity of a current testing framework to be more flexible and require less specialized personnel. We introduced a web application that complements it through the generation of test cases. The framework and web application are open sources and freely available which means that these software elements are a foundation that will allow future development teams to continue expanding their functionality and application areas within Industry 4.0.

Keywords: Industry 4.0, Verification and Validation, Testing, Message Sequence Specification, Aspect-Oriented Programming, Software

1 Introduction

Software quality has become one of the most important factors in determining the success of products or companies in the era of Industry 4.0. As stated by Oztemel et. al [9] "...the superior quality of the manufacturing industry strictly depends on its high-quality applied production technology..." and "...there are now companies having the largest part of businesses in their sector with only running a software...". Lee et. al [7] also highlighted the key role of software in this new industry: "The Fourth Industrial Revolution is ubiquitous and will increasingly transform and reshape operations/production, supply-chain, management, and governance as well as products and services. Whatever could be codified of the organizational life will be put into codes and software and embedded into cybernetics systems that will replace human work activities". Yang [8], as well, described that "Industry 4.0 has two key factors: integration and

ISBN 978 -987-633-574-4

interoperability. Integrated with applications and software systems, Industry 4.0 will achieve seamless operations across organizational boundaries and will realize networked organizations". To a great extent, the success of Industry 4.0 rests on the quality of the software, which has the responsibility to ensure that it is error-free.

In 2020 we present the latest version of our testing framework called TAPIR [5], which was designed to detect failures in the sequence of method calls. It was suitable for Industries 4.0, but it was only available as a tool for Java developers. In this paper, we expand the scope of the framework by adding a new component: a web application that helps the software developer or other professionals to generate test cases to test their software element regardless of the programming language. The web application was implemented using TypeScript and React. Keeping the TAPIR philosophy, all the development presented in this work is free and open source.

The rest of the paper is structured as follows. Section 2 provides background information about the TAPIR framework. Then, we present the contributions of this article in Section 3. We later show how it was possible to find an error using the web application. Finally, we conclude with a brief discussion on the limitations and advantages of our approach and the future work.

2 TAPIR

TAPIR [5] is a testing framework for object-oriented source code based on Message Sequence Specification (MSS) [3] and implemented using Aspect-Oriented Programming (AOP) [4]. AOP allows the framework to create test cases that execute automatically with each execution of the program under test without modifying its source code. The use of MSS allows the developer to describe a regular expression for each class, which represents its correct behavior. The framework executes the program, and it checks whether the methods are invoked according to the regular expression in the class specification. The first thing the developer must do to use the framework is to create the regular expressions associated with the classes under test. These regular expressions must specify the correct behavior or invocation order of the classes' methods. To simplify the regular expression writing, symbols (i.e. characters) are used instead of the actual names of the methods. However, to be able to interpret it, the developer must specify a mapping between the actual methods' names and their corresponding symbol. The regular expressions and the maps between methods and symbols are set in the *TestingSetup.java* class.

The framework consists of two main components: an aspect, and a java class. The aspect is named *TestingCore.aj* and it contains the implementation of the framework's core. Listing 1.1 shows the implementation of the *TestingSetup.java* class that describes to TAPIR the correct behavior of example classes *CA* and *CB*. In this case, the correct usage of class *CA* states that, after the creation of the object, there should be a call to f followed by a call to g. After that, there can be as many calls as desired to either g or h. The final call of the sequence must

ISBN 978 -987-633-574-4

be to *h*. To correct use class *CB*, there should be first a call to *alpha* followed by a call to *gamma*, or a call to *gamma* followed by a call to *beta*. Afterward, any method between *alpha*, *beta*, or *gamma* can be called.

Listing 1.1: TAPIR configuration for classes *CA* and *CB*

```
//Class CA: Definition of the methods and their corresponding symbols
mapObjectsToCallSequence = new HashMap<Integer, String>();
mapMethodsToSymbols = new HashMap<String, String>();
mapMethodsToSymbols.put("main.CA.<init>", "c");
mapMethodsToSymbols.put("main.CA.f", "f");
mapMethodsToSymbols.put("main.CA.g", "g");
mapMethodsToSymbols.put("main.CA.h", "h");
//Definition of the regular expression
regularExpression = Pattern.compile("cfg(g|h)*h");
//Initializing the regular expressions controller
matcher = regularExpression.matcher("");
//A TestingInformation instance stores all information related to how the class is tested
TestingInformation ti = new TestingInformation(CA.class.toString(),
    mapObjectsToCallSequence, mapMethodsToSymbols, regularExpression, matcher,
    true);
TestingCore.mapClassToTestingInformation.put(CA.class.toString(), ti);
//Class CB: Definition of the methods and their corresponding symbols
mapObjectsToCallSequence = new HashMap<Integer, String>();
mapMethodsToSymbols = new HashMap<String, String>();
mapMethodsToSymbols.put("main.CB.alpha", "a");
mapMethodsToSymbols.put("main.CB.gamma", "g");
mapMethodsToSymbols.put("main.CB.beta", "b");
//Definition of the regular expression
regularExpression = Pattern.compile("(ag|gb)(a|g|b)*");
//Initializing the regular expressions controller
matcher = regularExpression.matcher("");
//A TestingInformation instance stores all information related to how the class is tested
ti = new TestingInformation(CB.class.toString(), mapObjectsToCallSequence,
    mapMethodsToSymbols, regularExpression, matcher, false);
TestingCore.mapClassToTestingInformation.put(CB.class.toString(), ti);
```

In Listing 1.3, we can see the framework output when the code portion of Listing 1.2 corresponding to the *CA* class is executed. In this case, the last call to *f* does not follow the MSS specified for the *CA* class. As mentioned above, when an error is detected, TAPIR informs by console the class and object that produced the error. The method that violated the MSS, the MSS and the actual sequence of calls are also shown in the console. Finally, the system aborts the execution, as indicated by the last parameter of method *TestingInformation* in the configuration.

Listing 1.2: Two snippets of code showing examples of wrong usage of class *CA* and class *CB*.

```
CA ca1 = new CA();
ca1.f();
```

ISBN 978 -987-633-574-4


```
ca1.g();
ca1.h();
ca1.f();
CB cb1 = new CB();
```

```
cb1.alpha();
cb1.alpha();
cb1.gamma();
cb1.gamma();
```

Listing 1.3: Error example for the CA class. The execution is aborted when the error is found.

```
---      ERROR FOUND    ---
Class: class main.CA
Object Code: 977993101
Method Executed: main.CA.f
Regular Expression: cfg(g|h)*h
Execution Sequence: cfghf
------ SYSTEM ABORTING... ------
```

Listing 1.4: Error example for the CB class. The execution is allowed to continue when the error is found.

```
---      ERROR FOUND    ---
Class: class main.CB
Object Code: 859417998
Method Executed: main.CB.alpha
Regular Expression: (ag|gb)(a|g|b)*
Execution Sequence: aa
-- CONTINUING EXECUTION... ---
```

```
---      ERROR FOUND    ---
Class: class main.CB
Object Code: 859417998
Method Executed: main.CB.gamma
Regular Expression: (ag|gb)(a|g|b)*
Execution Sequence: aag
-- CONTINUING EXECUTION... ---
```

Listing 1.4 shows the framework output when the code portion of Listing 1.2 corresponding to the class *CB* is executed. In this case, the second call to *alpha* does not follow the MSS specification for class *CB*. As configured in Listing 1.2 , the last parameter in the call to method *TestingInformation* is false, indicating that the execution must continue despite the existing errors. Therefore, Listing 1.4 shows multiple errors.

For a more in-depth analysis of the framework and more usage examples, we recommend that the reader see [5].

3 Proposal

As we previously mentioned, TAPIR is developed in Java and can only be used in Java applications. Although programming knowledge is necessary, the framework can be used by developers who do not have specific knowledge of the testing area. These two restrictions present two opportunities for improvement, and these are the contribution proposal for this work.

The proposal in this work is to expand the development carried out in [5] to include a new piece of software, a web application that completes the technique previously presented. While the framework functionality is oriented to evaluate

ISBN 978 -987-633-574-4

the correct usage of a set of running classes, the web application allows the generation of test cases to test more methodically the behavior of a class against possible combinations of calls of its methods. This is useful for testing software components that are not yet part of a complete application. In this way, the web application generates documentation oriented to the unit testing of such components. This documentation is of great help to the developer and, since the web application is very easy to use, any member of the work team can generate the test cases.

The definition, design, and implementation of the application and its interactions with the user were conceived under a User-Centered Design strategy [2] and considering the work by Signoretti et. al [10]. Under this strategy, the user of a system has active participation in its design and development. In our case, we worked with users who did not have a computer profile but were familiar with the software industry and Industry 4.0 in general. With them, the design of the graphical interface of the system, the use of labels throughout the application, and the interactions were validated. Interdisciplinary teams [11] are an increasingly common way of working in the context of Industries 4.0, so the tools that ensure the quality of software products should be usable by all team members. In this way, it would be possible to test a greater part of the software or test the same as before but in less time. On the other hand, making the testing tool independent from the language and even from the software and hardware platforms would allow the same tool to be applied in different projects, reducing or eliminating training times in new methodologies or programs.

3.1 Implementation

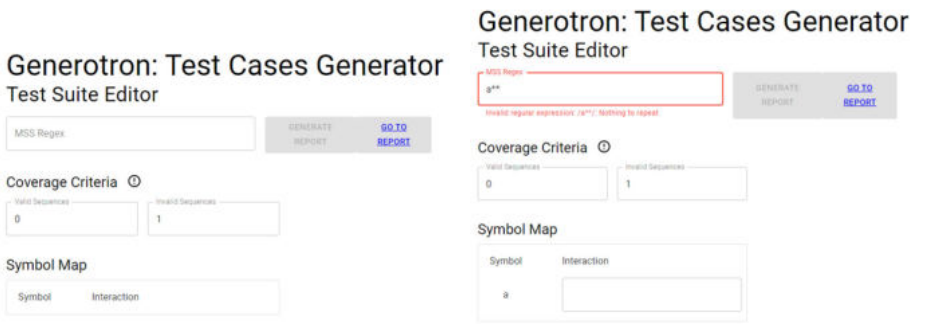
To avoid the MSS parsing logic being tied to the web application, the project was divided into two parts which are the MSS-Parser and the MSS-App modules. The MSS-Parser module validates that text strings are correct MSS and generates the test cases according to the provided coverage parameters. This module was not implemented from scratch but was a fork of the genex.js project repository, authored by Alix Axel. MSS-Parser was implemented entirely in TypeScript, to make it easier to use by providing a statically typed API. Like genex.js, MSS-parser uses the ret (Regular Expression Tokenizer) library to parse the regular expression associated with the MSS and return a tree of tokens. Then, this tree is traversed to generate the strings that represent test cases. The MSS-App was developed using React as the front-end framework, also in TypeScript, to be consistent with the MSS-Parser module and take advantage of static typing. The Material-UI web-component library was used because it offers a wide variety of components developed following the Material Design standards [1].

3.2 The front-end design

The front-end is divided into two parts. On the one hand, it offers an editor (Figure 1a) that allows entering the MSS, the coverage parameters, and an optional mapping between symbols and interaction names. Once the values are entered

ISBN 978 -987-633-574-4

and validated (Figure 1b), the application generates the test cases and enables the second part of the application. There the user can visualize the generated report (Figure 2a) and, for each test case the user can indicate if each step could be executed, and also if the test was successful or not. Each test case can contain a text note.



(a) Homescreen of Generotron, a web application for the generation of test cases based on MSS.

(b) Error found in the MSS input field. Every time the user types something in this field, it is checked

Fig. 1: Generotron: Test Case Generator

The application offers an extremely simple editor. Figure 1a shows a screenshot in its initial state. As seen in the figure, it is composed of three mandatory input fields in which the user enters the regular expression and the two values used as parameters for the coverage criteria. These last two are initialized by default with the values 0 and 1 since they are the minimum values allowed by definition. In turn, the input fields do not allow entering smaller values. Once a valid MSS expression has been entered, new fields are dynamically generated in which the user can enter the full name of the interaction, as shown in Figure 3.

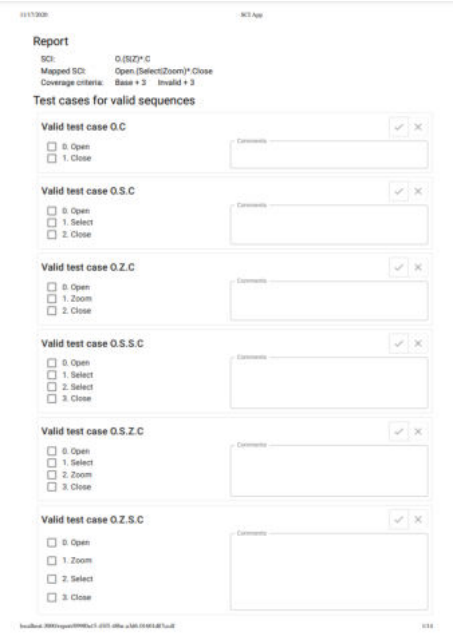
Symbol mapping is optional at the individual level; the user can add a more descriptive name for a symbol while omitting those where it is considered unnecessary. When viewing the report, the names added to the mapping are used to display more descriptive versions of the MSS expression and the list of methods to execute in each test case. It is important to note that although the mapping is optional, whenever abbreviated symbols are used, the ideal would be to provide one to improve the readability of the generated report.

In case an error is detected in the entered values, a message is displayed with a description (Figure 1b). The MSS-Parser module provides these error messages. Once the required values are entered, the Generate Report button is enabled. The web report was designed and implemented prioritizing simplicity over fanciness so that the same web format presented in the browser could be exported to a PDF using directly the universally provided printing functionality. At the top, it has a heading like the one in Figure 3, which shows the values

ISBN 978 -987-633-574-4

previously entered in the editor and used to generate the test cases in the report. The blue PDF icon is a button that allows exporting the document as a PDF file. Then the corresponding test cases are listed and grouped according to whether they are valid or invalid sequences of interactions. Each test case is contained by a box like the one shown in Figure 2a. The title includes the sequence of interactions from the MSS expression that compose the test case. As mentioned above, when a mapping was provided, the list of methods includes names instead of symbols as shown in Figure 6. The report includes a checkbox on each method of the test cases to indicate a successful execution. After conducting the test case, the user can register a successful or failed result in the upper right corner and write comments if necessary. A valid test case is successful if all the methods were successful. However, an invalid test case is successful if it fails at some point.

(a) A webform that shows the generated test cases. For each one of them, the user can indicate if each step could be executed, and also if the test was successful or not. Each test case can contain a text note.



(b) The webform with the test cases can be easily exported to PDF.

Fig. 2: Report generation in Generotron

ISBN 978 -987-633-574-4

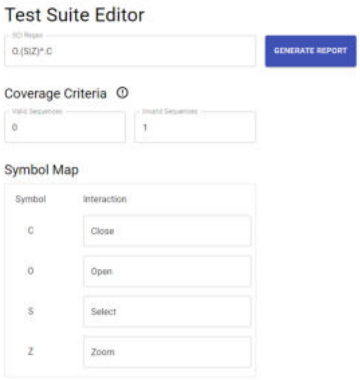


Fig. 3: For each symbol that is used in the MSS, it is possible to establish a mapping with the method it represents. This makes it easier to read the test cases.

4 Test Case. Rock.AR, a software solution for point counting

Point counting is the standard method to establish the modal proportion of minerals in coarse-grained igneous, metamorphic and sedimentary rock samples. This method requires taking observations at regular positions on the sample, namely grid intersections. Rock.AR [6] is an open-source visualization tool developed in Java that implements a semiautomatic point-counting method.

The implementation of Rock.AR includes a class called *CurrentTime* that provides the current time and is also used to measure the elapsed time between two moments. This class is part of a utility package that the application uses. There are only three methods available. *GetCurrentTime* returns the current time, conforming to format yyyy-MM-dd HH:mm:ss. *StartTimeFrame* is used to mark the beginning of a time frame, and *EndTimeFrame* marks the end of such time frame and returns the elapsed time between start and end.

The correct use of this class is described as follows: *GetCurrentTime* can be called at any time, and a time interval can be measured by first calling *StartTimeFrame* and then *EndTimeFrame*. Between a call to *StartTimeFrame* and *EndTimeFrame*, calls to *getCurrentTime* can occur. Using these symbols for each method; g for *GetCurrentTime*, s for *StartTimeFrame*, and e for *EndTimeFrame*, the following MSS describes the correct operation of the class:

$$((s^* \bullet g^* \bullet e)^* | g)^* \tag{1}$$

Using the Web Application to generate test cases for the *CurrentTime* class (see Figure 4), it was possible to detect a bug. The application generated combinations of calls to the methods of the class that caused the class to behave

ISBN 978 -987-633-574-4

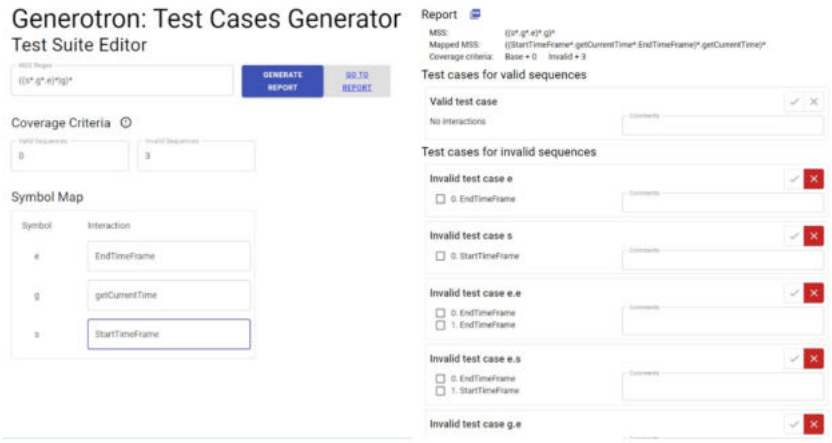


Fig. 4: Test cases generated for the *GetCurrentTime* class

incorrectly when tested. Although these sequences should not occur during normal execution, the class should be robust enough to withstand misuse, which is not. Note that a valid test case is successful if the sequence executes correctly; however, an invalid test case is successful if the sequence fails to execute.

5 Conclusions & Future Work

Industry 4.0 requires new methodologies to ensure the quality of its software, a key element in its production chain. A framework for testing Object-Oriented Software was developed for testing Java applications but was only available for this programming language and could only be used by someone with programming knowledge. We expanded this framework by introducing a web application that complements it. Although the framework can be used in any Java implementation without modifying the source code, it requires Java and programming knowledge. The web application is suitable for any implementation based on Object-Oriented Programming, regardless of the programming language and it can be used by anyone in the work team. All these tools were designed and implemented to detect, without modifying the source code, failures in the sequence of calls that objects make. As shown in the case studies, these tools help with the detection of errors that would otherwise be difficult to find. The framework is available for downloading at `http://cs.uns.edu.ar/~mll/lapaz/` and the web application can be used at `https://cs.uns.edu.ar/~dku/mss`. The source code is available and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

ISBN 978 -987-633-574-4

Acknowledgment

This work was partially supported by the following research projects: PGI 24/N050 and PGI 24/ZN35 from the Secretaría General de Ciencia y Tecnología, Universidad Nacional del Sur, Argentina.

References

1. Ian G Clifton. *Android user interface design: Implementing material design for developers*. Addison-Wesley Professional, 2015.
2. Roger Coleman, John Clarkson, and Julia Cassim. *Design for inclusivity: A practical guide to accessible, innovative and user-centred design*. CRC Press, 2016.
3. Shekhar Kirani and W. T. Tsai. Specification and verification of object-oriented programs. Technical report, Computer Science Department, University of Minnesota, 1994.
4. Ramnivas Laddad. *AspectJ in Action: Practical Aspect-Oriented Programming*. Manning Publications Co., Greenwich, CT, USA, 2003.
5. Martin Larrea and Dana Urribarri. Increasing confidence in industry 4.0 through new software verification and validation techniques. In *International Conference on Production Research ICPR Americas*, page In press. International Conference of Production Research, 2020.
6. Martín L Larrea, Silvia M Castro, and Ernesto A Bjerg. A software solution for point counting. petrographic thin section analysis as a case study. *Arabian Journal of Geosciences*, 7(8):2981–2989, 2014. doi:10.1007/s12517-013-1032-0.
7. MinHwa Lee, JinHyo Joseph Yun, Andreas Pyka, DongKyu Won, Fumio Kodama, Giovanni Schiuma, HangSik Park, Jeonghwan Jeon, KyungBae Park, KwangHo Jung, et al. How to respond to the fourth industrial revolution, or the second information technology revolution? dynamic new combinations between technology, market, and society through open innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, 4(3):21, 2018.
8. Yang Lu. Industry 4.0: A survey on technologies, applications and open research issues. *Journal of industrial information integration*, 6:1–10, 2017.
9. Ercan Oztemel and Samet Gursev. Literature review of industry 4.0 and related technologies. *Journal of Intelligent Manufacturing*, 31(1):127–182, 2020.
10. Ingrid Signoretti, Larissa Salerno, Sabrina Marczak, and Ricardo Bastos. Combining user-centered design and lean startup with agile software development: a case study of two agile teams. In *International Conference on Agile Software Development*, pages 39–55. Springer, 2020.
11. Alp Ustundag and Emre Cevikcan. *Industry 4.0: managing the digital transformation*. Springer, 2017.

ISBN 978 -987-633-574-4

Construcción de grafos de conocimiento a partir de especificaciones de requerimientos usando procesamiento de lenguaje natural

Luciana Tanevitch[1], Felipe Dioguardi[1], Juliana Delle Ville[1], Sebastián Villena[1], Francisco Herrera[1], Waldo Hasperue[2,4], Diego Torres[1,2,3], and Leandro Antonelli[1]

[1] LIFIA, Facultad de Informática, UNLP
[2] Comisión de Investigaciones Científicas (CICPBA)
[3] Departamento de Ciencia y Tecnología, UNQ
[4] Instituto de Investigación en Informática LIDI, Facultad de Informática, UNLP
luciana.tanevitch@lifia.info.unlp.edu.ar

Abstract. Los requerimientos cumplen un rol fundamental en el desarrollo de los sistemas informáticos, ya que un software bien diseñado y codificado no aporta ningún valor si no satisface los requerimientos. Relevar y especificar requerimientos no es una tarea fácil. La principal fuente de información son las personas y las técnicas basadas en lenguaje natural. Sin embargo, el lenguaje natural presenta muchas debilidades, por lo cual lograr una especificación de calidad es un gran desafío y esfuerzo. Así pues, se necesitan brindar herramientas para poder sintetizar las especificaciones de forma de poder identificar por ejemplo omisiones, ambigüedades y conflictos. Este artículo presenta una propuesta para construir un grafo de conocimiento a partir de una especificación en lenguaje natural. Se presenta tanto un proceso como una herramienta que automatiza el proceso. De esta forma, el grafo de conocimiento obtenido ofrece al analista una síntesis de los conceptos y las relaciones entre ellos.

Keywords: Grafo de conocimiento, Especificaciones de requerimientos, Lenguaje natural

1 Introducción

Un requerimiento [4] es la condición o capacidad que debe poseer un sistema o uno de sus componentes para satisfacer un contrato, un estándar, una especificación u otro documento formalmente impuesto. En otras palabras, un requerimiento describe las necesidades, deseos y expectativas [12] de los *stakeholders* (por ejemplo usuarios, o expertos del dominio) con el objetivo de lograr el producto adecuado para ellos [15]. De esta forma, los requerimientos cumplen un rol fundamental en el desarrollo de software, ya que los errores de la especificación de requerimientos se van a trasladar en los siguientes productos del ciclo de vida de desarrollo como por ejemplo diseño, código y casos de prueba. Un error en los requerimientos detectado cuando el software es entregado al usuario

ISBN 978 -987-633-574-4

implica volver a trabajar no solo de los requerimientos, sino también del diseño, codificación y testeo. Es así que esforzarse en lograr una buena especificación de requerimientos ayuda a reducir el volver a trabajar posteriormente. Es por eso que se recomienda que la especificación cumpla ciertos atributos de calidad como los siguientes: completa, concisa, coherente, consistente, no ambiguo y verificable entre otras [4].

De acuerdo a Loucopoulos et al. [12] la especificación de requerimientos es el producto final de un proceso con varias etapas: elicitación, modelización y validación. La elicitación es la etapa en la cual se obtiene el conocimiento necesario. Existen diferentes técnicas, sin embargo, lo más común es obtener el conocimiento de las personas. Esto se puede realizar a través de personas directamente utilizando entrevistas individuales o grupales, como así también en forma indirecta a través de cuestionarios o encuestas [15]. Para llevar a cabo este relevamiento es necesario que el equipo de desarrollo (los analistas) puedan comunicarse con los *stakeholders*. El medio de comunicación más utilizado es el lenguaje natural [3] ya que evita que los *stakeholders* aprendan formalismos que no les resultan naturales. De todas formas, el lenguaje natural presenta debilidades intrínsecas como ambigüedad, redundancia, incompletitud e inconsistencia, las que pueden ocasionar malos entendidos, que a su vez pueden ocasionar errores de especificación. Sea por ejemplo la frase "el pez esta listo para comer", la misma admite dos interpretaciones. Por un lado, se puede interpretar que el pez está listo para ser comido. Y por otro lado, se puede interpretar que el pez está listo para recibir su alimento. La diferencia entre ambas interpretaciones radica en si el pez es el sujeto que realiza la acción de comer, o en cambio una persona es quien come al pez (la persona es el sujeto mientras que el pez es el objeto).

Un grafo de conocimiento es un grafo que posee como propósito acumular y transmitir conocimiento sobre el mundo real, en el cual sus nodos representan entidades de interés y sus aristas relaciones entre esas entidades[10]. Para el ejemplo anterior, tanto "pez" como "persona" podrían ser nodos del grafo, mientras "comer" sería la relación entre ambos. De esta forma, quedaría muy claro que "la persona come al pez" y esto seria muy diferente de "el pez come un alga". Además de expresar el conocimiento de una forma más ordenada, los grafos de conocimiento le otorgan una semántica a cada nodo y cada relación. Es así que el significado sería preciso sin ocasionar ambigüedades ya que esta representación de conocimiento se basa en ontologías, que son formalizaciones que definen de forma estricta y no ambigua los conceptos que pretenden manejar [17]. Finalmente, un grafo de conocimiento brinda una especificación cuya interpretación puede automatizarse, es decir, se podrían utilizar algoritmos para inferir consecuencias a partir del grafo. Mas aún, los grafos de conocimiento permiten la implementación de técnicas para la resolución de sinónimos, búsquedas multi-idioma y resolución de ambigüedades. El objetivo de este artículo es presentar un proceso para procesar especificaciones en lenguaje natural y obtener un grafo de conocimiento. El artículo presenta una herramienta que da soporte al proceso.

ISBN 978 -987-633-574-4

El resto del artículo se organiza de la siguiente manera. La sección 2 analiza trabajos relacionados. La sección 3 describe grafos de conocimiento. La sección 4 presenta el proceso propuesto. Finalmente, la sección 5 discute conclusiones y trabajos futuros.

2 Trabajos relacionados

Delugach et al. [5] proponen un proceso para construir un grafo de conocimiento, sin embargo, el proceso es completamente manual y se basa en la interacción con los stakeholders. De todas formas, existen varios trabajos que proponen procesos automáticos o semi automáticos similares al nuestro. Yang et al. [23] proponen una herramienta para construir grafos de conocimiento para el lenguaje chino. Song et al. [19] también trabajan con el lenguaje chino y dadas las particularidades del mismo, ellos proponen una estrategia para construir mapas de conocimiento específicos de dominio a través de técnicas de deep learning. Qin et al. [16], quienes también trabajan con el lenguaje chino, proponen un método para construir un grafo de conocimiento a partir del lenguaje natural, pero específicamente para predecir enfermedades crónicas. Schlutter et al. [18] trabajan con oraciones simples para enriquecer un grafo de conocimiento y trabajan solo con el idioma inglés. Por su parte, Verma et al. proponen una técnica similar, sin embargo, ellos construyen mapas de conocimiento [22] o incluso generan ontologías luego de revisar las especificaciones y sugerir mejoras de redacción [21]. Otros trabajos utilizan los grafos de conocimiento como una herramienta para lograr otro fin. Hassan et al. [8] proponen una técnica para analizar requerimientos a partir de la construcción de grafos de requerimientos y finalmente lo convierten a un ontología. Ferrari et al. [6] sostienen la importancia del contexto para la detección de ambigüedades y para ello utilizan un grafo de conocimiento para representar el dominio y poder analizarlo. Mills et al. [13] además de utilizar lenguaje natura y grafos de conocimiento, ellos utilizan redes de Petri para enriquecer al análisis. Por otro lado, ellos convierten las oraciones en lenguaje natural, a una estructura "Agente, Acción, Paciente" lo cual es similar a nuestra estructura.

3 Grafos de conocimiento

En la actualidad, es necesario contar con estructuras que admitan el manejo de grandes volúmenes de datos, que puedan enriquecerse fácilmente, y que permitan, a través de su análisis, descubrir e inferir conocimiento implícito, más allá de las representaciones explicitadas en el esquema. Una manera de abordar esos requisitos es usando grafos de conocimiento [14]. Se puede visualizar a un grafo de conocimiento como una estructura de nodos que representan entidades, conectados a través de aristas que son las relaciones. Los grafos de conocimiento pueden, además, poser una definición del conocimiento con mayor nivel de formalismo a través de ontologías[10, 7]. Las ontologías son representaciones formales del conocimiento a través de taxonomías, que conceptualizan un dominio

ISBN 978 -987-633-574-4

determinado [9]. Su uso permite descubrir qué es una entidad, cómo deberían ser categorizadas, qué propiedades deberían tener, además de permitir realizar inferencias tales como "un perro es un animal, un animal es un ser vivo, los seres vivos tienen un tipo de alimentación, y por lo tanto un perro debe tener un tipo de alimentación". Para este tipo de descubrimiento, se requiere definir un conjunto de reglas a aplicar. Al introducir estos modelos semánticos, podemos ver que la semántica otorga significado a los datos, permite generar un contexto, permite manejar ambigüedades, lo que a su vez hace que la información en grafos de conocimiento pueda aprovecharse mejor, enriquecerse a lo largo del tiempo, y en consecuencia mejorar las consultas y análisis aplicables a ellos.

Al describir requerimientos necesitamos representar objetos del mundo real, no simples cadenas de texto que nombren un concepto, y para esto las propiedades de datos y de objetos ayudan a enriquecer la representación de un concepto. Podemos denotar a un grafo de conocimiento como un dataset de tripletas no ordenadas con la forma **entidad – relación – entidad**. Por ejemplo, para relaciones de elementos correspondientes a la administración de una provincia, podría existir la tripleta "La Plata – ubicada en – Buenos Aires", que determina que el nodo "La Plata" se relaciona con el nodo "Buenos Aires" a través de la relación "ubicada en". Gráficamente, las relaciones se pueden ver en un grafo (lo que le da el nombre a la estructura) como se muestra en la figura 1.

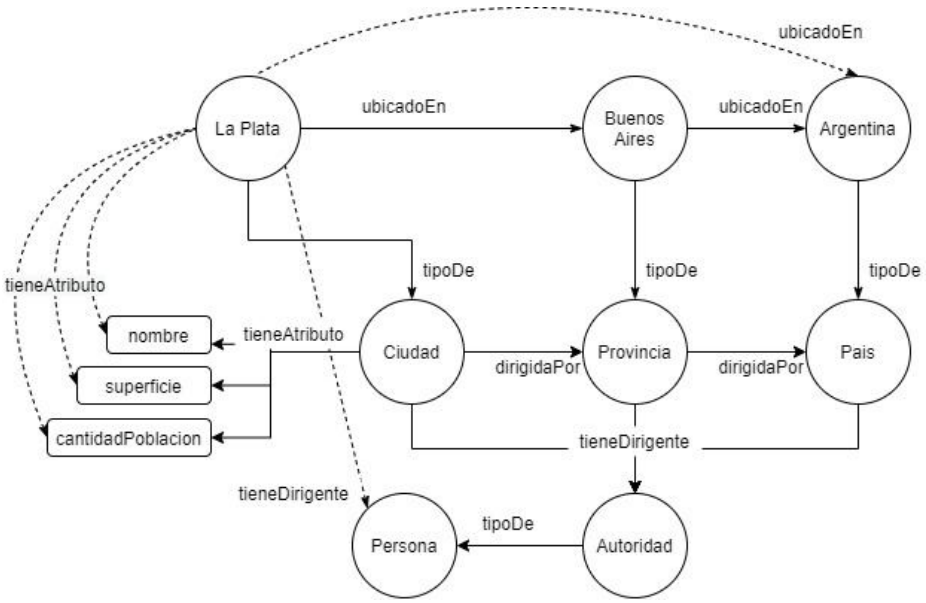


Fig. 1. Ejemplo grafo de conocimiento

Se pueden obtener ciertas conclusiones en base al modelo, tales como:

ISBN 978 -987-633-574-4

- La ciudad de La Plata está ubicada en la provincia de Buenos Aires.
- La provincia Buenos Aires está ubicada en el país Argentina.

Por lo tanto, si "ubicadoEn" es una propiedad transitiva, podemos inferir que la ciudad de La Plata está ubicada en el país Argentina.

- Una ciudad tiene una persona que es su autoridad.
- La Plata es una ciudad.

Entonces, la ciudad de La Plata tiene una persona que es su autoridad.

4 Proceso propuesto

Se plantea un proceso que recibe como entrada una especificación en lenguaje natural y produce como salida un grafo de conocimiento que sintetiza su información. Se propone la utilización de *kernel sentences* [11], frases simples creadas a partir de oraciones escritas en lenguaje natural, que mantienen la semántica original, y que siguen ciertas reglas sintácticas predefinidas. Determinar estas pautas no es un desafío menor, se deben producir expresiones fácilmente analizables por máquinas sin perder información relevante en el proceso. El proceso que se plantea en este trabajo consta de 4 etapas. En la primer etapa se separa el texto en oraciones. En la segunda etapa se reemplazan ciertas expresiones (pronombres personales, adjetivos demostrativos, etc.). En la tercera etapa se construyen kernel sentences que respetan una estructura simple del estilo "sujeto - verbo - objeto". La cuarta etapa identifica conceptos, propiedades y relaciones que usan para construir el grafo de conocimiento. La Figura 2 resume el proceso.

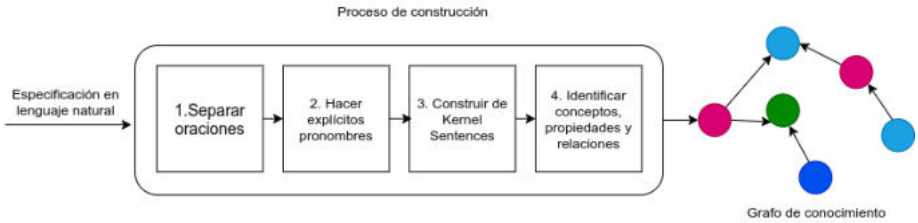


Fig. 2. Proceso de transformación

El resto de la sección amplía cada una de las etapas enumeradas. Para ello se utiliza la siguiente especificación.

Caso de estudio: *Sea una empresa que ofrece travesías en kayak. La misma ofrece distintas travesías que tienen diferente duración e itinerario. La empresa admite kayakistas inexpertos como también de experiencia. Cualquier persona le puede solicitar travesías a la empresa y la misma le informa el arancel.*

ISBN 978 -987-633-574-4

4.1 Separación de oraciones

El primer paso consiste en aplicar reglas que permiten delimitar dónde comienza una oración y dónde termina. La forma trivial es separarlas por signos de puntuación, por lo que, notando que la especificación contiene cuatro puntos, un primer procesamiento daría como resultado:

1. Sea una empresa que ofrece travesías en kayak.
2. La misma ofrece distintas travesías que tienen diferente duración e itinerario.
3. La empresa admite kayakistas inexpertos como también de experiencia.
4. Cualquier persona le puede solicitar travesías a la empresa y la misma le informa el arancel.

En algunas circunstancias, el punto puede ser utilizado para abreviaturas (*Sr.*, *Ud.*). Al igual que el signo de exclamación puede ser usado para interjecciones, y no para marcar una oración con las características que se analizaron (*¡Ay!*, *¡Viva!*). Se cuenta con herramientas de software que permiten reconocer estas situaciones, y así reducir la posibilidad de obtener expresiones que no cumplen con la estructura de una oración.

4.2 Explicitación de pronombres

A menudo, el lenguaje implica omisiones para evitar la repetición, el sujeto tácito es un ejemplo de esto. El sujeto tácito suele referir al núcleo de la expresión anterior, por lo que podría establecerse como regla de reemplazo de modo que el texto previo se reescriba de la siguiente manera:

1. Sea una empresa que ofrece travesías en kayak.
2. La empresa ofrece distintas travesías.
3. Las travesías tienen diferente duración e itinerario.
4. La empresa admite kayakistas inexpertos como también de experiencia.
5. Cualquier persona le puede solicitar travesías a la empresa y la empresa informa el arancel a la persona.

4.3 Construcción de kernel sentences

Las conjunciones agregan a las expresiones complejidad no deseada, por lo que se propone separar una oración que posea una conjunción en dos oraciones distintas. Se tuvieron en cuenta dos posibles escenarios: que la conjunción aparezca en el sujeto, o que aparezca en el objeto. En cualquiera de los casos se deben analizar las dependencias entre palabras para evitar la pérdida de información relevante al hacer el desglose. Por ejemplo, si hubiera modificadores del núcleo del objeto o del sujeto, podría nombrarse tácitamente luego de la conjunción, y si este fuera el caso debe tenerse en cuenta conservar el núcleo en las oraciones que se generen.

La especificación analizada tomaría la siguiente forma:

1. Sea una empresa que ofrece travesías en kayak.

ISBN 978 -987-633-574-4

2. La empresa ofrece distintas travesías.
3. Las travesías tienen diferente duración.
4. Las travesías tienen diferente itinerario.
5. La empresa admite kayakistas inexpertos
6. La empresa admite kayakistas de experiencia.
7. La persona puede solicitar travesías a la empresa.
8. La empresa informa el arancel a la persona.

En la frase "La empresa admite kayakistas inexpertos como también de experiencia.", la palabra *kayakista*, el núcleo del objeto, debe aparecer en las oraciones extraídas. De no ser así, podría ser complejo inferir qué es lo que la empresa admite "de experiencia".

La oración "Cualquier persona le puede solicitar travesías a la empresa y la empresa informa el arancel a la persona.", se puede separar en la conjunción y obtener dos sentencias que cumplen el requisito "sujeto - verbo - objeto".

4.4 Identificación de conceptos, propiedades y relaciones

Para lograr identificar las relaciones de dependencia se utilizan herramientas de procesamiento de lenguaje natural que permiten asignar etiquetas a cada una de las palabras para que las máquinas puedan reconocerlas y analizarlas. Partiendo de una estructura sintáctica del tipo "sujeto - verbo - objeto" se pueden aplicar nuevas reglas que permitan reconocer entidades, y las relaciones entre ellas. Entonces, podemos ver al núcleo del sujeto como el ejecutor de una acción, quien será identificado como una entidad. El objeto directo será reconocido como una entidad en caso de que el objeto indirecto no esté presente. Caso contrario el objeto indirecto, el receptor de una acción, será una entidad. El verbo "tener" indica una relación de posesión o pertenencia por parte de la entidad que refiere el sujeto; este posibilita reconocer como propiedades de una entidad aquellas expresiones involucradas con el verbo (regularmente, el objeto directo). Las relaciones pueden ser un verbo, una frase verbal, o la combinación del verbo con el objeto directo (en caso que esté presente el objeto indirecto). Para permitir comprender mejor este proceso, se describirá en la siguiente sección una herramienta que lo lleva a cabo.

5 Herramienta implementada

Con el fin de dar soporte automatizado al proceso propuesto, se implementó una herramienta Web en Python que utiliza la librería Spacy [20] como soporte al procesamiento de lenguaje natural[5]. En la figura 3 se puede ver una captura de la página principal de la herramienta. A continuación se describe el funcionamiento de esta herramienta.

[5] El código se encuentra disponible en `https://github.com/cientopolis/requirements-knowledge-graph`

ISBN 978 -987-633-574-4

Fig. 3. Herramienta, home page

En primera instancia, toma el texto ingresado y utiliza Spacy para obtener el rol sintáctico de cada palabra. A continuación, se aplican las reglas de transformación propuestas en la sección 4 para obtener las kernel sentences equivalentes a las oraciones originales.

A partir de estas expresiones, la herramienta identifica entidades, propiedades y relaciones siguiendo lo definido en la sección 4.4, y construye un grafo de conocimiento que almacena en formato RDF [1]. Luego permite realizar gráficos representativos y consultas mediante el lenguaje SPARQL [2].

Para comprender cómo trabaja la herramienta, se analizará su funcionamiento dadas siguientes kernel sentences:

1. La empresa ofrece travesías en kayak.
2. Las travesías en kayak tienen arancel.
3. Los kayakistas contratan travesías en kayak.
4. La empresa informa el arancel a los kayakistas.

Oración 1: se identifica *empresa* como entidad por ser el núcleo del sujeto, y *travesías en kayak* por objeto directo, en ausencia del objeto indirecto. La empresa realiza la acción de *ofrecer*, que sirve como vínculo con la entidad *travesías en kayak*. Es interesante notar que se selecciona *travesías en kayak*, siendo *en kayak* un modificador del sustantivo, porque agrega precisión y contexto. Ignorarlo produciría un resultado demasiado genérico, que es lo que se busca evitar.

Oración 2: se identifica la entidad *travesías en kayak*, y su atributo *arancel*, siguiendo las reglas propuestas en el proceso que indican reconocer como propiedades aquellas expresiones que se asocien al verbo "tener".

Oración 3: se identifica la entidad *kayakistas*, que realiza la acción *contratar* que la vincula con la entidad *travesías en kayak*.

Oración 4: se identifica la entidad *empresa* por ser núcleo del sujeto, y como hay presencia de objeto directo, también lo será *kayakistas*, el núcleo del objeto indirecto. Entonces, la relación será *informa arancel*, siendo la combinación del verbo con el objeto directo.

La figura 4 muestra el resultado del análisis realizado. Allí puede verse que se detectan las entidades y las acciones que las vinculan. Además se detectan algunas características que pueden describirse en términos de programación orientada a objetos como es el caso de Empresa como una clase candidata.

ISBN 978 -987-633-574-4

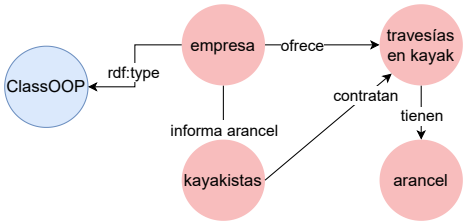


Fig. 4. Herramienta de soporte del proceso de transformación

6 Conclusiones

Este articulo presentó un proceso que tiene como finalidad la construcción de un grafo de conocimiento a partir de una especificación en lenguaje natural. El proceso cuenta con una herramienta que le da soporte a cada una de las cuatro etapas que lo conforman. El proceso propuesto permite realizar una síntesis de las conceptos y las relaciones a través de un grafo, de forma que el analista pueda identificar los aspectos que necesitan ser mejorados en la especificación. La herramienta presentada tiene como finalidad el tratamiento del lenguaje español y se basa en una librería de terceros, pero tuvo que ser enriquecida para ciertas particularidades del lenguaje español, que dada lo rico que es y las distintas variantes para expresar las ideas, lo convierten en un lenguaje muy complejo de analizar. Nuestros próximos trabajos tienen varias arista. Por un lado mejorar los aspectos de procesamiento del lenguaje español. Por otro lado, queremos vincular el grafo de conocimiento producido por nuestro enfoque con otros grafos de conocimientos ya elaborados, de forma de poder aprovechar el conocimiento en los otros grafos. Finalmente, estamos trabajando en analizar el grafo de conocimiento, para identificar automáticamente debilidades y poder sugerir al analista que aspectos de la especificación necesitan ser mejorados. También se está trabajando en detectar a partir de las especificaciones más elementos del paradigma orientada a objetos como posibles métodos y atributos.

References

1. RDF. `https://www.w3.org/RDF/`, accedido: 2020-08-01
2. SPARQL. `https://www.w3.org/TR/sparql11-query/`, accedido: 2020-08-01
3. Alzayed, A., Al-Hunaiyyan, A.: A bird's eye view of natural language processing and requirements engineering
4. Committee, I.C.S.S.E.S., Board, I.S.S.: Ieee recommended practice for software requirements specifications, vol. 830. IEEE (1998)
5. Delugach, H., Lampkin, B.: Acquiring software requirements as conceptual graphs. In: Proceedings Fifth IEEE International Symposium on Requirements Engineering. pp. 296–297 (2001)
6. Ferrari, A., Gnesi, S.: Using collective intelligence to detect pragmatic ambiguities. In: 2012 20th IEEE International Requirements Engineering Conference (RE). pp. 191–200. IEEE (2012)

ISBN 978 -987-633-574-4

7. Guarino, N., Oberle, D., Staab, S.: What Is an Ontology?, pp. 1–17. Springer Berlin Heidelberg, Berlin, Heidelberg (2009), https://doi.org/10.1007/978-3-540-92673-3_0
8. Hassan, T., Hassan, S., Yar, M.A., Younas, W.: Semantic analysis of natural language software requirement. In: 2016 Sixth International Conference on Innovative Computing Technology (INTECH). pp. 459–463 (2016)
9. Hogan, A.: The Web of Data. Springer (2020), https://doi.org/10.1007/978-3-030-51580-5
10. Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G.D., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., Ngomo, A.C.N., Polleres, A., Rashid, S.M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., Zimmermann, A.: Knowledge graphs. ACM Comput. Surv. 54(4) (Jul 2021), https://doi.org/10.1145/3447772
11. Katz, B.: From sentence processing to information access on the world wide web
12. Loucopoulos, P., Karakostas, V.: System Requirements Engineering. McGraw-Hill, Inc., USA (1995)
13. Mills, M., Psarologou, A., Bourbakis, N.: Modeling natural language sentences into spn graphs. In: 2013 IEEE 25th International Conference on Tools with Artificial Intelligence. pp. 889–896 (2013)
14. Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., Taylor, J.: Industry-scale knowledge graphs: Lessons and challenges. Commun. ACM 62(8), 36–43 (Jul 2019), https://doi.org/10.1145/3331166
15. Pohl, K.: Requirements Engineering Fundamentals, 2nd Edition: A Study Guide for the Certified Professional for Requirements Engineering Exam - Foundation Level - IREB compliant. Rocky Nook-Ips (2016), https://books.google.com.ar/books?id=1VsUDgAAQBAJ
16. Qin, S., Xu, C., Zhang, F., Jiang, T., Ge, W., Li, J.: Research on application of chinese natural language processing in constructing knowledge graph of chronic diseases. In: 2021 International Conference on Communications, Information System and Computer Engineering (CISCE). pp. 271–274 (2021)
17. Saorín, T.: Grafos de conocimiento y bases de datos en grafo: conceptos fundamentales a partir de una" obra maestra" del museo del prado. Anuario Think EPI 13 (2019)
18. Schlutter, A., Vogelsang, A.: Knowledge representation of requirements documents using natural language processing (2018)
19. Song, Y., Rao, R.N., Shi, J.: Relation classification in knowledge graph based on natural language text. In: 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). pp. 1104–1107 (2018)
20. Vasiliev, Y.: Natural Language Processing with Python and SpaCy: A Practical Introduction. No Starch Press (2020)
21. Verma, K., Kass, A.: Requirements analysis tool: A tool for automatically analyzing software requirements documents. In: International semantic web conference. pp. 751–763. Springer (2008)
22. Verma, R.P., Beg, M.R.: Representation of knowledge from software requirements expressed in natural language. In: 2013 6th International Conference on Emerging Trends in Engineering and Technology. pp. 154–158 (2013)
23. Yang, X., Zhao, S., Cheng, B., Wang, X., Ao, J., Li, Z., Cao, Z.: A general solution and practice for automatically constructing domain knowledge graph. In: 2020 IEEE 6th International Conference on Computer and Communications (ICCC). pp. 1675–1681 (2020)

ISBN 978 -987-633-574-4

Ingeniería de Requisitos para Organizaciones Enfocadas en los Procesos

Gladys Kaplan[1], Juan Pablo Mighetti[1], Gabriel Blanco[1]

[1]Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de La Matanza. Florencio Varela 1903, (B1754JEC) San Justo, Buenos Aires
{gkaplan,jpmighetti,gblanco}@unlam.edu.ar

Resumen. Las organizaciones determinan su desempeño a través de una cadena de procesos. Contar con información tangible y concreta de cada proceso es una forma de evitar fallos. Así, muchas organizaciones están comenzando a enfocarse en sus procesos dejando la tradicional verticalidad. Esta información se encuentra ordenada y fuertemente alineada a la organización. En el presente artículo se describe un mecanismo para utilizar este conocimiento organizacional en la Ingeniería de Requisitos, específicamente en el Proceso de Requisitos basado en Escenarios. De esta manera se reduce significativamente el costo de construcción de los requisitos en aquellas estrategias que analizan el contexto actual antes de especificar los servicios del nuevo sistema de software. También se ha detectado que la relación entre los modelos de procesos y los modelos de requisitos permiten una validación cruzada, ya que la omisión o inconsistencias en un modelo es rápidamente detectada al construir el otro.

Palabras Clave: ingeniería de requisitos, enfoque por procesos, estrategias de construcción.

1 Introducción

Durante el siglo XX las organizaciones se caracterizaron por ser piramidales y jerárquicas, mientras que las del siglo XXI tienden a ser organizaciones más horizontales [1]. La verticalidad supone una gran influencia de la jerarquía en el funcionamiento de la organización y, por el contrario, la horizontalidad presupone la capacidad de que todas las personas de la organización puedan planificar, organizar, dirigir sus actividades y a la vez autosupervisarse. Las organizaciones verticales se centran en aspectos estructurales, por ejemplo, cual es la cadena de mando o definir la función de cada departamento mientras que las horizontales se orientan a desarrollar la misión de la organización, mediante la satisfacción de las expectativas de los involucrados (clientes, proveedores, accionistas, etc.). Estas organizaciones horizontales consideran que toda la organización se puede concebir como una red de procesos interrelacionados o interconectados [2] y su gestión está basada por concepción [3]. Empresas líderes han aplicado este cambio organizativo,

ISBN 978 -987-633-574-4

individualizando sus procesos, eligiendo los procesos relevantes, analizándolos y mejorándolos y finalmente utilizando este enfoque para transformar sus organizaciones. Luego de los buenos resultados logrados, aplicaron la experiencia obtenida para optimizar el resto de sus procesos en toda la organización. Entre las empresas de IT más conocidas están Amazon, Spotify, Google, Deloitte, etc.

Algunas organizaciones definen los procesos con la intención de mejorar su funcionamiento individual, con lo cual su estructura sigue siendo por silos, departamental y jerárquica, sin compromiso hacia una transformación integral. Aún entre las organizaciones que están cambiando su estructura, existen diferentes niveles de madurez en el enfoque por procesos, o sea, manejan diferentes niveles de información o de gestión acerca de sus procesos. Muchas de las organizaciones terminan abandonando sus planes de transformación debido a la complejidad de un cambio tan radical para el cual no están preparadas. Por el contrario, existen otras organizaciones que logran ver el valor agregado de tener una gestión integral por procesos e implementan BPM (Business Process Management) [4] [5] para realizar el modelado y la automatización correspondiente.

En este marco las organizaciones definen sus propios modelos de procesos con información validada y consensuada, la cual puede ser absorbida por la IR para mejorar sus propios procesos y alinearse mejor con la organización. Existen tres formas de abordar estos contextos: desde el punto de vista de la IR, desde la propuesta organizacional o en un mix de ambas. En el primer caso, utilizar el punto de vista de la IR, la información de procesos de la organización se utiliza como una fuente de información más. En el segundo caso, punto de vista de la organización, el cambio para la IR es más radical, ya que debe realizar todo el proceso de requisitos utilizando la representación organizacional. En este caso se debe asegurar que los modelos utilizados han sido probados para corroborar su efectividad para contener los requisitos del software. En la tercera opción, ambos puntos de vista colaboran aportando un análisis diferente de la información del contexto para llegar a los requisitos del software. Es en esta opción donde se centra el presente artículo, donde se utilizan los mapas de procesos dentro del Proceso de Requisitos basado en Escenarios [6]. De esta manera se aprovecha todo el conocimiento existente para enriquecer y asegurar la calidad de los requisitos del software.

En la sección 2 se describe qué es un enfoque por procesos. En la sección 3 se presenta el proceso de requisitos basado en escenarios, en el cual se basa la presente propuesta. En la sección 4 se detallan las estrategias para realizar la IR en organizaciones con un enfoque inicial por procesos y se describe un pequeño ejemplo. Finalmente, en la sección 5 se presentan las conclusiones y trabajos futuros.

ISBN 978 -987-633-574-4

2 Enfoque por Procesos

El cumplimiento de los objetivos de una organización, está directamente vinculado a la correcta creación y ejecución de sus procesos, como así también a su correcta actualización y ampliación de los mismos [3]. Según la ISO 9001 la definición de procesos en la organización, proporciona múltiples beneficios potenciales:

- Aumento de la capacidad de centrar los esfuerzos en los procesos clave y en las oportunidades de mejora.
- Resultados coherentes y previsibles mediante un sistema de procesos alineados.
- Optimización del desempeño mediante la gestión eficaz del proceso, el uso eficiente de los recursos y la reducción de las barreras interdisciplinarias.
- Posibilidad de que la organización proporcione confianza a las partes interesadas en lo relativo a su coherencia, eficacia y eficiencia

Para comprender los procesos se genera un **mapa de procesos,** el cual es un diagrama que permite ver de forma gráfica todos los procesos que se llevan a cabo en una organización| y sus interrelaciones. Con ello se obtiene la visión conjunta de todos los aspectos relacionados con cada proceso. Además, se pueden ver todas las interrelaciones existentes entre sus fases, todo esto reflejado en un solo gráfico que es el mapa de procesos. Es decir, el gráfico presenta la visualización general del sistema tal y como está organizada toda la empresa. Estos mapas deben ser concisos y muy claros para que puedan lograr su objetivo: obtener una visión general de todo lo que ocurre en la empresa. En los últimos años, liderado por iniciativas y normas como ISO, se ha incrementado el interés por la gestión por procesos [5], siendo adoptado por muchas organizaciones, tanto locales como internacionales, con el afán de lograr una mejora continua y un incremento en la calidad proporcionada a sus clientes. Este enfoque consiste en la detección y gestión sistemática de los procesos desarrollados en la organización y en particular en el nexo que existe entre ellos. El enfoque por procesos hace uso de herramientas informáticas que facilitan la gestión de los procesos de negocio, estas herramientas se conocen también con el nombre de BPM (Business Process Management). En Argentina, la gestión por procesos, paso a ser tan relevante que dio origen a diferentes modelos de excelencia en la gestión como el Premio Nacional a la Calidad Argentina (Ley 24127/92) para la promoción, desarrollo y difusión de los procesos y sistemas destinados al mejoramiento continuo de los productos y servicios. En Argentina existen aproximadamente unas 8000 empresas con certificaciones ISO 9001, lo que indica que cada vez más las organizaciones identifican el valor agregado de adoptar un enfoque de estas características.

ISBN 978 -987-633-574-4

3 Proceso de Requisitos basado en Escenarios

El Proceso de Requisitos basado en Escenarios [6] tiene una estrategia de construcción secuencial divida en tres etapas: Comprender el UdeD[1] actual, Proyectar el UdeD futuro y Explicitar los Requisitos del Software. La primera etapa propone conocer el dominio en estudio antes de generar una propuesta para el nuevo sistema de software. Para ello se elicita información del dominio y se modela el proceso de negocio tal como existe en el mismo momento de iniciar la IR. El conocimiento obtenido es utilizado para la segunda etapa, Proyectar el UdeD futuro, donde se deben tomar las próximas decisiones acerca de los servicios que tendrá el nuevo sistema de software. La complejidad de esta etapa se debe a la necesidad de proyectar cómo será el proceso del negocio con el sistema de software incluido. Para ello se modelan todas las situaciones involucradas con el nuevo sistema de software, siendo estos modelos los anfitriones de los requisitos. Finalmente, en la tercera etapa, estos requisitos son extraídos y explicitados en una ERS. El formato de este documento dependerá de las políticas organizacionales, de los estándares nacionales o internacionales que se utilicen, siendo uno de los más utilizados el estándar internacional IEEE 830.

Este proceso de requisitos se basa en construir básicamente dos modelos: el Léxico Extendido del Lenguaje (LEL) [7] [8] y los escenarios [9]. El LEL es un glosario cuyo objetivo es describir el léxico del dominio para mejorar la comunicación con el cliente y asegurar la comprensión de todos los artefactos producidos. Los escenarios son narrativas estructuradas de situaciones del contexto, centrando la atención en su comportamiento. Estos modelos pueden representar diferentes puntos de vista dependiendo del momento en el cual se construyen. El LEL es el glosario del UdeD actual y evoluciona al LEL_R [10] en el UdeD futuro. Lo mismo sucede con los escenarios que pueden ser actuales (EA) o futuros (EF) [11]. Estos últimos tienen empotrados los requisitos del software y, por lo tanto, son el modelo central del proceso. Cuando una situación más pequeña está contenida en otra, aparece un sub-escenario. Cuando un escenario tiene una mirada global del contexto y todos sus episodios son escenarios, se transforma en un escenario integrador. Estos últimos se construyen una vez completos, verificados y validados todos los escenarios. Puede observarse en la **Fig. 1** esta jerarquía de escenarios.

[1] Universo de Discurso: "Todo el contexto en el cual el software será desarrollado y operado. Incluye todas las fuentes de información y todas las personas relacionadas con el software. Se utiliza el término Universo de Discurso con el mismo significado que lo utiliza Michael Jackson en [13].

ISBN 978 -987-633-574-4

Fig. 1. Jerarquía de escenarios.

4 IR en organizaciones con Enfoque por Procesos

La propuesta del presente artículo es mostrar diferentes formas de utilizar la información de procesos disponibles en algunas organizaciones para colaborar con el proceso de requisitos. La cantidad y calidad de la información proporcionada por el enfoque por procesos dependerá del grado de madurez en este tipo de gestión [12]. Cabe aclarar que esta propuesta se ha pensado para organizaciones horizontales o departamentales que están comenzando con un enfoque hacia los procesos, por lo tanto, los modelos que se utilizan son los mapas de procesos y las fichas de procesos que se generan al iniciar este nuevo enfoque. En este contexto, cuando es necesario construir un nuevo sistema de software, se hace indispensable que la IR aproveche este conocimiento fuertemente alineado a la organización. La propuesta se basa en el Proceso de Requisitos basado en Escenarios, descripto en la sección anterior, y el conocimiento organizacional existente con el objetivo de mejorar y facilitar la Comprensión del UdeD actual. Además, utilizar estos modelos permite aprovechar el entrenamiento en el uso de estos modelos de procesos que ya tienen los clientes y usuarios y que son parte de la comunicación interna de la organización.

La diferencia sustancial entre la estrategia original de la sección 3 y la que se presenta en esta sección es fundamentalmente pasar de un enfoque middle out a uno top down. La estrategia original consiste en construir la primera versión de los escenarios a partir de la información del LEL y estudiando el contexto observable. Esto produce una descripción de situaciones del contexto de nivel de detalle medio y bajo en la jerarquía de escenarios. Cuando todos los EA están completos y validados, recién entonces se construyen los escenarios integradores. Puede observarse que este orden de construcción se aleja de la estrategia top down promovida por Harlan Mills y Niklaus Wirth en la década del 70, que es ir de lo general a lo particular. Una estrategia top down para el proceso de requisitos comienza construyendo un escenario general (EG) antes de describir los EA. Luego, tomar esta información como guía para elicitar y ordenar el resto de los escenarios. Crear un EG requiere tener conocimiento del contexto muy temprano en el proceso de requisitos o de lo contrario, genera un esfuerzo adicional de actualización cada vez que se incorpora nueva información, con el riesgo de seguir caminos erróneos. Construir el EG a partir de los mapas de procesos permite

ISBN 978 -987-633-574-4

construir un escenario seguro aprovechando los beneficios de la estrategia top down. Utilizar los modelos de proceso guían la elicitación de conocimiento y ayuda al ingeniero de requisitos a insertarse rápidamente en el dominio en estudio con menos esfuerzo cognitivo y con menor probabilidad de asumir erróneamente aspectos del contexto que cree conocer. Trabajar con modelos construidos por la misma organización reduce las subjetividades. Todo esto repercute directamente en una reducción del costo de construcción. También se asegura el alineamiento entre el nuevo sistema de software y la organización. Estos modelos suponen haber resuelto y consensuado los conflictos existentes en el dominio. Por ejemplo, un conflicto entre punto de vistas ("deber ser" y "es"). Cuando se analiza la literatura del dominio se está en presencia del punto de vista de la organización, o sea el "deber ser". Luego, cuando se analiza cómo se opera, pueden aparecer otras formas de realizar las tareas, generando conflictos entre lo que espera la dirección y lo que realmente se hace. De no ser detectados a tiempo, llegarán al nuevo sistema de software y se harán visibles cuando se encuentre en producción. Cuando esto sucede suele ser muy perturbador y el costo de corrección se eleva aún más.

En la **Fig. 2** se presenta la relación que existe entre los modelos de procesos y los de requisitos. Puede observarse que el Objetivo General del Sistema se alimenta de las mejoras definidas para los procesos ya que en muchas organizaciones la definición de procesos está en el marco de un Sistema de Gestión de Calidad (SGC), como es el caso de la ISO 9001. En estos contextos, las mejoras deben ser absorbidas por el Objetivo General de Sistema para ser tratadas adecuadamente cuando se proyecte el UdeD futuro. Sin lugar a dudas, la utilización de los modelos de procesos mejora la definición de los modelos de requisitos del UdeD actual, actuando uno como validación del otro.

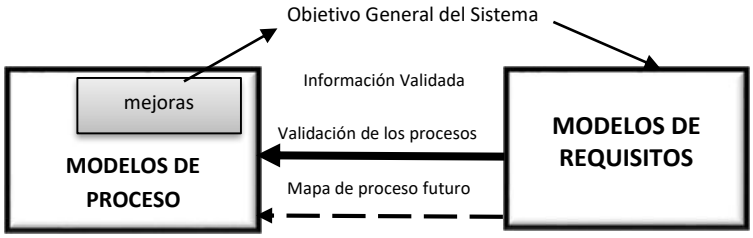


Fig. 2. Relación entre modelos de procesos y los modelos de requisitos.

Algunas ventajas de utilizar los modelos de procesos en la IR:

- Garantiza una correcta comprensión del UdeD actual.
- Reduce los costos y el tiempo de elicitación, modelado y análisis (V&V).
- Asegura un mejor alineamiento con la organización.
- Reduce los conflictos del dominio, ya que se espera que los diferentes puntos de vista estén resueltos.

ISBN 978 -987-633-574-4

Se puede observar en la Fig. 3 que el proceso de requisitos se puede realizar completo (opción 1) o se puede realizar un mix entre los modelos organizacionales y los modelos de requisitos (opción 2). En el primer caso se cuenta solo con un mapa de procesos. En el segundo caso existen diferentes modelos de procesos (mapas, fichas de procesos, etc.) con información suficiente para comprender el UdeD actual. En ambos casos se construye el LEL como primera actividad para homogeneizar el léxico utilizado. Cabe destacar que la propia construcción del LEL sirve para verificar las definiciones de procesos y detectar posibles omisiones e inconsistencias. Con el LEL completo, en la opción 1 de la Fig. 3, se aparean los impactos de los símbolos Sujetos, los cuales definen actividades o tareas, con las actividades del mapa de procesos. El objetivo de este paso es identificar los procesos involucrados y teniendo en cuenta que el LEL se construye en el marco del Objetivo General del Sistema, los procesos seleccionados a partir de él también lo están. Con esta información se construye el EG (escenario general) el cual será la guía para elicitar y completar los EA. El resto del proceso de requisitos que se definió en la sección 3, no se modifica.

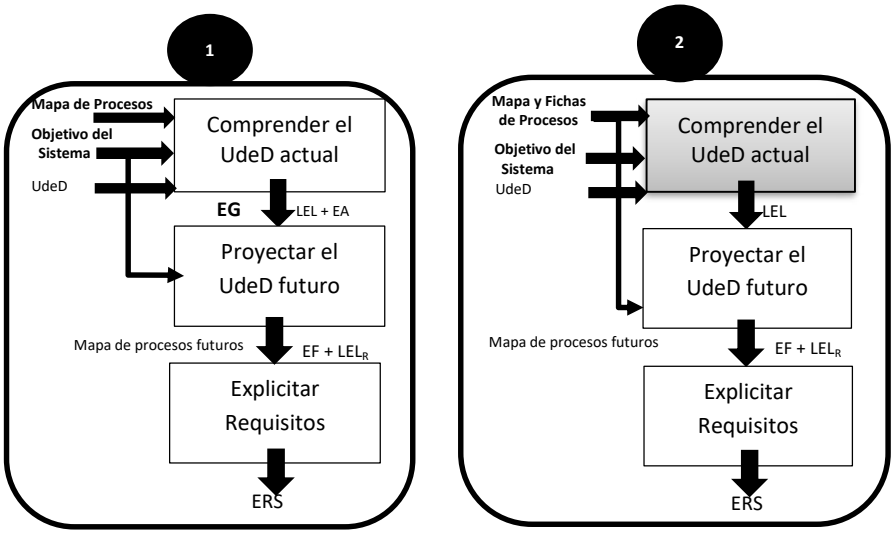


Fig. 3. Estrategias de IR utilizando la definición de procesos organizacionales.

En la opción 2 de la Fig. 3 también se realiza un enfoque top down y se utiliza nuevamente un mix de conocimiento pero con más fuerza en los modelos de procesos. Esta opción es la correcta cuando existen definiciones más detalladas de las actividades, como es el caso de las fichas de procesos u otra representación. Se plantea comprender el UdeD actual utilizando los modelos de procesos existentes, ya que construir los EA se reduce a duplicar dicho conocimiento. En este caso para seleccionar las actividades involucradas se deben analizar los modelos de procesos en el contexto del Objetivo

ISBN 978 -987-633-574-4

General del Sistema y para construir los EF se deben recorrer los procesos-actividades observando cómo se modifican en el UdeD futuro. Esta construcción es procedural si los cambios en el proceso del negocio son menores o un enfoque por objetivos cuando son significativos. En la práctica, se espera que el enfoque de construcción sea híbrido, o sea cualquiera de los dos o ambos dependiendo de cómo impactan los cambios en cada actividad. Si bien cada organización puede utilizar diferentes modelos de procesos, es muy común encontrar mapas de procesos que permiten segmentar los procesos y visualizarlos en conjunto con sus relaciones asociadas. La definición del mapa de procesos se transforma en sí misma en una forma de comunicación dentro de la organización, lo que puede aportar un doble beneficio a la IR. Por un lado, al utilizar modelos previamente validados en la organización se abre un camino seguro al ingeniero de requisitos para comprender el contexto en estudio y por el otro lado, mejora la comprensión del cliente durante la construcción de los requisitos de software. Los *mapas de procesos futuros* permiten una validación cruzada entre los EF construidos y la representación de procesos existentes, ya que la construcción de este mapa permite validar estos EF y ese proceso actúa como un mecanismo de autoevaluación del mapa.

4.1 Ingeniería de Requisitos aplicada al caso Norpak

Este ejemplo corresponde a una empresa denominada Norpak la cual fabrica cajas de cartón corrugado. En este ejemplo se aplicó la Opción 1 de la Fig. **3**. Por cuestiones de espacio, el presente ejemplo tuvo que ser recortado.

En un primer momento se generó el mapa de procesos que se presenta en el *Paso 1* de la Fig. 4. Luego, se tomó un LEL existente y se separaron los impactos de los Sujetos. Con esta información se creó la tabla del *Paso 2* donde se relaciona el LEL con el mapa de procesos. Se debe recordar que los impactos del LEL son actividades del dominio y, por consiguiente, son parte de algún proceso. Una vez identificado dicho proceso se obtuvieron las actividades desde el mapa. Con esta información se construyeron varios EG, pero en este ejemplo solo se describe “Planificar la Producción” (*Paso 3* de la Fig. 4). Puede observarse que se construyeron tantos EA como episodios tiene el EG (*Paso 4* de la Fig. 4). Este EG se utilizó como guía para describir los escenarios. Del mapa de procesos se obtuvieron los responsables de los procesos a quienes se entrevistaron para completar la primera versión de los EA y quien determinó cómo elicitar la información faltante, o sea con otras entrevistas o en documentos existentes. Cabe recordar que una vez identificados los EA, el resto del proceso de requisitos es igual al descripto en la sección 3.

La construcción de un mapa de procesos futuros a partir de los EF es decisión de cada organización. Así cómo es posible construir un EG desde el mapa de procesos, el camino inverso puede ser automatizado sin ningún inconveniente.

ISBN 978 -987-633-574-4

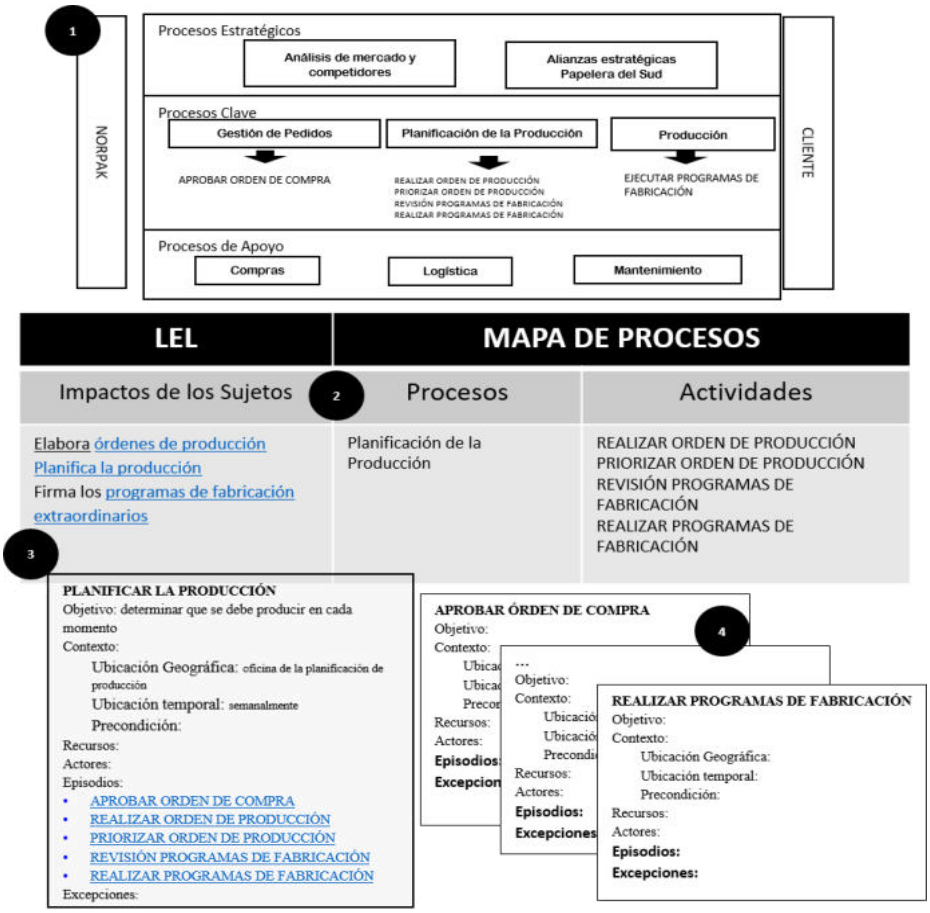


Fig. 4. Ejemplo de una IR con mapas de procesos.

5 Conclusiones

El presente artículo se centra en organizaciones que han comenzado su transformación hacia la horizontalidad. Por la diversidad de casos, se decidió utilizar los mapas de procesos por ser un modelo común a casi todas. La reutilización de los modelos organizacionales reduce significativamente el esfuerzo de elicitación y validación para comprender el contexto actual. Además, asegura un fuerte alineamiento con la organización. Otro aspecto de gran valor es mantener la forma de comunicación existente. De esta manera se aprovecha el entrenamiento de los clientes y usuarios en el uso de esos modelos, mejorando su participación y cooperación en el nuevo proyecto de software. En resumen, las estrategias presentadas en este artículo, permiten realizar

ISBN 978 -987-633-574-4

una IR de mayor calidad impactando favorablemente en el tiempo y costo de los requisitos de software.

6 Trabajos Futuros

Se espera probar las estrategias en organizaciones con diferente grado de madurez en el enfoque por procesos. De esta manera se podrá analizar la importancia de la retroalimentación de conocimiento entre los modelos de proceso y los de requisitos. También se espera estudiar cómo se integran los modelos de procesos en el diseño.

Referencias

1. Pardo, Isabel, "Organización vertical versus horizontal", ESIC MARKET. nº 117. 182-197, 2004.
2. Hammer, M. (1996). Beyond Reengineering: How the Process– Centered Organization is Changing Our Work and Our Lives. New York: Harper Collins.
3. Mallar Miguel Ángel, "LA GESTIÓN POR PROCESOS: UN ENFOQUE DE GESTIÓN EFICIENTE", Universidad Nacional de Cuyo, "Visión de Futuro" Año 7, Nº1 Volumen Nº13, Enero - Junio 2010 http://www.fce.unam.edu.ar/revistacientifica/index.php?option=com_content&view=article&id=184&Itemid=51
4. R.G. Lee , B.G. Dale, Business process management: a review and evaluation, Business Process Management Journal, ISSN: 1463-7154, 1998
5. Jan vom Brocke and Michael Rosemann, Handbook on Business Process Management, editores, Springer Link, ISBN: 978-3-642-45100-3, 2015.
6. Leite, J.C.S.P., Doorn, J.H., "Perspectives on Software Requirements: An introduction", en el libro "Perspectives on Software Requirements", Kluwer Academic Publishers, EEUU, ISBN: 1-4020-7625-8, Capítulo 1, 2004.
7. Leite, J.C.S.P., Franco, A.P.M., "O Uso de Hipertexto na Elicitaçao de Linguagens da Aplicaçao", Anais de IV Simpósio Brasilero de Engenharia de Software, SBC, SBC, pp 134-149, 1990.
8. Hadad, G.D.S., Doorn, J.H., Kaplan, G.N., "Creating Software System Context Glossaries", Encyclopedia of Information Science and Technology, Idea Group Publishing, 2º edición, 2007.
9. Leite, J.C.S.P., Hadad, G.D.S., Doorn, J.H., Kaplan, G.N., "Scenario Construction Process", Requirements Engineering Journal, Springer-Verlag London Ltd., Vol.5, Nº1, pp. 38-61, 2000.
10. Kaplan G., Doorn J., Gigante, N., "Evolución Semántica de los Glosarios en los Procesos de Requisitos", CACIC14. 2013.
11. Doorn, J.H., Hadad, G.D.S., Kaplan, G.N., "Comprendiendo el Universo de Discurso Futuro", WER'02 - Workshop en Ingeniería de Requisitos, Valencia, España, pp.117-131, noviembre 2002. http://www.cucsur.udg.mx/sites/default/files/iso_9001_2015_esp_rev.pdf
12. Gabriel Páez, Claudia Rohvein, Diana Paravie, Mario Jaureguiberry, "Revisión de modelos de madurez en la gestión de los procesos de negocios", Ingeniare. Revista chilena de ingeniería, vol. 26 N° 4, 2018, pp. 685-698.
13. Jackson, M., "Software Requirements & Specifications. A lexicon of practice, principles and prejudices", Addison Wesley, ACM Press, 1995.

ISBN 978 -987-633-574-4

Evaluación de metodologías para la validación de requerimientos

Sonia R. Santana[1], Leandro Antonelli[2], Pablo Thomas[3]

(1) Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos
sonia.santana@uner.edu.ar
(2) Laboratorio de Investigación y Formación en Informática Avanzada (LIFIA), Facultad de Informática, Universidad Nacional de La Plata
leandro.antonelli@lifia.info.unlp.edu.ar
(3) Instituto de Investigación en Informática (III-LIDI). Facultad de Informática, Universidad Nacional de La Plata - Centro Asociado CIC. Buenos Aires, Argentina
pthomas@lidi.info.unlp.edu.ar

Abstract. Una de las fases más importantes en el desarrollo de un proyecto de software es la validación de los requerimientos. Los requerimientos erróneos, si no se detectan a tiempo, pueden causar problemas, como costos adicionales, incumplimiento de los objetivos esperados y retrasos en las fechas de entrega. Por estas razones, es beneficioso dedicar un esfuerzo a esta tarea. Con el fin de utilizar la metodología adecuada, en este trabajo se realizó un relevamiento de las principales tendencias de la validación de requerimientos del software desde el año 2007 hasta el año 2021. Se seleccionaron y evaluaron exhaustivamente cuatro metodologías a partir de un marco "Way-of" o "forma de" y de un modelo de referencia para revisiones técnicas.

Keywords: Ingeniería de requerimientos, validación de requerimientos, metodología para la validación de requerimientos.

1 Introducción

En el marco de la Ingeniería de Requerimientos (RE por sus siglas en inglés Requirements Engineering) la validación de los requerimientos es una tarea fundamental en cualquier proyecto de Ingeniería de Software y debe ser un proceso continuo en el ciclo de vida del desarrollo del sistema. El principal objetivo de la validación de requerimientos es confirmar que los requerimientos especificados sean representaciones de las necesidades y expectativas de los usuarios [1] [2] [3] y que además sean completos, correctos y consistentes [4] entre otras características.

Según Kotonya la validación de requerimientos se refiere a verificar la coherencia, integridad y corrección del documento de requerimientos [3], y también se establece que los requerimientos deben ser: validos, comprensibles, consistentes, trazables, íntegros, reales y verificables [5]. Según Bahill [6] el proceso de validación de requerimientos consiste en primer lugar en asegurar que un conjunto de requerimientos sean: correctos, completos y consistentes; en segundo lugar si se puede crear un modelo que cumpla con los requerimientos, y por último que se pueda construir y probar una solución de software en el mundo real para demostrar que cumple con los requerimientos de las partes interesadas.

Trabajar en la validación de requerimientos se está convirtiendo en un desafío para los equipos, clientes y usuarios. Existen diferentes causas que imponen problemas de comunicación, control, intercambio de conocimientos, confianza y retrasos en el desarrollo del software [7].

ISBN 978 -987-633-574-4

Este trabajo se enfoca en las principales tendencias de la validación de requerimientos del software y presenta una evaluación exhaustiva de cuatro metodologías para la validación de requerimientos.

El resto del artículo está organizado de la siguiente manera, en la sección 2 se realiza una revisión de la literatura para seleccionar las metodologías para la validación de requerimientos. En la sección 3 se evalúan las metodologías seleccionadas con un marco "Way-of" o "forma de" y un modelo de referencia para revisiones técnicas. En la sección 4 se analizan los resultados obtenidos. Finalmente, en la sección 5 se expresan conclusiones y trabajo futuro.

2 Revisión bibliográfica

En esta sección se realizó un proceso de tres fases: búsqueda, selección y evaluación de trabajos para una revisión bibliográfica de metodologías para la validación de requerimientos.

La pregunta a responder para el proceso de revisión es ¿Cuál es la tendencia de las prácticas relacionadas con el proceso de validación de requerimientos? Las búsquedas se realizaron en artículos publicados en las fuentes de información IEEE, Elsevier, Springer y ACM Digital Library, publicados entre 2007 y 2021.

Los criterios de inclusión y exclusión son los siguientes: a) Los artículos seleccionados deben estar directamente relacionados con el tema de validación de requerimientos en el área de RE. b) Los artículos no deben discutir la validación de requerimientos en la fase de prueba del software desarrollado o en la implementación en prueba con respecto a los requerimientos. c) Los artículos no deben utilizar términos como "Revisión sistemática de la literatura", "Revisión de la literatura", "Análisis sistemático".

Luego, se aplicaron los criterios de evaluación presentados en la Tabla 1. La lista de criterios de evaluación fue adaptada de [8] y permite seleccionar los artículos finales en base a seis criterios que se enumeran.

Tabla 1. Criterios de evaluación.

Sección	*Criterios*
Introducción	1. ¿La introducción proporciona una descripción general del aporte para la validación de requerimientos? 2. ¿Qué tipo de aporte es? Una metodología, enfoque, método, técnica, marco o herramienta. 3. ¿Está claramente definido el propósito / objetivo de la investigación?
Metodología	4. ¿Está claramente definida la metodología de investigación?
Resultados	5. ¿Están los hallazgos claramente establecidos? ¿Los resultados ayudan a resolver los problemas de validación de requerimientos?
Conclusión	6. ¿Existen límites o restricciones impuestas a la afirmación de conclusión?

Esta revisión bibliográfica permitió preseleccionar 38 artículos de los cuales posteriormente se seleccionaron solo aquellos que aportan una metodología para el proceso de validación de requerimientos. El resultado de la selección final se presenta en la Tabla 2.

Tabla 2. Artículos seleccionados de la revisión bibliográfica.

Ref. Bibliog.	*Ref. Estudio*	*Año*	*Nombre*	*Dominio de aplicación*
[9]	M1	2007	FBCM	Empresa
[10]	M2	2009	From Informal Requirements to Property-Driven Formal Validation	Sistema con seguridad crítica.
[11]	M3	2011	CoReVDO	Sistemas distribuídos.
[12]	M4	2020	A Methodology of Requirements Validation for Aviation System Development	Sistemas de aviación civil

ISBN 978 -987-633-574-4

3 Evaluación de las metodologías

En esta sección se evalúan las metodologías seleccionadas con un marco “Way-of” o “forma de” y un modelo de referencia para revisiones técnicas para conocer sus características, necesidades de información y restricciones.

3.1 Definición del marco de referencia para evaluar las metodologías.

Existen metodologías que introducen modelos con poco conocimiento, otras metodologías ofrecen algoritmos, o al menos procedimientos explícitos para construir un modelo específico o verificarlo. Sin embargo, otras metodologías brindan sugerencias informales pero prácticas para obtener un modelo. Por tanto, es prudente distinguir entre la "forma de modelar" y la "forma de trabajar" de una metodología. La “forma de control” está en esencia relacionada con el control de tiempo, costos y calidad del proceso de desarrollo de los sistemas de información y sus productos.

Según Seligmann [13], las metodologías se pueden describir diferenciando entre una forma de modelar, una forma de trabajar y una forma de control. Sin embargo, para comprender realmente una metodología Sol [14] y Kensing [15] consideran necesario conocer su filosofía subyacente o “forma de pensar” utilizada para mirar las organizaciones y los sistemas de información. El marco “Way-of” o “forma de” que se utiliza para evaluar las metodologías con cinco aspectos diferentes, cada uno de ellos se describe con explicaciones sobre sus contribuciones para la gestión flexible del proceso de desarrollo [16]. Los cinco aspectos del marco “Way-of” o “forma de” son:

- **Forma de Pensar:** Define lineamientos de desarrollo de la metodología, por lo tanto provee una perspectiva del dominio del problema y hace explicita las suposiciones, principios y estrategias sobre la misma.
- **Forma de Modelar:** Provee información de los conceptos para el modelado. Proporciona formalismo y notación para expresar modelos de proceso de la metodología.
- **Forma de Trabajar:** Define la estructura del proceso de la metodología, las actividades, tareas y la secuencia en el que se deben llevar a cabo.
- **Forma de Controlar:** Define los medios que ofrece una metodología para determinar cómo se debe controlar y evaluar el modelo de la metodología.
- **Forma de Soportar/Apoyar**: Se refiere a las técnicas, herramientas y/o ayudas de trabajo que apoyan la ejecución del proceso de la metodología.

A partir de estos aspectos se elaboraron un conjunto de preguntas para poder evaluarlos. La Tabla 3 presenta estas preguntas.

Tabla 3: Conjunto de preguntas para evaluar los aspectos del marco “Way of” o “forma de”

Forma de	*Preguntas de evaluación*
Pensar *(Enmarca a la metodológica)*	• ¿Cuál es la función de la metodología? • ¿Cuáles son los componentes de la metodología? • ¿Cuál es el entorno de la metodología? • ¿Cuáles son los componentes del entorno de la metodología? • ¿Cuáles son las características de la metodología y su entorno?
Modelar *(Proceso operativo orientado al producto de la metodología)*	• ¿Qué modelo utiliza la metodología? • ¿La metodología describe los componentes y sus relaciones dentro de los modelos utilizados? • ¿La metodología describe las relaciones entre los modelos utilizados?
Trabajar *(Proceso operativo orientado al proceso de la metodología)*	• ¿Cuáles son las actividades y tareas de la metodología? • ¿Cuáles son las responsabilidades de las actividades y tareas de la metodología?
Controlar *(Proceso orientado a controlar y evaluar el modelo)*	• ¿Cuál es la forma de control que utiliza la metodología? • ¿Cuáles son los objetivos medidos a través de indicadores?
Soportar/Apoyar *(Proceso de apoyo para la ejecución de la metodología)*	• ¿Cuáles son las técnicas, herramientas y/o ayudas de trabajo que apoyan la metodología? • ¿Qué estándares utiliza la metodología?

ISBN 978 -987-633-574-4

3.2 Evaluación de las metodologías según el marco de referencia.

Esta subsección describe el análisis de cada una de las metodologías seleccionadas en la Tabla 2. El análisis consiste en responder a cada una de las preguntas de la Tabla 3. Las tablas 4, 5, 6 y 7 presentan las respuestas para cada metodología: M1, M2, M3 y M4.

Tabla 4. Evaluación de la Metodología M1

Forma de	***Metodología M1***
Pensar	**Función**: definir y validar requerimientos comerciales que se utilizan como requerimientos funcionales de software. **Componentes**: modelos de colaboración de hechos. **Entorno**: sistemas empresariales. **Componentes del entorno**: planificación de sistema, análisis y especificación de requerimientos. **Características**: evaluar la integridad de las metas y objetivos fundamentales de la organización para el desarrollo del sistema de tecnología de la información.
Modelar	**Modelo**: BSC (Balanced Score Card) se utiliza para desarrollar estrategias en las empresas. **Componentes**: Perspectiva: puntos de vista del objetivo estratégico- Procesos críticos – Objetivos estratégicos de la empresa – Indicador clave de rendimiento (KPI) – Relación causa-efecto entre los objetivos estratégicos. **Relación entre los componentes**: Visualizar la estrategia de BSC, adicionar objetivos por observación, analizar la estructura de estrategia, evaluar la validez de la estructura de la estrategia y extraer funciones del sistema.
Trabajar	La metodología propone un enfoque dividido en 5 pasos: 1. Visualizar la estrategia de BSC. Generar un árbol de análisis de objetivos. 2. Adicionar objetivos por observación 3. Analizar la estructura de estrategia. Asignar KPI a cada objetivo estratégico. 4. Evaluar la validez de la estructura de la estrategia, mediante análisis estadístico de los datos de KPI. 5. Extraer resultados y refinar el árbol de análisis de objetivos.
Controlar	**Control**: biblioteca de 700 KPI para que los desarrolladores elijan KPI fácilmente divididas en 4 perspectivas: financiera, del cliente, de procesos de negocio, de aprendizaje y conocimiento.
Soportar/ Apoyar	**Técnicas**: árbol de análisis de objetivos, tarjetas de observaciones de campo, mapa de estrategias y matriz de colaboración.

Tabla 5. Evaluación de la Metodología M2

Forma de	***Metodología M2***
Pensar	**Función**: validar una especificación de requerimientos escrita en lenguaje informal. **Componentes**: modelos formales para la validación de requerimientos. **Entorno**: sistemas complejos de seguridad crítica. **Componentes del entorno**: especificación de los requerimientos escrita en un lenguaje informal. **Características**: Formalizar los requerimientos en el uso del lenguaje de modelado unificado (UML) y en el uso de un lenguaje natural controlado (CNL), basado en un subconjunto del lenguaje de especificación de propiedades (PSL).
Modelar	**Modelo**: lenguaje de modelado unificado (UML), lenguaje natural controlado (CNL) y lenguaje de especificación de propiedades (PSL). **Componentes**: diagrama de clases, máquinas de estado y diagramas de secuencia. **Relación entre los componentes**: En primer lugar, la metodología prevé un análisis informal del documento de requerimientos para categorizar cada requerimiento. Luego, cada fragmento de requerimiento se formaliza según la categorización mediante diagramas UML y el uso de un Lenguaje Natural Controlado. Por último, se realiza un análisis formal automático para identificar posibles fallas en los requerimientos formalizados.
Trabajar	La metodología propone un enfoque dividido en 3 fases: 1. Fase de análisis de los requerimientos informales basados en inspecciones para identificar fallas. 2. Fase de formalización. Cada fragmento de requerimiento identificado en la fase de análisis informal especificando los conceptos y diagramas de UML correspondientes y / o las restricciones de CNL. Vincular los elementos UML con los requerimientos textuales. 3. Fase de validación formal. Verificar, reducir y validar los fragmentos de requerimientos formalizados.
Controlar	**Control:** *Comprobación de vacuidad*: comprobar si una propiedad dada se mantiene de forma permanente. *Comprobación de la cobertura*: comprobar qué elementos del fragmento de requerimiento considerado formalizado han sido estimulados (cubiertos) por una traza generada. *Análisis de seguridad*: identificar las causas que conducen a la violación de una propiedad, es decir, identificar variables de interés que son causas de una infracción específica para que algoritmos avanzados pueden recopilar una descripción de las causas y organizarlas en forma de árbol de fallas.

ISBN 978 -987-633-574-4

Forma de	*Metodología M2*
Soportar/ Apoyar	**Técnicas**: diagrama de clases, diagrama de secuencias, máquinas de estado y lenguaje natural controlado. **Herramientas**: desarrolladas en base a estándares industriales: IBM Rational RequisitePro (RRP), interconectado con Microsoft Word, e IBM Rational Software Architect (RSA), para soportar la fase de análisis informal y la trazabilidad del vínculo entre los fragmentos de requerimientos informales y sus contrapartes.

Tabla 6. Evaluación de la Metodología M3

Forma de	*Metodología M3*
Pensar	**Función**: verificar, negociar y validar los requerimientos distribuidos mediante un conjunto de actividades. **Componentes**: múltiples puntos de vista y competencias cognitivas de las partes interesadas en un proceso distribuido y colaborativo. **Entorno**: sistemas distribuidos. **Componentes del entorno**: documento de especificación de requerimientos. **Características**: trabajar con equipos distribuidos geográficamente e incluir al cliente en el proceso de validación colaborativo.
Modelar	**Modelo**: Lenguaje de Modelado Unificado (UML). **Componentes**: diagrama de actividad. **Relación entre los componentes**: Presenta tres actividades principales: Organización, Verificación distribuida y Validación colaborativa.
Trabajar	La metodología propone un enfoque dividido en 3 fases: 1. Organizativa. Responsabilidad: Analista. Equipo de RE. 2. Verificación Requerimientos Funcionales y no Funcionales. Responsabilidad: Equipo de RE. 3. Validación colaborativa. Responsabilidad: Equipo de RE.
Controlar	**Control**: tasa de motivación y compromiso entre las partes interesadas. Si la tasa es positiva, el acuerdo entre las partes es fuerte; de lo contrario, es moderado o negativo. Coherencia global donde un requerimiento no debe contradecir otros requerimientos establecidos después de la integración de requerimientos globales.
Soportar/ Apoyar	**Técnicas**: cheklist, orientado al punto de vista, inspección, revisión y prototipos. **Estándar**: IEEE 830 para garantizar la calidad de los requerimientos y los entregables.

Tabla 7. Evaluación de la Metodología M4

Forma de	*Metodología M4*
Pensar	**Función**: validar una especificación de requerimientos y cumplir con la certificación ARP4754A, consideración especial del ciclo de vida de desarrollo de aeronaves y sistemas civiles. **Componentes**: modelos formales para la validación de requerimientos. **Entorno**: sistemas de aviación. **Componentes del entorno**: especificación de los requerimientos. **Características**: validar los requisitos considerando las características del diseño de aeronaves y las regulaciones de certificación durante el ciclo de vida del desarrollo del producto.
Modelar	**Modelo**: Modelo de proceso de validación basado en el cumplimiento de los objetos requeridos por los reguladores de certificación SAE ARP 4754A. **Componentes**: Plan de validación, rigor de validación, verificaciones de corrección y exactitud, matriz de validación (inicial y final) e informe de validación. **Relación entre los componentes**: En primer lugar, la metodología desarrolla un plan de validación, luego crea una matriz de validación donde verifica la corrección y comprueba la integridad para realizar su actualización. Por último genera un resumen para revisar las actividades de la validación.
Trabajar	La metodología propone un proceso dividido en: 1. Crear plan de validación. 2. Crear una matriz de validación de requerimientos. 3. Generar una lista de verificación de corrección de requerimientos. 4. Generar una lista de verificación de integridad de requerimientos. 5. Actualizar la matriz de validación de requerimientos. 6. Resumen de validación de requerimientos. 7. Revisar actividades de validación de requerimientos.
Controlar	**Control:** *Listas de verificación* comprueba la exactitud e integridad de los requerimientos.
Soportar/ Apoyar	Técnicas: cheklist, pruebas, inspección y revisión. Herramientas: Para la gestión de requisitos, ReQtest, IBM Rational Doors, Visure Requirements. Estándar: EIA 632 para el proceso de validación de requerimientos.

La Tabla 8 resume la evaluación de las metodologías a partir de la información de las Tablas 4, 5, 6 y 7. Para cada metodología de validación se indica con el vocablo SI que la

metodología cumple con el proceso del marco de referencia para el desarrollo de metodologías.

Tabla 8. Evaluación de las metodologías según el marco de referencia.

Metodologías / *Proceso*	*M1*	*M2*	*M3*	*M4*
Definición *(Forma de pensar)*	SI	SI	SI	SI
Modelado *(Forma de modelar)*	SI	SI	SI	SI
Operación *(Forma de trabajar)*	SI	SI	SI	SI
Control *(Forma de controlar)*	SI	SI	SI	SI
Soporte *(Técnicas/Herramientas)*	SI	SI	SI	SI
Dominio de aplicación	Sistemas para empresas	Sistemas complejos con seguridad crítica.	Sistemas distribuidos	Sistemas de aviación civil

3.3 Identificación de procesos complementarios al marco "Way-of" a través de un modelo para la evaluación de las metodologías.

En este apartado se presenta un modelo un modelo de referencia para revisiones técnicas utilizado por Pressman [17] para identificar los procesos complementarios al marco "Way-of" en la evaluación de las metodologías.

En este modelo, las actividades son: roles de los individuos, planeación y preparación, estructura de la reunión y corrección y verificación.

Figura 3. Modelo de referencia para revisiones técnicas [17].

Cada una de las características del modelo de referencia ayuda a definir el nivel de formalidad de la revisión. La formalidad de una revisión se incrementa cuando: 1) se definen explícitamente roles distintos para los revisores, 2) hay suficiente cantidad de planificación y preparación para la revisión, 3) se define una estructura distinta para la revisión (incluso tareas y productos internos del trabajo) y 4) el seguimiento por parte de los revisores tiene lugar para cualquier corrección que se efectúe [17].

En este contexto, se pueden definir tres actividades complementarias para el proceso de validación de requerimientos:

Planificación. Actividad que permite definir con anticipación los objetivos que se obtendrán, los requerimientos que se revisarán, las personas que participarán, el procedimiento, técnicas y herramientas que se utilizarán. Si se encuentran estos elementos sugeridos se lo puede denominar plan de validación de requerimientos.

Gestión de defectos. Comprende la identificación y documentación de defectos como resultado de la aplicación del proceso de validación, pero adicionalmente se debe establecer las acciones para su corrección, lo que finalmente lleva a realizar un seguimiento y el control de estado del defecto.

Aceptación. La aceptación de requerimientos es la última actividad del proceso de validación, esta actividad es el punto donde se conecta la Ingeniería de Requerimientos con la siguiente fase del ciclo de vida del desarrollo de software. Esta actividad asegura contar

con la aceptación del cliente sobre las especificaciones, condiciones, restricciones y parámetros de calidad que se deberá verificar posteriormente en el producto de software.

3.4 Evaluación de las metodologías según las actividades complementarias del modelo de referencia.

En la Tabla 9 se presenta la evaluación de las metodologías según las actividades complementarias obtenidas del modelo de referencia, en la cual se indica con el vocablo SI que la metodología contribuye al proceso para la validación de requerimientos, mientras que el vocablo NO indica que la metodología no contribuye al proceso para la validación de requerimientos.

Tabla 9. Evaluación de las metodologías según procesos complementarios del modelo de referencia.

Metodologías / *Actividad*	*M1*	*M2*	*M3*	*M4*
Planificación	NO	NO	SI	SI
Gestión de defectos	SI	SI	SI	SI
Aceptación	NO	NO	NO	SI

4. Resultados

Del análisis de los resultados obtenidos en la evaluación de las metodologías indicados en la Tabla 9 y los aspectos evaluados en las metodologías en las Tablas 4, 5, 6, y 7 se extraen diferentes características que son calificadas como puntos de evaluación para las metodologías seleccionadas. En la Tabla 10 se enumeran las características identificadas, en la cual se indica con el vocablo SI que la metodología contribuye a la característica, mientras que el vocablo NO indica que la metodología no contribuye a la característica.

Tabla 10. Evaluación de las metodologías para validación de requerimientos.

Metodologías / *Característica*	*M1*	*M2*	*M3*	*M4*
Definición				
Define la funcion	SI	SI	SI	SI
Define los componentes	SI	SI	SI	SI
Define el entorno	SI	SI	SI	SI
Define componentes del entorno	SI	SI	SI	SI
Define caracteristicas	SI	SI	SI	SI
Modelado				
Especifica el modelo	SI	SI	SI	SI
Especifica componentes del modelo	SI	SI	SI	SI
Especifica relaciones entre los componentes del modelo.	SI	SI	SI	SI
Operación				
Define una planificacion.	NO	NO	SI	SI
Define fases, actividades y tareas.	SI	SI	SI	NO
Define requerimientos a validar.	SI	SI	SI	SI
Define roles de las tareas.	NO	NO	SI	SI
Define funciones y responsalidades de los roles.	NO	NO	SI	SI
Control				
Define la forma de control.	SI	SI	SI	SI
Define indicadores de rendimiento.	SI	NO	NO	NO
Define objetivos medidos por indicadores.	SI	NO	NO	NO
Gestión de defectos				
Identifica los defectos.	SI	SI	SI	SI
Documenta los defectos.	SI	SI	SI	SI
Seguimiento de defectos	SI	SI	SI	SI
Aceptación				
Considera la aceptación del cliente sobre las especificaciones	NO	NO	SI	SI
Considera la aceptación del cliente sobre las condiciones	NO	NO	SI	SI
Considera la aceptación del cliente sobre las restricciones	NO	NO	SI	SI
Considera la aceptación del cliente sobre los parámetros de calidad.	NO	NO	SI	SI
Valida los requerimientos en el ciclo de vida del desarrollo de software	NO	NO	NO	SI

Metodologías / *Característica*	M1	M2	M3	M4
Soporte				
Define técnicas.	SI	SI	SI	SI
Define herramientas.	NO	SI	NO	SI
Utiliza estandares.	NO	NO	SI	SI
Dominio de aplicación				
Define dominio de aplicación.	SI	SI	SI	SI

Del análisis comparativo realizado entre las metodologías seleccionadas se extraen los siguientes resultados:

1. *En el proceso de definición, modelado gestión de defectos y definición de dominio de aplicación* las metodologías cumplen con todas las características propuestas.
2. *En el proceso de operación las metodologías*: M3 y M4 definen un proceso de planificación para la validación de requerimientos. M1, M2, y M3 definen una estructura compuesta por fases, actividades y tareas. Mientras que M4 solo define una estructura compuesta por actividades y tareas. M2, M3 y M4 validan los requerimientos sobre la especificación de requerimientos y M1 valida los requerimientos comerciales. M3 y M4 definen roles y responsabilidades.
3. *En el proceso de soporte las metodologías:* M1, M2, M3 y M4 definen técnicas para validación de requerimientos. Además, M2 y M4 especifican herramientas. M3 implementa el estándar IEEE830 [18] para la especificación de requerimientos y M4 el EIA632 [19] para el proceso de validación de requerimientos.
4. *En el proceso de control las metodologías:* M1, M2, M3 y M4 implementan una forma de control para la validación de requerimientos, pero solo M1 define indicadores de rendimiento y objetivos basados en dichos indicadores.
5. *En el proceso de aceptación las metodologías:* M3 y M4 consideran la aceptación del cliente sobre las especificaciones, restricciones, condiciones y parámetros de calidad. M4 valida los requerimientos en todo el ciclo de vida del desarrollo del software.

5. Conclusiones y Trabajo Futuro

Este trabajo se ha focalizado en un proceso fundamental de la Ingeniería de Requerimientos: la Validación. Se han preseleccionado 38 trabajos para posteriormente centralizarse en 4 de ellos. Para responder a las preguntas de evaluación de las cuatro metodologías seleccionadas se identificaron las contribuciones de los trabajos en el proceso de Validación de Requerimientos.

Las metodologías M1, M2, M3 y M4 proporcionan correctamente sus funciones, componentes, entornos y características, Además detallan los componentes del modelado, las relaciones entre ellos e identifican, documentan y realizan el seguimiento de los defectos.

Adicionalmente, las metodologías M3 y M4 aportan la actividad de planificación compuesta por fases, actividades y tareas, donde se definen roles y describen funciones y responsabilidades. Ambas metodologías emplean técnicas de control, incorporan el uso de estándares y reúnen la aceptación del cliente/usuario en el proceso de validación de requerimientos.

Todas las metodologías fueron desarrolladas para dominios específicos de aplicación.

Si bien las metodologías cumplen con la mayoría de las características del marco “Way of” o “forma de” y del modelo de referencia para revisiones técnicas utilizado para su

evaluación se encontraron algunos problemas que deben considerarse al momento de aplicarlas.

Las metodologías proporcionan diferencias en el enfoque de dominio para validar los requerimientos con diversos grados de éxito. Este éxito depende de la naturaleza de la organización en sí y del conocimiento profundo del negocio para adaptar la metodología a las necesidades del negocio y del usuario.

Se observa la escasa participación de los usuarios/clientes en el proceso de validación de requerimientos, integrado con la falta de enfoque por parte de las metodologías sobre la gestión del nivel de responsabilidad, el grado de toma de decisiones y el equilibrio entre usuarios y desarrolladores.

Las metodologías realizan la validación de requerimientos sobre la especificación de requerimientos, es decir, no se aplican en las diferentes etapas del ciclo de vida del desarrollo del software. Si bien la metodología M4 emplea la validación de los requerimientos en todo el ciclo de vida del sistema esto se debe al uso de estándar de aviación ARP4754A [20], donde respalda la certificación de sistemas de aeronaves, abordando "el ciclo completo de desarrollo de aeronaves, desde los requerimientos del sistema hasta la verificación de los sistemas". En el marco de la observación anterior, el cumplimiento de estándares es insuficientes en el tratamiento de las características implícitas esperadas en el desarrollo de software profesional.

Se evidencia el escaso control y seguimiento de los defectos en los requerimientos a través del uso de indicadores de rendimiento y la exigua definición de objetivos a ser evaluados por dichos indicadores.

Para fortalecer el proceso de validación de requerimientos, como trabajo futuro se propone crear una metodología en base a los problemas mencionados anteriormente. La metodología, en primer lugar, debería describir un marco de referencia para el desarrollo de la estructura y en segundo lugar establecerá un modelo para el proceso de validación de requerimientos a nivel de caja blanca en la estructura interna del diseño de los componentes del desarrollo del sistema. Esto permitiría garantizar la calidad del producto o sistema, enfocándose en las diferentes etapas del ciclo de vida del software.

Referencias

1. P. A. Laplante: Requirements Engineering for Software and Systems, CRC Press (2019).
2. B. H. C. Cheng, J. M. Atlee: Current and Future Research Directions in Requirements Engineering, Design Requirements Engineering A Ten-Year Perspective, Lecture Notes in Business Information Processing, vol. 14, pp. 11–43 (2019).
3. G. Kotonya, I. Sommerville: Requirements Engineering: Processes and Techniques, JohnWiley & Sons, England, (1998).
4. S. L. Pfleeger: Software Engineering – Theory and Practice, Prentice Hall (1998).
5. G. Kotonya, I. Sommerville, Requirements Engineering: Processes and Techniques, John Wiley & Sons, (2000).
6. A.Terry Bahill, Steven J. Henderson: Requirements development, verification, and validation exhibited in famous failures, Systems Engineering. 8. 1 - 14. 10.1002/sys.20017 (2005).
7. P. Loucopoulos, V. Karakostas: System Requirements Engineering, McGraw-Hill, London, ISBN 0-07-707843-8 (1995).

8. Moketar Nor Aiza, Kamalrudin Massila, Mokhtar Mohd Yusof, Safiah Sidek: A review on requirements validation for software development, Journal of Theoretical and Applied Information Technology, (2018).
9. Atsushi Kokune, Masuhiro Mizuno, Kyoichi Kadoya, Shuichiro Yamamoto: FBCM: Strategy modeling method for the validation of software requirements, Journal of Systems and Software, Volume 80, Issue 3, Pages 314-327, (2007).
10. A. Cimatti, M. Roveri, A. Susi, S. Tonetta: From Informal Requirements to Property-Driven Formal Validation. In: Cofer D., Fantechi A. (eds) Formal Methods for Industrial Critical Systems. FMICS 2008. Lecture Notes in Computer Science, vol 5596. Springer, Berlin, Heidelberg, (2009).
11. M. D Sourour, N. Zarour: A methodology of Collaborative Requirements Validation in a cooperative environment, 10th International Symposium on Programming and Systems, Algiers, Algeria, pp. 140-147, (2011).
12. X. Fei, C. Bin, Z. Siming: A Methodology of Requirements Validation for Aviation System Development, Chinese Control And Decision Conference (CCDC), págs. 4484-4489, (2020).
13. P.S Seligmann, G.M Wijers,, H.G Sol: Analyzing the structure of IS methodologies, an alternative approach, Proceedings of the First Dutch Conference on Information Systems, Amersfoort, the Netherlands, (1989).
14. H.G Sol: A Feature Analysis of Information Systems Design Methodologies: Methodological Considerations, T.W. Olle, H.G. Sol, C.J. Tully (Eds.), Information Systems Design Methodologies: A Feature Analysis, North-Holland, Amsterdam, The Netherlands, (1983).
15. F. Kensing.: Towards Evaluation of Methods for Property Determination, The M.A. Bemelmans Ed., Beyond Productivity: Information Systems Development for Organizational Effectiveness, North-Holland, Amsterdam, The Netherlands, pp.325-338, (1984).
16. Wijers G.M., Heijes H.: Automated support of the modelling process: A view based on experiments with expert information engineers. Steinholtz B., Sølvberg A., Bergman L. (eds) Advanced Information Systems Engineering. CAiSE 1990. Lecture Notes in Computer Science, vol 436. Springer, Berlin, Heidelberg, (1990).
17. Pressman S.R.: Ingeniería del software. Un enfoque práctico, Séptima edición, México D.F., Mc Graw Hill, (2010).
18. IEEE Recommended Practice for Software Requirements Specifications, IEEE Std 830-1998 vol., no., pp.1-40, (1998).
19. J. N. Martin: Overview of the EIA 632 standard: processes for engineering a system, 17th DASC. AIAA/IEEE/SAE. Digital Avionics Systems Conference. Proceedings (Cat. No.98CH36267), pp. B32-1, (1998).
20. SAE international Group. APR4754A Guideline for Development of Civil Aircraft and Systems. vol. 4970, (2011).

ISBN 978 -987-633-574-4

Tecnología CASE para Modelado Específico de Dominio en Sistemas de Información Sanitaria basado en Estándar de Interoperabilidad Clínica

Juan Cesaretti[1], Lucas Paganini[1], Arián Calabrese[1], Martín Lunasco[1], Leandro Rocca [1], Leopoldo Nahuel[1], Roxana Giandini[1,2],

[1] GIDAS, Grupo de I&D Aplicado a Sistemas informáticos y computacionales, UTN - FRLP, La Plata, Argentina
[2] CIC, Centro de Investigación, LIFIA, UNLP - Facultad de Informática, La Plata, Argentina
{jcesaretti, lpaganini, acalabrese, mlunasco, leorocca, lnahuel, rgiandini}@frlp.utn.edu.ar

Abstract. Los sistemas de información sanitaria plantean dos grandes retos: por un lado, deben adaptarse a las constantes actualizaciones tecnológicas, y por otro, deben posibilitar la integración de toda la información y su disponibilidad en cada punto en que se necesite acceder a ella. La primera dificultad se abordó con el enfoque del Modelado Específico de Dominio (DSM). El poder de abstracción que provee el DSM permite a los ingenieros de software manejar la complejidad creciente de una manera rápida y clara. Por eso resulta beneficioso disponer de un Lenguaje Específico de Dominio (DSL) como el que aquí se propone: SIS_Static, complementado por un DSL dinámico: SIS_Dynamic. Para solucionar el segundo problema (comunicación entre distintos sistemas), se utilizó como referencia un estándar de interoperabilidad clínica: FHIR. Así, se implementó una herramienta de software basada en DSM que permite crear especificaciones gráficas de alto nivel y producir código fuente de manera automatizada, en distintos lenguajes de programación.

Keywords: Modelado específico de dominio (DSM), lenguaje específico de dominio (DSL), Fast healthcare interoperability resources (FHIR), desarrollo dirigido por modelos (MDD).

1 Introducción

La Ingeniería Dirigida por Modelos (MDE: Model-Driven Engineering) se enfoca fuertemente en los modelos, a los que considera los elementos centrales de la Ingeniería de Software.

Un modelo es una representación simplificada de la realidad, desacoplada de los detalles de implementación. Esto provee una gran adaptabilidad frente a la evolución de las tecnologías utilizadas. En particular, el Desarrollo de Software Dirigido por Modelos [1], [2] abarca todas las propuestas y mecanismos para producir aplicaciones a partir de la transformación de modelos, proporcionando un alto nivel de abstracción.

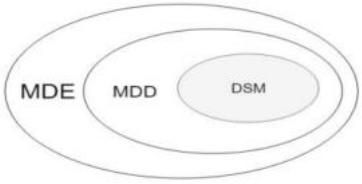


Fig. 1. Especialización de enfoques de la MDE

ISBN 978 -987-633-574-4

El Modelado Específico del Dominio o DSM (del inglés: Domain Specific Modeling), es una disciplina que, en el contexto del MDD, trabaja con lenguajes propios, restringidos a cada dominio o ámbito de interés. Son los llamados Lenguajes Específicos de Dominio, también conocidos por su acrónimo: DSL (Domain-Specific Language). Esta especialización permite una mayor automatización, que no podría lograrse usando un lenguaje de modelado de propósito general, como UML [17]. En la Fig. 1 se representa con un diagrama de Venn la relación de inclusión del enfoque DSM en el MDD, en el encuadre más general de la MDE.

Un modelo puede constituirse con diagramas, utilizando símbolos gráficos definidos en un lenguaje de modelado. En un DSL, cada bloque de construcción representa algún concepto propio del dominio considerado. Luego, este tipo de lenguaje es más claro y manejable para los usuarios, y las particularidades tecnológicas son transparentes para los mismos [1].

La sintaxis abstracta de un lenguaje gráfico de modelado se define en un metamodelo que especifica los elementos de modelado, las relaciones entre ellos y las reglas de buena formación de los modelos. Dado que un metamodelo es también un modelo, debe estar expresado en un lenguaje bien definido, conocido como metalenguaje. Este determina qué elementos pueden ser usados y cómo pueden vincularse. Algunos metalenguajes conocidos son: MOF (Meta-Object Facility) [3], ECORE [4] y GOPPRR [5].

La sintaxis concreta de un lenguaje gráfico determina el aspecto visual del mismo, estableciendo una colección de símbolos que pueden utilizarse para construir los diagramas.

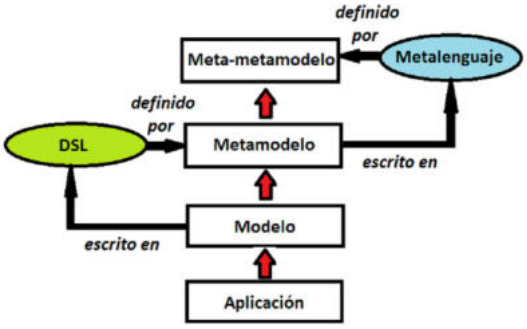


Fig. 2. Arquitectura de cuatro niveles de abstracción

La organización OMG propone una arquitectura de cuatro niveles, como muestra la Fig. 2:

Nivel M0 (Aplicación): Consta de los objetos instanciados por la aplicación.

Nivel M1 (Modelo): En el que se dibujan los diagramas para modelar el sistema.

Nivel M2 (Metamodelo): Es donde se definen los bloques de construcción (elementos y relaciones) de los lenguajes de modelado, utilizando un metalenguaje.

Nivel M3 (Meta-metamodelo): En el que se define la sintaxis abstracta de cada metalenguaje.

La solución que planteamos parte de dos vistas, una estática y otra dinámica. Se desarrolló un DSL para construir cada una de ellas: SIS_Static y SIS_Dynamic, respectivamente. Los metamodelos (nivel M2) en los que se sustentan estos lenguajes se expresaron en GOPPRR, que es el metalenguaje que emplea MetaEdit+ [6], la herramienta Meta CASE que utilizamos. La expresividad de los lenguajes que presentamos fue validada con distintos modelos (nivel M1), como el que se expone en la sección 4, a modo de ejemplo de aplicación.

Un dominio con una complejidad creciente es el de los Sistemas de Información Sanitarios (SIS). En los países avanzados, la esperanza de vida va en aumento. Y las personas mayores padecen distintas enfermedades crónicas, debido al sedentarismo, la obesidad y otros problemas causados por su estilo de vida. Se estima que el 17% de los pacientes en EE.UU. tienen más de seis afecciones crónicas. Así, una misma persona

ISBN 978 -987-633-574-4

consulta a distintos especialistas, y se plantea la necesidad de integrar toda la información registrada por ellos. Y esta información de atención médica debe ser accesible desde distintas organizaciones y puntos geográficos, en virtud de la movilidad de los pacientes [7]. Además, los SIS necesitan adecuarse constantemente a las nuevas tecnologías. Por lo tanto, el abordaje del DSM resulta muy conveniente en este caso.

Para intercambiar la información entre distintos sistemas de gestión sanitaria, existen estándares de interoperabilidad que definen la estructura de los datos, para que puedan compartirse [8]. Diversas organizaciones se ocupan de uniformar criterios de interoperabilidad, como ser: HL7 Internacional, HIMMS o NEMA.

El último estándar de interoperabilidad clínica de HL7 es FHIR (Fast Healthcare Interoperability Resources) [9]. Es de código abierto, y amalgama lo mejor de los estándares más utilizados en la actualidad. El elemento fundamental de FHIR es el "recurso", definido como la unidad básica de interoperabilidad. Cada recurso especifica un concepto del dominio sanitario: paciente, médico, problema de salud, entre otros.

Este trabajo presenta una solución DSM basada en FHIR. A través de ella, un ingeniero de software, aún sin tener un conocimiento previo de FHIR, puede crear diagramas compatibles con dicho estándar. Y esas especificaciones gráficas con alto nivel de abstracción, pueden ser transformadas en código de manera automatizada.

El presente artículo se organiza de la siguiente manera: en la Sección 2 se presentan la definición de los DSLs y cómo se construyó un editor basado en los mismos, en la Sección 3 se detalla la aplicación de la herramienta en un caso de estudio, en la Sección 4 se describen los antecedentes y los avances, y finalmente, en la Sección 5, se abordan las conclusiones y líneas de trabajo futuro.

2 Desarrollo

En esta sección se abordará el detalle de los dos DSLs diseñados cada una de las vistas construidas (estática y dinámica), el editor gráfico que se construyó como herramienta de modelado y los mecanismos utilizados para lograr generar código fuente a partir de modelos construidos con los DSLs.

2.1 Vista Estática

En primer lugar, se definió un DSL para modelar los aspectos estructurales de los sistemas de información sanitaria, al que denominamos SIS_Static. Sentamos sus bases en un estándar específico del dominio: FHIR. Del mismo, se seleccionó un subconjunto relevante de recursos. Cada uno de estos recursos fue tomado como un bloque de construcción del lenguaje de modelado SIS_Static, a saber:

Patient: Paciente, sujeto que recibe atención sanitaria.

Practitioner: Personal sanitario que provee servicios de atención médica, de enfermería u otras áreas afines.

Person: Abstracción que agrupa la información demográfica, tanto de pacientes como del personal sanitario.

Organization: Organización dentro de la que se proveen servicios sanitarios (hospital, clínica, sala de atención primaria, etc.).

Encounter: Encuentro entre un paciente y personal sanitario, por ejemplo: una práctica de salud (curación, rehabilitación, colocación de férulas y yesos, entre otras), una consulta ambulatoria, entre otras.

Episode of care: Episodio de atención, o asociación temporal entre una organización sanitaria responsable y un paciente, durante la cual pueden ocurrir uno o más encuentros.

ISBN 978 -987-633-574-4

Las propiedades de estos elementos también fueron tomadas de FHIR, al igual que la única relación válida entre dichos elementos: “Reference”. Asimismo, los nombres de los roles con que cada elemento participa en una relación se extrajeron del mismo estándar.

Este subconjunto de elementos escogidos permite modelar sistemas básicos de información sanitaria, para gestionar atenciones ambulatorias de demanda espontánea, servicios de guardia e internaciones no programadas. Y puede ser extendido fácilmente para aumentar el alcance de los sistemas modelados.

La Fig. 3 muestra el metamodelo en GOPPRR que define la sintaxis abstracta del lenguaje SIS_Static. La semántica de cada elemento es la misma que la especificada para los recursos correspondientes de FHIR.

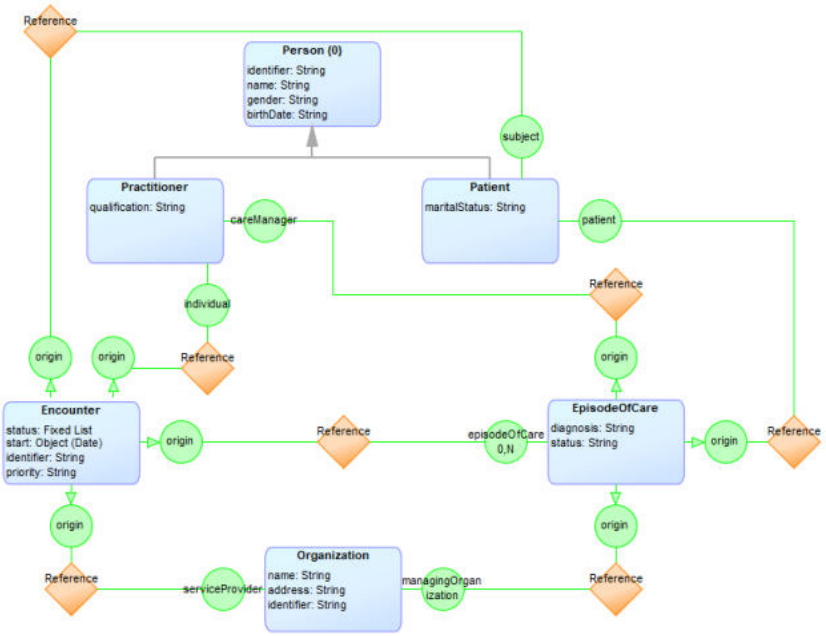


Fig. 3. Metamodelo en GOPPRR del DSL SIS_Static

2.2 Herramienta DSL

Para poder crear, visualizar y editar modelos, usando el lenguaje SIS_Static, se construyó un editor gráfico. Para ello, se utilizó el metaeditor MetaEdit+.

La sintaxis concreta del DSL se definió a través del editor de símbolos que está integrado a dicho metaeditor. La Fig. 4 presenta la paleta de elementos y relaciones del DSL SIS_Static. Los elementos pueden arrastrarse hasta el espacio de trabajo del diagrama, y pueden vincularse entre sí. La herramienta detecta e impide que se formen relaciones ilegales. Esto se debe a que las reglas de correctitud del dominio se materializaron por medio de “bindings”, que establecen las conexiones válidas entre elementos. El metaeditor también permite almacenar otras restricciones de conectividad, ocurrencia, unicidad, etc.

Fig. 4. Paleta de elementos y relaciones construídas para SIS_Static

ISBN 978 -987-633-574-4

2.3 Generación automática de código

Para transformar a texto (código fuente de un lenguaje de programación) los diagramas construidos con el DSL presentado, se utilizó el editor generado que MetaEdit+ posee integrado. El mismo cuenta con un lenguaje propio: MERL (MetaEdit Report Language). Este permite recorrer los diagramas, y producir una salida de texto, a partir de la información extraída de los elementos, sus propiedades, sus roles y relaciones por las que va navegando.

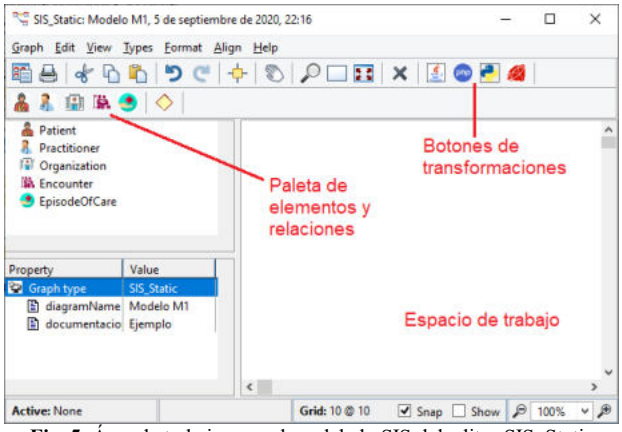


Fig. 5. Área de trabajo para el modelado SIS del editor SIS_Static

En la Fig. 5 se muestra la ventana principal del editor, con la paleta de elementos y relaciones, y los botones añadidos a la barra de herramientas para generar código automáticamente en distintos lenguajes de programación, como: Java, Php, Python y Ruby.

Aprovechando el generador integrado que posee MetaEdit+, se realizaron transformaciones de modelo a texto (ver Fig. 6) para producir código automáticamente, con sólo presionar un botón de la barra de herramientas (Botones de transformaciones en Fig. 5). Esto fue posible gracias a un lenguaje propio del metaeditor: MERL (MetaEdit Report Language), que permite navegar por los elementos de los diagramas, generando salidas con distintos formatos.

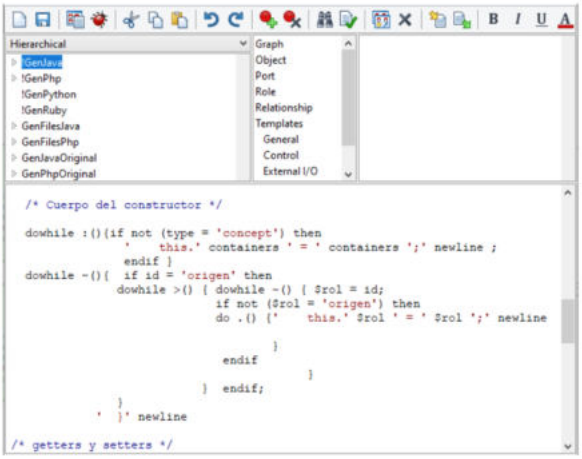


Fig. 6. Generador de código de MetaEdit+ a partir de modelos construidos con el editor SIS_Static

ISBN 978 -987-633-574-4

En la Fig. 7 se muestra un archivo de salida, obtenido a partir de un diagrama.

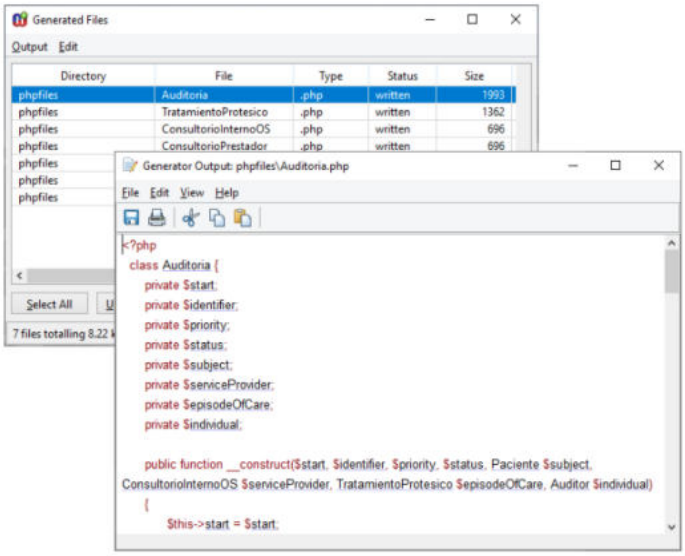


Fig. 7. Código fuente generado a partir de modelos gráficos construidos con el editor SIS_Static

2.4 Vista dinámica

Por último, se definió un DSL para capturar aspectos de comportamiento del dominio de los sistemas de información sanitaria. Lo denominamos SIS_Dynamic, y está basado en el enfoque de las máquinas de estado de UML [17]. Permite representar los diferentes estados por los que puede transitar un "Encounter" (Encuentro entre un paciente y personal sanitario), que es el elemento neurálgico de los modelos de este dominio acotado. Dichos estados se restringieron a una lista de valores posibles, que se halla especificada en FHIR.

Los elementos del lenguaje SIS_Dynamic son:

State [Encounter]: Se trata de cada uno de los estados por los que puede pasar un encuentro (planned, arrived, triaged, in-progress, onleave, cancelled, finished, entered-in-error y unknown). La herramienta solamente permite seleccionar uno de esos valores, definidos en FHIR. Y se agregó una propiedad "name", para describir el estado en lenguaje natural, para que sea más intuitivo.

InitialState: Es el estado temporal de inicio. Al igual que en UML, se restringe a una sola instancia por diagrama.

FinalState: Es el estado final, que cierra el ciclo de vida.

Action: Es una acción desencadenada al producirse la transición a la que se asocia. De ella puede extraerse información para agregar funcionalidad al código generado a partir de SIS_Static.

ISBN 978 -987-633-574-4

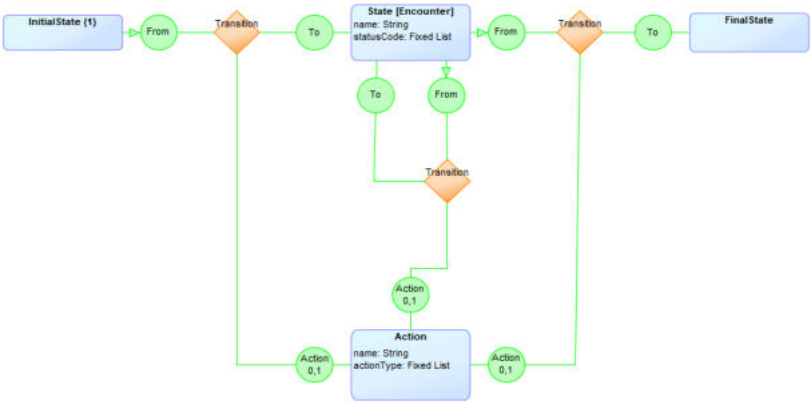


Fig. 8. Metamodelo en GOPPRR del DSL SIS_Dynamic

La única relación válida entre estos elementos es: "Transition" (transición). Esta representa el pasaje de un estado a otro.

La Fig. 8 muestra el metamodelo de SIS_Dynamic en GOPPRR, que define su sintaxis abstracta.

Del mismo modo que se hizo con la vista estática, se construyó un editor gráfico en MetaEdit+. La Fig. 9 presenta la paleta correspondiente de elementos y relaciones.

Fig. 9. Paleta de elementos y relaciones de SIS_Dynamic

Los metamodelos de los DSL SIS_Static y SIS_Dynamic se vincularon mediante una estructura de explosión. La Fig. 10 muestra cómo un tipo de objeto Encounter, componente de un diagrama estático en SIS_Static, puede ser refinado en otro diagrama más específico: un diagrama dinámico en SIS_Dynamic, que detalla sus cambios de estado y su transición.

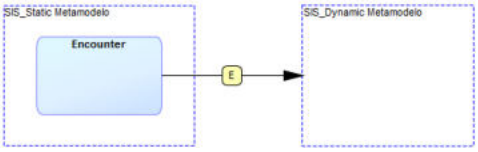


Fig. 10. Relación entre la vista estática y la vista dinámica

3 Resultados obtenidos a partir de un caso de estudio

Para probar el poder expresivo y la suficiencia de la solución DSM propuesta, se modeló la atención de pacientes en una Unidad Febril de Urgencia (UFU). Estas unidades son espacios anexados por el Ministerio de Salud de la ciudad de Buenos Aires a gran parte de los hospitales de agudos y pediátricos, en el contexto de la pandemia de COVID-19 [19].

Cuando una persona, presuntamente contagiada de COVID-19, se presenta en una UFU, primero es atendido por un enfermero. Si este determina que se trata de una emergencia, el paciente se remite directamente a la guardia. Si no, luego de llevar a cabo el protocolo de triage, el enfermero lo deriva a un médico. Este último lo evalúa de nuevo, y si resuelve que es una emergencia, lo envía a la guardia. Si no, le realiza un examen médico y epidemiológico, y registra los resultados en un informe. Luego es derivado a una Unidad Transitoria de Aislamiento (UTA), donde aguarda de forma segura hasta ser trasladado a un

ISBN 978 -987-633-574-4

hotel, una sala de hospital, o a una Unidad de Tratamientos Intensivos (UTI), según la gravedad del cuadro de la persona.

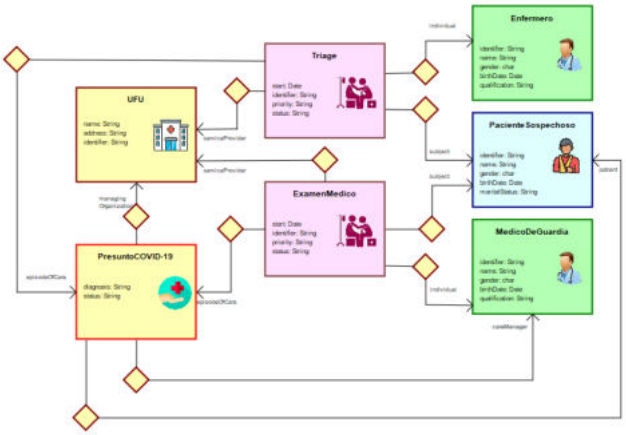


Fig. 11. Diagrama estático de Atención en UFU realizado con el DSL SIS_Static

La Fig. 11 muestra el diagrama estático, realizado con el lenguaje SIS_Static. El triage llevado a cabo por un enfermero, y el examen realizado por un médico, se representan con dos elementos de tipo Encounter. Ambos están referidos a un mismo Episode of care, que es la presunción de contagio del virus.

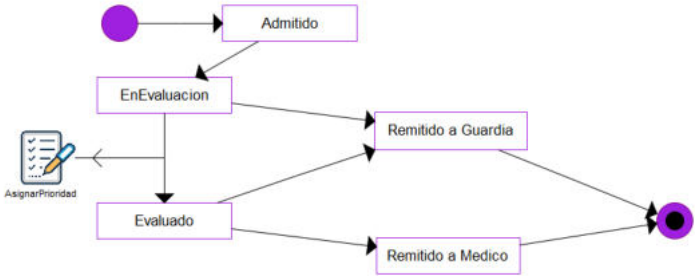


Fig. 12. Diagrama dinámico del Triage en una UFU realizado con el DSL

En la Fig. 12 se presenta un diagrama dinámico correspondiente al Encounter del Triage efectuado por un enfermero.

Los diagramas y archivos de código fuente generados automáticamente, para éste y otros casos de estudio, se encuentran disponibles en un repositorio [20]. Para conocer más detalles de la solución y tecnología utilizada, se ofrece un video demostrativo [21].

4 Discusión

Existe un metamodelo de HL7, llamado MIF (Model Interchange Format) [10], el cual es muy extenso y complejo, y cuyo aprendizaje significa un costo muy alto para los ingenieros de software. Además, MIF tiene múltiples versiones coexistentes, y esto puede causar problemas de incompatibilidad [11]. Por otra parte, algunos investigadores de HL7 Internacional han publicado un perfil de UML ajustado a MIF [12]. No obstante, el propósito de este trabajo es presentar una solución DSM, con un metamodelo más simple, fundado en un subconjunto relevante de recursos de FHIR. Se trata de una vista estática complementada con otra dinámica, que permite adicionar funcionalidad a los sistemas de información modelados.

Se han publicado experiencias previas en el uso de HL7 en el marco de MDE [13], [14]. Pero en ninguna de ellas se produjo una herramienta para automatizar la generación de código fuente. Esto se debe a que su propósito fue extender o especializar un lenguaje de propósito general como UML, o lograr transformaciones de modelos entre HL7 y UML. Con nuestra solución DSM, es posible esta automatización, siendo uno de los aspectos más significativos de esta propuesta.

En trabajos previos [15], [16], se desarrolló un DSL acotado para este dominio, a través de un metamodelo construido con el metalenguaje ECORE, reutilizando algunos elementos del metamodelo de UML [17]. Se implementó un editor gráfico basado en el DSL propuesto. Para ello se utilizó Sirius [18], un framework de código abierto, que aprovecha las tecnologías EMF y GMF de la plataforma Eclipse. Este editor solamente permitía crear, visualizar y modificar diagramas, respetando la sintaxis del DSL, pero no contaba con la posibilidad de realizar transformaciones de modelo a código fuente.

5 Conclusiones

En primer lugar, destacamos las fortalezas de la solución DSM propuesta. Se probó la expresividad del editor construido, modelando algunos casos concretos de situaciones reales y de interés actual por la pandemia Covid-19. Uno de esos casos (atención de pacientes en Unidad Febril de Urgencia en contexto de evaluación de contagios Covid-19) fue presentado en la sección anterior, mostrando vistas: estática y dinámica. Además, se realizaron transformaciones de modelo a texto, para obtener código en diferentes lenguajes de programación, de manera automatizada.

En cuanto al DSL SIS_Static, podemos observar que ofrece un mecanismo de abstracción muy adecuado, para manejar la complejidad del dominio. Es muy útil para simplificar el trabajo de los analistas de negocios y diseñadores de sistemas, en las primeras fases del proceso de desarrollo del software. Cada elemento de la paleta de edición se identifica con algún objeto reconocible del dominio, y es además una unidad de interoperabilidad.

Otro aspecto importante es que el editor verifica automáticamente los modelos construidos y detecta tempranamente posibles errores en el proceso de modelado, evitando la creación de especificaciones ilegales o no deseadas (según las reglas de negocio del dominio). Esto último constituye un aporte significativo en automatizar las reglas de buena formación de relaciones válidas entre los elementos del dominio.

El DSL SIS_Dynamic, por su parte, se adicionó para capturar el comportamiento funcional de los sistemas de información, permitiendo refinar elementos definidos en la vista estática, aumentando así las capacidades de la herramienta propuesta.

También es importante subrayar las ventajas observadas en el metalenguaje y el metaeditor utilizados. GOPPRR, al ser diseñado específicamente para describir lenguajes de modelado, aporta los mismos beneficios que usar DSM en cualquier dominio, contribuyendo principalmente en la sencillez y precisión. Y en MetaEdit+ se observan ventajas comparativas, con respecto a otros modeladores. Comparándolo con las herramientas utilizadas anteriormente en la plataforma Eclipse, se puede afirmar que el tiempo invertido para construir el editor es notablemente menor en MetaEdit+. Además, su entorno es mucho más intuitivo y simple. Ofrece un buen soporte a la evolución del lenguaje, facilitando la propagación de cambios del metamodelo a sus instancias. Y posee un editor de símbolos y un generador de código, ambos integrados en la herramienta, lo que resulta de suma utilidad para el diseño e implementación de un DSM.

Existen antecedentes de proyectos que abordaron el mismo dominio desde la perspectiva de MDE, tomando como base algún estándar de HL7. Sin embargo, ninguno de ellos se realizó en el marco del DSM, para explotar los beneficios que ofrece frente a un proceso tradicional de desarrollo de sistemas informáticos: ocultamiento de la complejidad y automatización.

Como trabajo futuro, dando continuidad a este proyecto, se plantea adicionar a las vistas estática y dinámica, una vista de interfaz gráfica de usuario (DSL SIS_GUI). También se procura agregar eventos a los diagramas dinámicos, y perfeccionar la definición de los elementos de tipo Action. Al mismo tiempo, se espera realizar un test de usabilidad y calidad interna del entorno construido: herramienta de software DSM para SIS.

Referencias

1. J. García Molina, F. O. García Rubio, V. Pelechano, A. Vallecillo, J. M. Vara y C. Vicente-Chicote, (2013). "Desarrollo de software dirigido por modelos: Conceptos, métodos y herramientas". Madrid, España: Ra-Ma
2. Kelly, S., Tolvanen, J., (2008). Domain-Specific Modeling: Enable Full Code Generation. Hobaken, Estados Unidos: Wiley-IEEE Computer Society
3. Especificación de MOF (MetaObject Facility). Recuperado de https://www.omg.org/mof/
4. Definición del Package ECORE. Recuperado de http://download.eclipse.org/modeling/emf/emf/javadoc/2.6.0/org/eclipse/emf/ecore/package-summary.html
5. Especificaciones del lenguaje de modelado de GOPPRR. Recuperado de https://www.metacase.com/support/45/manuals/mwb/Mw-1_1.html
6. MetaEdit+ Workbench – Build your own modeling tool. Recuperado de https://www.metacase.com/mwb/
7. M. L. Braunstein., (2018). "Health Care in the Age of Interoperability: The Potential and Challenges", IEEE Pulse, vol(9), 34-36. doi: 10.1109/MPUL.2018.2856941
8. Foro de la OMS sobre la Estandarización y la Interoperabilidad de los Datos Sanitarios. (2012). Organización Mundial de la Salud. Ginebra, Siuza.
9. Resourcelist – FHIR v4.0. Recuperado de https://hl7.org/fhir/resourcelist.html
10. Spronk, R., Ringholm, C. (2010). The HL7 MIF - Model Interchange Format. Recuperado de http://www.ringholm.com/docs/03060_en_HL7_MIF.htm
11. Villegas, A., Olivé, A. (2013). UML Profile for MIF Static Models. Version 1.0. Recuperado de http://www.vico.org/HL7_Tooling/Submission/MIF.pdf
12. Model Interchange Format, HL7. Recuperado de https://wiki.hl7.org/index.php?title=Model_Interchange_Format
13. E. R. Pfaff et al. (2019). "Fast Healthcare Interoperability Resources (FHIR) as a Meta Model to Integrate Common Data Models: Development of a Tool and Quantitative Validation Study". JMIR medical informatics, vol. (7) doi:10.2196/15199
14. M. A. Olivero, F. J. Domínguez-Mayo, C. L. Parra-Calderón, M. J. Escalona, y A. Martínez-García. (2020). "Facilitating the design of HL7 domain models through a model-driven solution", BMC medical informatics and decision making, vol. (20). doi:10.1186/s12911-020-1093-4
15. Cesaretti, J., Paganini, L., Rocca, L., Caputti, M., Zugnoni, I. (2019). Herramienta basada en Lenguaje Específico de Dominio para Sistemas elementales de Información Sanitaria. JAIIO 48. Salta, Argentina.
16. Rocca, L., Caputti, M., Zugnoni, I., Paganini, L., Cesaretti, J., Nahuel, L., Giandini, R. (2018). Marco de trabajo para el diseño y desarrollo de Herramientas de modelado conceptual basado en DSL utilizando tecnologías GMF. CIITI. Buenos Aires, Argentina.
17. OMG Unified Modeling Language (OMG UML). Version 2.5.1. December 2017.Recuperado de https://www.omg.org/spec/UML/2.5.1/PDF
18. Sirius – The easiest way to get your own Modeling Tool. Recuperado de https://www.eclipse.org/sirius/
19. Unidades Febriles de Urgencia (UFU). Recuperado de https://www.buenosaires.gob.ar/coronavirus/unidades-febriles-de-urgencia-ufu
20. Repositorio de modelos y código generado para casos de estudio. Recuperado de https://drive.google.com/drive/folders/1_9pxCOqDRH8qtxNVAKuyLlir_-R8zpGG
21. DSL-SIS Un Lenguaje de Modelado Especifico del Dominio de Sistemas de Información Sanitaria. Recuperado de https://youtu.be/WTs6f_ZNqf8

Refining a Software System Deployment Process Model: A Case Study

Marisa Panizzi[1,2,3 *], Marcela Genero[4], Rodolfo Bertone[1]

[1] School of Computer Science, Computer Science Research Institute LIDI (III-LIDI), Universidad Nacional de La Plata, La Plata, Argentina
marisapanizzi@outlook.com; pbertone@lidi.info.unlp.edu.ar
[2] Department of Engineering Information Systems, Universidad Tecnológica Nacional – Facultad Regional Buenos Aires, Argentina.
[3] Institute of Technology and Engineering, Universidad Nacional de Hurlingham, Hurlingham, Argentina
[4] Department of Technologies and Information Systems, University of Castilla-La Mancha, Ciudad Real, Spain
marcela.genero@uclm.es

Abstract. Software system deployment describes the activities associated with ensuring that a software system is available for its end users. Every company, regardless of its size, requires an efficient and effective software system deployment process to ensure the customer will accept the system software successfully. Small and Medium Enterprises (SMEs) often operate on limited resources and with strict time constraints, and need to improve their processes. For this reason, the existing proposals for deployment processes are not usually useful for SMEs. This fact led us to propose DepProMod (Deployment Process Model) to help SMEs to execute the deployment process of software systems in a systematized and controlled manner. The initial version of DepProMod has subprocesses, activities and tasks defined in addition to a capability-level architecture which allow its implementation in a step-by-step manner. This paper presents the results of a case study we carried out in order to examine the feasibility of the implementation of the initial version of DepProMod in a real environment with the purpose of refining it (if necessary) and completing it. We worked with the deployment process documentation of the "Company creation" module of a management system of advertising agencies for Latin America, in a software development SME in Argentina, to analyze the information requirements of the deployment procesgs and thus move towards the design of templates. In addition, a set of good practice recommendations was designed, not only for the deployment process but also for the rest of the company's software processes.

Keywords: Software Processes, Software System Deployment Process Model, Case Study.

1 Introduction

Small an Medium Enterprises (SMEs) need efficient and effective software engineering solutions. But the proper implementation of software engineering techniques is a difficult task for SMEs as they often operate on limited resources and with strict time constraints [1]. For this type of organizations, it is crucial to improve their processes and work methods because they account for the highest percentage of software development companies all over the world [2].

* Corresponding author: Marisa Panizzi

Deployment is a crucial process of the software development life cycle because its result will determine whether the client successfully accepts, or not, the software system that has been delivered. There are automation solutions to improve the last stages of the life cycle [3], among which we can mention new techniques/practices such as DevOps [4] and Continuous Deployment [5] in the context of agile methodologies. Google, Amazon, Netflix, LinkedIn, Facebook, and Spotify are some examples of successful companies whose DevOps practices have been reported and disclosed in IT books, blogs and events [6]. These emerging solutions are not viable for a large number of SMEs due to the lack of human resources and infrastructure that would allow them to adopt such solutions.

Before starting the design of the model, a systematic mapping study of the literature (SMS) was perfomed in order to review the state of the art and identify models, methodologies or methods which might serve as a guide for SMEs when deploying software systems [7]. As a result of the SMS, two process models were identified which could guide SMEs during the deployment process. Such models have the limitation that they delegate the responsibility of making decisions on a series of deployment-related aspects to the organizations that apply them. These aspects include tasks, artifacts, techniques, methods, tools and role definitions. This delegated responsibility potentially hinders the application of these models in SMEs since this type of organizations require more detailed and descriptive processes and, therefore, more easily appliable. In order to supplement the SMS, with the purpose of gathering evidence on the current state of the software system deloployment process practice in SMEs in Argentina, a survey-based exploratory study was conducted [8]. The results of the survey confirmed the need for a software system deployment process model which helps SMEs to conduct deployments in a systematized manner by means of: a) the execution of well-defined activities and tasks, b) the use of guiding templates, c) the assignment of specific roles which possess the necessary competences to execute the deployment, and d) the use of tools to automate some of the process activities in order speed up the process.

All of the above considerations led us to define the objective of our long-term research, which is to propose a holistic software system deployment process model to help SMEs execute the deploymemt process of software systems in a systematized and controlled manner. Our preliminary version of the model was called Model of a Computer Systems Implantation Process (MoProIMP) [9], but since it was not compatible with the international terminology or with the methodologies that refer to this phase of the software development life cycle, we decided to rename it to DepProMod (Deployment Process Model) and this acronym will be used hereinafter for the entire paper. DepProMod was developed to respond to the software system deployment process problem in SMEs in Argentina, although the feasibility of extending it to the international context will be studied later.

The preliminary version of DepProMod has subprocesses, activities and tasks. Our model differs from the existing proposals in its way of application; it allows SMEs to execute it in a step-by-step manner because its architecture is based on the capability levels of the CMMI-DEV standard [10]. The advantage of this application modality is the increased quality of the deployment process as well as the growth and improvement of the knowledge of the human resources of the SMEs. Since the DepProMod structure includes subprocesses, activities and management tasks for the

deployment process based on [11], it provides SMEs with more stabilized work methods. Another advantage is that it can be coupled with the development methodology used by the SME.

This paper presents the results of a case study we carried out in order to examine the feasibility of the implementation of the preliminary version of DepProMod in a real environment with the purpose of refining it (if necessary) and completing it. We worked with the deployment process documentation of the "Company creation" module of a management system of advertising agencies for Latin America, in a software development SME in Argentina to analyze the information requirements of the deployment process and thus move towards the design of templates.

The paper is organized as follows: Section 2 presents an overview of the DepProMod. The case study design and results are presented in Section 3 and, finally, our conclusions and proposals for future work are set out in Section 4.

2 Overview of DepProMod

The preliminary version of DepProMod has a life cycle model that adopts the 5 PMIBOK process groups [11]. These groups are: Initiating, Planning, Executing, Monitoring and Controlling and Closing. The reason for this choice is that PMIBOK is a globally recognized standard for use in the software industry. Each of these processes in DepProMod is called a "*subprocess*".

For the definition of the activities of DepProMod, a set of processes of the ISO / IEC / IEEE 12207 standard [12] were considered. The processes extracted from the standard are the technical management processes: risk management, configuration management, project management, and other technical processes: verification and validation. In our model, these processes are called "*activities*". At the "*tasks*" level, the model adopts a group of tasks proposed in the Metrica v3 [13] methodology as it is considered one of the most complete methodologies at the level of the tasks that are executed in the deployment process and those used in Spain and Latin America. In addition, a series of activities proposed in the "transition" technical process of the ISO / IEC / IEEE 12207 standard [12] were considered.

In order to implement the model in a step-by-step manner, three of the capability levels were adopted from the CMMI-DEV standard [10]. These levels are: level 1 = Done, level 2 = Managed and level 3 = Defined. Level 0 = Incomplete was not considered since it means the non-completion or partial completion of that process in the organization. These levels were analyzed and defined at a granularity level of the tasks considered in the model. The choice to consider capability levels rather than maturity levels is due to the fact that not all software development companies have reached maturity levels 4 and 5. This tiered architecture offers the advantage that software development companies can implement it in a step-by-step manner and, as they manage to stabilize the process at one level and achieve the necessary knowledge for its implementation, they can scale it to the next level.

The process pattern used for the representation model is the one proposed in the Competisoft model [14], since it is a process improvement model for Ibero American software industry SMEs with some adaptations to the needs of the DepProMod definition.

3 Case Study Description

In this section, we present the detailed description of the case study, following the guidelines proposed in [15, 16].

3.1 Case study design and research questions

The main goal of our case study is to examine the feasibility of the implementation of the DepProMod preliminary version in a real environment with the purpose of refining it (if necessary) and completing it. This case study is of an exploratory type [16] because it makes it possible to find out what is happening in the deployment process, seeking new points of view and generating ideas and hypotheses for our research. We worked with the documentation of the deployment process of the "Company creation" module of a management system for advertising agencies for Latin America to analyze the information requirements for the software system deployment process and thus move forward towards the design of the templates necessary for our model. We believe the case study is suitable because we wish to find the information requirements of the software system deployment process.

To achieve our goal, we posed the following research questions (RQ):

RQ1: Is it necessary to refine the model to adapt it to the existing needs in the industrial context?

Through this question, we sought to obtain the information needs for the execution of the tasks carried out by the consulting company in the deployment process to compare them with our model in order to refine it and complete it.

RQ2: Was the implementation of the model useful for the company?

With this question we tried to determine how the consulting company can strengthen its software system deployment process. For this purpose, we will provide a set of specific recommendations for the process as well as Software Engineering practices in general.

This a single embedded case study [16] according the classification of Yin (2002) with the following characteristics:

- Context: although our model arose in response to the need for SMEs to improve and stabilize their software system deployment process, the case study that we had available involved an SME (55 employees), located in Argentina, which offers consulting products and services. This company uses a development methodology with an iterative-incremental life cycle model, with the conventional stages: Analysis and Design, Construction, Testing and Implementation. In each stage, product/s-artifact/s are built to continue with the next stage. They also incorporate some practices of agile methodologies such as extreme programming (XP), pair programming. The first author of this work had access to the company's facilities and project documentation subject to an agreement not to disclose the company's name as well as a commitment to inform about the findings and recommendations to be considered.
- Case: deployment of the "Company creation" module (in a new country) of the management system of advertising agencies for Latin America.This module corresponds to a management system called "T&C" that has the following

modules: customers, suppliers, accounting, treasury, administration and parameters (module where master entities are created and the system is configured), expense reports and security. The module “Company creation” of the T&C management system that was implemented contains the following global features: creation of the company in the system and preparation of initial information and parameters to operate it. The features in detail are: upload the general data of the company, enter the provinces or states, create the divisions, upload the people master file, upload the suppliers master file, upload the clients / brands / products / projects master file, set up holidays, set up working days, create departments, create hierarchies, create work groups, assign modules to groups, create and assign administrative functions to people, assign people to work groups, assign menu options to the employees, assign clients to work groups, fill the parameters module with information, enter the accounts plan with its respective additional information.

- Unit of analysis: deployment documentation of the "Company creation" module of the advertising agency management system called “T&C”.

3.2 Preparation for data collection

The third grade collection technique was used according to the classification proposed in [16]. Qualitative data were collected from the documentation used in the deployment of the "Company creation" module of the T&C management system, which was obtained from different sources and / or repositories of the project.

In order to facilitate the preparation of the documentation to be collected, a data collection template was defined with a coding scheme according to the template approach mentioned in [16]. The template coding scheme is made up of a set of 5 groups, each of which coincide with the 5 subprocesses of DepProMod (Initiating, Planning, Executing, Monitoring and Controlling, and Closing).

For each group, a series of categories and their description were defined. In Table 1, an extract of the coding scheme is presented. The rest of the coding scheme used is presented in the Appendix [17].

Table 1. Extract of the coding scheme for data collection.

Group Code	Category Name	Description
S1 – Initiating subprocess	PRO: Project	The project plan, software requirements and software architecture are explored.
	ORG: Organization	The organization’s communication aspects, documentation protocols and configuration management handling are explored.
	RES: Resources	The organization’s human resources, the users and technological resources are identified.

3.3 Analysis and interpretation of results

Since this is an exploratory study, the "Hypothesis Generation" technique was used to analyze the data [16]. In this case study, we consider that the research could be verifying the following:

- Based on knowing the information required in the software system deployment process in a real context, DepProMod is refined and completed with the definition of templates for its tasks and,
- It is possible that from the analyzed documentation, the company is provided with a set of recommendations of good practices to improve its deployment process.

In a first instance, the drawing of conclusions from the collected data was carried out by the first author as part of the research process of the doctoral thesis and then agreed with the other authors, the thesis supervisors.

Two columns were added to the template designed to collect the study data. The first, called "comments", was used to record additional information in the analyzed document. The second column was called "recommendations" and was used to record recommendations for the deployment process analyzed (of the case). The information collected and analyzed is presented in the Appendix [17].

Within the reviewed documentation, the content of the emails found in the Incident Follow-up System (IFS) was also analyzed, since this allowed the acquisition of information on relevant milestones of the project.

In total, twenty one documents were analyzed. The review was developed in a systematic way, each document was associated with the defined coding, seeking traceability of its use in the different groups defined in the coding. Each group corresponded to the suprocess defined in our model and each category corresponded to an aspect to consider in its subprocesses, such as: aspects of the project, the organization, etc. This method of analysis allowed us to contrast the information needs of a real case with our model and simultaneously reflect on good practices to recommend to the consulting company.

3. 4 Results

Table 2 shows the traceability of the documents reviewed for each DepProMod subprocess.

Table 2. Traceability of the documents reviewed for each DepProMod subprocess.

Documents/ Subprocesses	Subprocess 1: Inttiating	Subprocess 2: Planning	Subprocess 3: Executing	Subprocess 4: Monitoring and Controlling	Subprocess 5: Closing
T&C Project	x				
General documentation	x	x			
User requirements	x				

Requirements procedure	x				
Project standards	x				
Work plan		x	x	x	x
Requirements for the installation site		x	x		
Instructions to structure the submission		x			
Installation test procedure		x			
Acceptance test procedure		x			
User's manual			x		
Smoke test instructions			x		
Data entry instructions			x		
Acceptance test instructions			x		
New company application form			x		
Installation script			x		
Progress report			x		
Meeting memo				x	
Smoke test results			x		
Acceptance test results			x		
Installation completion report					x

The results related to the research questions formulates for this case study are as follows:

RQ1: Is it necessary to refine the model to adapt it to the existing needs in the industrial context?

Based on the documentation analyzed, a series of requirements were obtained to complete the definition of DepProMod, which are presented below according to the subprocess structure:

Subprocess 1: Initiating. Five documents were reviewed. There was incomplete or inaccurate information which made it impossible to associate it with the deployment tasks. From this analysis, we consider that, in our model, it is necessary to design templates that allow the information to be documented to be unified, with a clear objective of use, distribution and the definition of a person responsible for its creation, modification and approval.

Subprocess 2: Planning. Six documents with the information related to this subprocess were reviewed. There was information that could not be analyzed either because it was not found or was incomplete. In the documentation reviewed, only the use of two metrics, time and effort, was found. These are considered in our model along with others, such as productivity and error rate of installation tests. In contrast to our model, it was not possible to obtain new information because DepProMod will contemplate more specific metrics.

Subprocess 3: Executing. Eleven documents with information related to this subprocess were reviewed. There was no information related to data migration because it was the deployment of a new system module. For this subprocess, the model is enhanced by building the following templates in the previous subprocess (planning) which will be used in this subprocess: "deployment strategy", "guide for site preparation", "installation guide", "data migration", "data upload", "test specifications", "user acceptance testing", "required human resources", "required

technological resources", "competencies of the technical team", "users to be trained", "metrics", "measurement report", "deployment risks" and "contingency plan". Within this process, the following templates will be designed: "end user assistance report", "technical team assistance report" and "activity report".

Subprocess 4: Monitoring and Controlling. Two documents with information related to this subprocess were reviewed. There was information that was not found or was insufficient to contrast with our model. There was documentation that reflected the monitoring of activities (work plan) and the meeting memo was also reviewed, which includes the decisions made by the project participants. There was no information regarding those who participated in the training activities (users, trainers and technicians). DepProMod will incorporate two templates that allow registering of the activities carried out as part of the "activity report" deployment that will be shared with the client and the information from the "report of risks occurred" and the "measurement report" is updated.

Subprocess 5: Closing. The two reviewed documents containing information related to this subprocess, one of them is the updated deployment plan and the other contains information of the installation activities. There was no evidence of the closing of the training activities, the closing of the deployment team or the learned lessons. DepProMod proposes " acceptance document", "closing report" and to register lessons learned in a a knowledge base.

RQ2: *Was the implementation of the model useful for the company?*

The company found the DepProMod implementation useful since we provided a report with a set of recommendations to improve its deployment process for future projects as well as suggestions for good Software Engineering practices in general. These recommendations can be listed as follows:

- Use appropriate tools for the administration of the project plan, since the project plan was managed with Excel.
- Analyze the deployment process strategy through a feasibility study.
- Expand the definition of metrics for the deployment process as well as for the rest of the software development processes since the only metrics they use are time and effort.
- Define risk management and its mitigation procedure.
- Effectively delegate the activities to be carried out by the client, since the preparation of the installation site was carried out by the client without adequate supervision by the consulting company.
- Create an institutional space to share knowledge not only regarding the deployment process, but also the rest of the processes of the software development life cycle.

3.5 Threats to validity

To analyze the validity of the study, the factors proposed in [16] were taken into account:

- Construct validity. Results were obtained in relation to the information needs of a deployment process in a real context, which allowed us to answer the defined research questions, determining their pertinence and suitability for the case.

- Internal validity. The documentation used belongs to a real case, a deployment of a module of an advertising agency management system (T&C). To achieve greater precision and validity of the studied process, the need to combine the data source (project documentation) with another type of source, such as interviews and / or focus group to ensure a "Data (Source) Triangulation", is recognized. Furthermore, the collected and analyzed qualitative data could be combined with quantitative data resulting from the project thus ensuring a "Methodological Triangulation".
- External validity. The use of a single case study may limit the generalization of the results. However, reporting on these first findings is considered necessary, as it serves as an incentive for other researchers to replicate our study in different case studies.
- Reliability. The study data were collected by a single researcher. Although they were analyzed with the thesis supervisors, this can be considered a threat to the research. To add a higher degree of reliability, it would be advisable for another researcher to apply the template with the coding created here in another case study.

4 Conclusions and Future work

This paper presented the results of a case study we carried out in order to examine the feasibility of the application of the initial version of DepProMod in a real environment with the purpose of refining it (if necessary) and completing it. After carrying out the case study, we can conclude that:

- RQ1 allowed us to identify the information (stated in section 3.6.) required in the deployment process in a real context and, given the diversity of the documentation structure, we consider it is necessary to create templates to complete our model. This will allow the information to be documented to be unified, with a clear objective of use, distribution and the definition of a person responsible for its creation, modification and approval.
- RQ2 allowed us to create a set of recommendations (presented in section 3.6.) for the company to improve its deployment process as well as to introduce good practices for the rest of the software project processes.

Our future work will consist of refining and completing the DepProMod in order to allow SMEs to systematize the deployment process of their software systems and provide detailed guidance on the subprocesses, activities, tasks, templates, roles, techniques/practices and tools and a definition of levels so that it can be implemented in stages. Our model can be coupled to the software development methodologies used by these SMEs. Next, we plan to carry out case studies in Argentine software development companies to test the usefulness of DepProMod, especially in SMEs.

Acknowledgements. The research work presented in this paper is framed within the following projects: GEMA (JCCM, SBPLY/17/180501/000293) and AETHER-UCLM (MICINN, PID2020-112540RB-C42).

References

1. Felderer M., Ramler R. Risk orientation in software testing processes of small and medium enterprises: an exploratory and comparative study. Software Quality, 24, pp. 519–548 (2016).
2. Mishra D., Mishra A. Software Process Improvement in SMEs: A Computer Science and Information Systems, 6, pp. 111 – 140 (2009).
3. Fuggetta A., Di Nitto, E. Software Process. In: Proceedings of the 36th International Conference on Software Engineering - Future of Software Engineering (FOSE'14), pp. 1-12 (2014).
4. Bass L., Weber I., Zhu L. DevOps: A Software Architect's Perspective (2015).
5. Scaled Agile. https://www.scaledagileframework.com/continuous-deployment. Last accessed 2020/03/03.
6. Díaz J., Pérez J., Yague A., Villegas García A., de Antona A. DevOps in Practice - A preliminary Analysis of two Multinational Companies. In: Proceeding of the 20 th Product-Focused Software Process Improvement. (PROFES 2019), pp. 323-330 (2019).
7. Panizzi M., Genero M., Bertone R. Software system deployment process: A systematic mapping study In: Proceedings of the XXIII Iberoamerican Conference on Software Engineering, CIbSE 2020, Curitiba, Paraná, Brazil, November 9-13, pp. 138-151 (2020).
8. Panizzi M., Genero M., Bertone R. Encuesta para analizar las necesidades con respecto al proceso de despliegue de las PyMES en Argentina. In: Proceedings of the XXIII Iberoamerican Conference on Software Engineering, CIbSE 2021, San José, Costa Rica, Agosto 30 – Setiembre 3. (2021).
9. Panizzi M., Bertone R., Hossian A. (2019) Proposal for a Model of a Computer Systems Implantation Process (MoProIMP). In: Pesado P., Aciti C. (eds) Computer Science – CACIC 2018. CACIC 2018. Communications in Computer and Information Science, vol 995. Pages 157-170. Springer, Cham. https://doi.org/10.1007/978-3-030-20787-8-11.
10. CMMI® Institute. CMMI Development V1.3 (2010).
11. A Guide to the Project Management Body of Knowledge. (PMIBOK® Guide) – Fifth Edition. Project Management Institute (2013).
12. IEEE ISO/IEC/IEEE 12207:2017(E). Systems and software engineering - Software life cycle processes (2017).
13. Portal de Administración Electrónica. Gobierno de España. Métrica versión.3. https://administracionelectronica.gob.es/pae_Home. Last accessed 2020/03/07 (2001).
14. Competisoft. Mejora de Procesos para Fomentar la Competitividad de la Pequeña y Mediana Industria del Software de Iberoamérica. Last accessed 2021/04/03. https://alarcos.esi.uclm.es/competisoft/web/completo/index.htm
15. Runeson P., Höst M. Guidelines for conducting and reporting case study research in software engineering. Empir Softw Eng 14:131–164 (2009).
16. Runeson P., Höst M., Rainer A., Regnell B. Case study research in software engineering: guidelines and examples. Wiley Publishing, Hoboken (2012).
17. Panizzi M., Genero M., Bertone. Appendix - Refining a Software System Deployment Process Model: A Case Study. https://doi.org/10.6084/m9.figshare.15000642.v1

ISBN 978–987–633–574–4

Modelado Conceptual de Juegos Serios: Revisión sistemática de la literatura

Andrés Daniel Chimuris Gimenez[1], Juan Cristian Daniel Miguel[1], Matias Leonel Bassi[1], Nicolás Matías Garrido[1], Gabriela Velazquez[1], Marisa Daniela Panizzi[1]

[1] Programa de Maestría en Ingeniería en Sistemas de Información. Escuela de Posgrado. Universidad Tecnológica Nacional. Regional Buenos Aires (UTN-FRBA). Medrano 951. (C1179AAQ). CABA, Argentina.
chimuris@gmail.com; juancristianmiguel@gmail.com; bassimatias@yahoo.com; garridonm@gmail.com; gav.sistemas@gmail.com; marisapanizzi@outlook.com

Resumen. Los Juegos Serios (Serious Games o SG) son todos aquellos cuyo objetivo no es, únicamente, promover un mero entretenimiento, sino también estimular el aprendizaje o la adquisición de un conocimiento o habilidad. Actualmente, hay una tendencia en el mercado a la generación de este tipo de juegos. Dada la importancia de la conceptualización del dominio de un problema y de su solución, en este trabajo se presenta el desarrollo de un mapeo sistemático de la literatura (en inglés, systematic mapping study o SMS) con el propósito de identificar el estado del arte y descubrir las contribuciones que existen en relación con el modelado conceptual de juegos serios. Se realizó una búsqueda en las librerías digitales Scopus, IEEE Xplore y ACM desde enero del año 2010 a junio del año 2021. De un total de 558 artículos encontrados, se analizaron 31 estudios primarios. Se evidenció que UML[1] es el lenguaje de modelado predominante para el modelado de Juegos serios, aunque se utilizan otros lenguajes como UP4EG, DSML, Deterministic Finite Automaton (DFA), Discrete Event System Specification (DEVS) y Fuzzy Inference Systems (FIS). Dentro de los diagramas UML, los predominantes son los diagramas de clases y diagramas de actividad. El 30% de los estudios primarios proponen un framework y en el mismo porcentaje (30 %) de los artículos propone una metodología para el desarrollo de Juegos serios. De los frameworks encontrados, la mayoría no especifican la manera para realizar el modelado conceptual.

Palabras clave: Modelado conceptual, Juegos serios, Mapeo sistemático de la literatura.

1 Introducción

Un modelo conceptual es una consolidación concisa y deliberada de un conjunto de conceptos que se presentan mediante términos en un formato lingüístico predefinido [1].

El modelado conceptual es una técnica de análisis de requisitos y de diseño de bases de datos. Como técnica de análisis de requisitos ayuda a identificar problemas en los requisitos antes de comenzar el desarrollo, evitando gastos innecesarios. Como técnica de diseño de bases de datos, permite representar de forma abstracta los conceptos y

[1] UML: Lenguaje unificado de modelado.

hechos relevantes del dominio del problema y transformarlos posteriormente en un esquema de una base de datos concreta [2].

El modelo del sistema es una conceptualización del dominio del problema y de su solución. El modelo se focaliza sobre el mundo real: identificando, clasificando y abstrayendo los elementos que constituyen el problema y organizándolos en una estructura formal. La abstracción es una de las principales técnicas con la que la mente humana se enfrenta a la complejidad. Ocultando lo que es irrelevante, un sistema complejo se puede reducir a algo comprensible y manejable. Cuando se trata de software, es sumamente útil abstraerse de los detalles tecnológicos de implementación y tratar con los conceptos del dominio de la forma más directa posible. De esta forma, el modelo de un sistema provee un medio de comunicación y negociación entre usuarios, analistas y desarrolladores que oculta o minimiza los aspectos relacionados con la tecnología de implementación [3].

Michel & Chen definen el término juego serio (JS) como una forma de combinar videojuegos y educación, donde el objetivo principal es la educación (en cualquiera de sus formas), y cuyas componentes principales son: objetivos, reglas, retos e interacción. Los JS habilitan otro mecanismo para llevar adelante la enseñanza y aprendizaje, a la vez que extiende los objetivos de entrenamiento y genera no solo condiciones para que el jugador (estudiante) aprenda, sino que además pueda aplicar y demostrar lo aprendido [4].

Los Juegos serios son aquellos cuyo principal objetivo no se centra en la diversión, sino en el aprendizaje o adquisición de un conocimiento o habilidad. Hoy en día son utilizados para la formación de conocimientos dentro del ámbito militar, político, empresarial, salud y educación. Este “concepto de juegos serios busca potenciar el aprendizaje, la estimulación del pensamiento crítico, el entrenamiento, la alfabetización digital, cambios de actitud y generación de emociones, lo cual va más allá del componente lúdico propio de los juegos” [5]. Cabe destacar que también “...se potencia el aprendizaje activo y se capacita en competencias complementarias como la toma de decisiones, el trabajo en equipo, habilidades sociales, liderazgo y colaboración…” [6].

Este artículo se desarrolla en el marco del Seminario de Modelado Conceptual de la Maestría en Ingeniería de Sistemas de Información de la Universidad Tecnológica Nacional, Regional Buenos Aires. La elección del tema ha sido motivada por los tópicos de interés del área de “Aplicaciones avanzadas y multidisciplinarias” propuestas en la 40 edición del Congreso Internacional de Modelado Conceptual (ER 2021) [7]. En este artículo se presenta un mapeo sistemático de la literatura (SMS) para analizar el estado del arte y descubrir las contribuciones que existen en relación con el modelado conceptual de juegos serios. Para realizar el SMS se siguieron los lineamientos propuestos por Kitchenham et al. [8] y por Petersen et al. [9].

El artículo se estructura de la siguiente manera: en la Sección 2 se describe la planificación del SMS, en la Sección 3 se describe su ejecución. Los resultados se presentan en la Sección 4. En la Sección 5 se presenta un análisis de las amenazas a la validez y, finalmente, en la Sección 6 se exponen las conclusiones.

2 Planificación del SMS

En esta sección se presenta la definición del protocolo del SMS: preguntas de investigación (PI), estrategia de búsqueda, selección de los estudios, criterios y proceso de selección, formulario de extracción y el proceso de síntesis de los datos.

El objetivo de este SMS es responder la siguiente pregunta de investigación (PI): *¿Qué características tiene el modelado conceptual de los Juegos Serios?* Esta pregunta principal se descompone en un conjunto de sub-preguntas (PI1-5), las cuales se presentan en la Tabla 1 junto con su motivación.

Tabla 1. Preguntas de investigación (PI) y motivación.

Pregunta (PI)	Motivación
PI1: ¿Qué contribuciones realiza respecto al modelado conceptual de los Juegos serios?	Encontrar y comprender qué tipo de aportes otorgan en cuanto modelado conceptual.
PI2: ¿En qué ámbitos se utilizan los Juegos serios?	Identificar el ámbito en los que se utilizan los Juegos serios
PI3: ¿Cuál es el Lenguaje de Modelado que se utiliza para proyectos de Juegos serios?	Determinar el lenguaje de modelado utilizado para afrontar el modelado de un juego serio.
PI4: ¿Qué diagramas se consideran para el modelado en proyectos de Juegos serios?	Identificar qué diagramas se utilizan para el modelado de un juego serio.
PI5: ¿Cuáles son los tipos de investigación encontrados en los artículos?	Identificar los tipos de investigación de los estudios de acuerdo con la clasificación propuesta por Wieringa [10].

Se decidió realizar una búsqueda automática en las librerías y plataformas digitales Scopus, IEEE Xplore y ACM por tratarse de las bibliotecas más utilizadas en el campo de la Ingeniería de software, considerando artículos de congresos y artículos de revistas. La búsqueda se realizó en el período comprendido entre enero del año 2010 hasta junio del año 2021.

Para el armado de las cadena de búsqueda se consideraron como términos principales “Serious Games” y “Conceptual modelling”, incluyendo sus términos alternativos. La cadena de búsqueda resultante es:

((“Serious game” OR “Serious games” OR “SG”) AND (“Concept” OR “Conceptual modeling” OR “conceptual modelling”))*

Los criterios de inclusión y exclusión utilizados para el proceso de selección de artículos se presentan en la Tabla 2.

El proceso de selección de los estudios consistió en los siguientes pasos: 1) realizar la búsqueda en las fuentes definidas aplicando la cadena en el título y/o en el resumen, 2) eliminar los artículos duplicados, 3) aplicar los criterios de inclusión y exclusión en el título, resumen y palabras clave, 4) aplicar los criterios de inclusión y exclusión al

texto completo. Este proceso permitió la selección de los estudios primarios que se analizaron para dar respuesta a las preguntas de investigación (PI) formuladas.

Para dar respuesta a cada una de las preguntas de investigación (PI) se definió un esquema de clasificación, que por restricciones de espacio se presenta en un apéndice [11], junto con el formulario de extracción de datos. Se utiliza una síntesis temática basada en el esquema de clasificación que se representará a través de tablas.

Tabla 2. Criterios de inclusión y exclusión.

Criterios de inclusión.
I1. Artículos duplicados: si hay varios artículos de un mismo autor que contemple la misma investigación, se considerará el más completo y el más reciente.
I2. Artículos en idioma inglés.
I3. Artículos publicados entre enero de 2010 y junio de 2021.
I4. Artículos que contengan cadenas candidatas en el título, palabras clave y/o en el resumen.
Criterios de exclusión.
E1. Artículos cuya óptica sea ajena al ámbito de software.
E2. Toda literatura gris, a saber: informes técnicos, tesis, presentaciones en power point, entre otros.
E3. Artículos a los cuales no se tenga acceso.
E4. Artículos cuyo contenido no se enfoque en el modelado conceptual.

3 Ejecución del SMS

En esta sección, se presenta la búsqueda realizada en las librerías y plataformas digitales, la selección de estudios primarios de acuerdo con lo definido en el protocolo de revisión del SMS.

Se aplicó la cadena de búsqueda en las librerías con algunas adecuaciones necesarias en función de las particularidades de cada una que se encuentran en [11].

De un total de 558 artículos encontrados, se analizaron 31 estudios primarios. El listado de los estudios analizados se presenta en [11].

4 Síntesis del SMS

En la Tabla 5 se presenta una síntesis de los resultados del análisis de los estudios primarios en base a lo establecido en el esquema de clasificación definido (Ver apéndice, Tabla 1). A continuación, se pretende dar respuesta a las preguntas de investigación en base al material recolectado.

Tabla 5. Síntesis de los resultados obtenidos.

Id	Resultados por cada PI				
	Contribución (PI1)	**Rubros (PI2)**	**Lenguaje de modelado (PI3)**	**Diagramas (PI4)**	**Tipos de Investigación (PI5)**
[EP1]	Metodología	Medicina	UML	Diagrama de clases	Propuesta de solución
[EP2]	Framework	Educación	No menciona	No menciona	Validación
[EP3]	Procedimiento, Técnica	Educación	No menciona	Otros	Experiencia personal
[EP4]	Metodología	Educación	No menciona	Otros	Experiencia personal
[EP5]	Metodología	Militar	Otros	Otros	Propuesta de solución
[EP6]	Metodología	No menciona	Otros	Otros	Propuesta de solución
[EP7]	Técnica	Ingeniería	No menciona	Otros	Validación
[EP8]	Framework	Otros	No menciona	Otros	Propuesta de solución
[EP9]	Framework	Educación	Otros	Diagrama de Actividad Otros	Experiencia personal
[EP10]	Metodología	Educación	UML	Diagrama de clases	Evaluación
[EP11]	Framework	Educación	UML	Otros	Evaluación
[EP12]	Otros	Medicina	UML	No menciona	Evaluación
[EP13]	Procedimiento	Medicina	No menciona	No menciona	Propuesta de solución
[EP14]	Herramienta	Otros	UML DSML	Diagrama de Dominio Diagrama de Actividad Diagrama de objetos	Evaluación
[EP15]	Otros	Ingeniería	Otros	Otros	Propuesta de solución
[EP16]	Procedimiento	Educación	No menciona	No menciona	Validación
[EP17]	Framework	Educación, Otros	UML DSML	Diagrama de clases	Propuesta de solución
[EP18]	Procedimiento	Educación	No menciona	No menciona	Propuesta de solución
[EP19]	Framework	Educación	UML Otros	Diagrama de clases	Propuesta de solución
[EP20]	Otros	No menciona	No menciona	No menciona	Validación
[EP21]	Metodología	Educación	No menciona	No menciona	Evaluación
[EP22]	Metodología	Educación	No menciona	No menciona	Evaluación
[EP23]	Metodología	Educación	No menciona	No menciona	Evaluación

[EP24]	Metodología	No menciona	UML	Diagrama de Dominio Diagrama de Actividad	Evaluación
[EP25]	Framework	Educación	No menciona	No menciona	Propuesta de solución
[EP26]	Otros	Educación	UML	Otros	Propuesta de solución
[EP27]	Otros	Educación, Otros	Otros	Diagrama de Actividad	Experiencia personal
[EP28]	Lenguaje	No menciona	DSML	Diagrama Dominio Diagrama de Actividad	Propuesta de solución
[EP29]	Framework	Educación	No menciona	Otros	Propuesta de solución
[EP30]	Framework	No menciona	Otros	Otros	Propuesta de solución
[EP31]	Framework	No menciona	Otros	No menciona	Propuesta de solución

PI1: ¿Qué contribuciones realiza respecto al modelado conceptual de los Juegos serios?

En el artículo de Céspedes-Hernández et al. [EP1] se publica un metamodelo conceptual específicamente diseñado para asistir el desarrollo de Juegos serios orientados al tratamiento de discapacidades auditivas. Alserri et al. [EP22], tras un análisis exhaustivo de la literatura disponible, proponen un modelo conceptual con el fin de incrementar el interés del público femenino en materias de ciencias de la computación.

Durk-Jouke van der Zee y Bart Holkenborg [EP2] proponen un framework para modelado conceptual para Juegos serios de simulación, el cual está basado en una secuencia de pasos y actividades ordenadas e iterativas. Por su parte, Bellotti et al. [EP8] promueven un framework con un modelo conceptual que provee un margen consistente de desarrollo, desde el diseño del contenido hasta su implementación.

Una gran parte de los estudios primarios proponen metodologías referentes a algún aspecto concreto de los Juegos serios. Martin et al. [EP5] presentan el uso de sistemas-L en la generación de escenarios. Por su parte, Chaffin y Barnes [EP3], Asuncion et al. [EP4], Baldeón et al. [EP9], Zaki et al. [EP21], Rocha et al. [EP6] y Amab et al. [EP23] exponen sobre el desarrollo en sí y su ciclo de vida como software.

Bennis et al. [EP7] evalúan y comparan cinco modelos de diseño aplicados a Juegos serios. Como extensión de su trabajo, proponen el desarrollo de una herramienta orientada a resolver los problemas hallados en el modelo DICE en [EP11].

Perrin et al. [EP10] presentan cómo puede utilizarse una arquitectura Modelo-Vista-Controlador (MVC) para la incorporación de un maestro virtual en un juego serio.

Hirdes y Leimeister [EP24] definen una serie de requisitos que debe cumplir un juego serio, con el propósito de definir un lenguaje de modelado con una estructura que los soporte, y que permita reutilizar lo desarrollado en otros proyectos.

PI2: ¿En qué ámbito se utilizan los Juegos serios?

La gran mayoría de los estudios primarios se focalizan en el ámbito de la Educación, totalizando un 54% de los estudios. En el resto de los estudios primarios, se observa cierta homogeneidad, entre medicina e ingeniería. Es importante remarcar que hay un total de 19% de estudios que no especifican cuál es el ámbito donde se utilizan los Juegos serios.

En la literatura revisada se reconocen artículos con propuestas de frameworks especializados para ciertos ámbitos. Ejemplos de ello son los estudios de Céspedes-Hernández et al. [EP1], Martin et al. [EP5], Bellotti et al. [EP8], Mayr et al. [EP13], Abdelgawad et al. [EP15] y Huynh et al. [EP27].

PI3: ¿Cuál es el Lenguaje de Modelado que se utiliza para proyectos de Juegos serios?

Una gran cantidad de estudios primarios no arrojan respuesta puntual para este interrogante, se observa que muchos de ellos recurren como base al Lenguaje de Modelado Unificado (UML), a saber: Céspedes-Hernández et al. [EP1], Perrin et al. [EP10], Bennis et al. [EP11], Avila-Pesantez et al. [EP12], Nurhadi et al. [EP14], Hamiye et al. [EP17], Roungas y Dalpiaz [EP19], Hirdes y Leimeister [EP24]. Sin embargo, Nurhadi et al. [EP14], Hamiye et al. [EP17] y Zahari et al. [EP28] argumentan que los lenguajes de modelado conceptual existentes tienen limitaciones para soportar todos los requerimientos de los Juegos serios y proponen extensiones de lenguajes específicos con base en el dominio que contemplen los modelos estructurales y lógicos necesarios para implementar los procesos de aprendizaje y dinámicas de juego en un mismo marco de trabajo. Por otro lado, catorce artículos no hacen referencia acerca del lenguaje de modelado utilizado. Tal es el caso de Durk-Jouke van der Zee y Bart Holkenborg [EP2], Chaffin y Barnes [EP3], Asuncion et al. [EP4], Bennis et al. [EP7], Bellotti et al. [EP8], Mayr et al. [EP13], Biloshchytskyi et al. [EP16], Mestadi et al. [EP18], Uskov y Sekar [EP20], Zaki et al. [EP21], Alserri et al. [EP22], Arnab et al. [EP23], Hall et al. [EP25], Mettler y Pinto [EP29]. Por otro lado, Chaffin y Barnes [EP6] proponen el uso de Deterministic Finite Automaton (DFA), Discrete Event System Specification (DEVS) y Fuzzy Inference Systems (FIS).

PI4: ¿Qué diagramas se consideran para el modelado en proyectos de Juegos serios?

Mientras que Nurhadi et al. [EP14], Hirdes y Leimeister [EP24], Zahari et al. [EP28] utilizan diagramas de dominio para representar los modelos pedagógicos y constructivos del juego, se observa que Céspedes-Hernández et al. [EP1], Perrin et al. [EP10], Hamiye et al. [EP17], Roungas y Dalpiaz [EP19] utilizan diagramas de clase para tal fin. Adicionalmente, para los flujos que definen las mecánicas de juego, Baldeón et al. [EP9], Nurhadi et al. [EP14], Hirdes y Leimeister [EP24], Zahari et al. [EP28] utilizan diagramas de actividades o derivados de este. Un caso especial lo

constituye el estudio de Melero et al. [EP26], el cual propone dos diagramas basados en el lenguaje de modelado UML, pero no contemplados en dicho estándar. No obstante, Durk-Jouke van der Zee y Bart Holkenborg [EP2], Avila-Pesantez et al. [EP12], Mayr et al. [EP13], Biloshchytskyi et al. [EP16], Mestadi et al. [EP18], Uskov y Sekar [EP20], Zaki et al. [EP21], Alserri et al. [EP22], Arnab et al. [EP23], Hall et al. [EP25] y Carvalho et al. [EP31] no especifican el uso de diagramas de modelado. Por su parte, Glenn et al. [EP5] propone el uso de cierto diagrama utilizando la gramática de sistemas funcionales L. Se destaca el uso de Storyboards en los artículos de Chaffin y Barnes [EP3], Asuncion et al. [EP4] y Rocha et al. [EP6]. Bennis et al. [EP11] menciona el uso de diagramas de nivel (Level Diagram).

PI5: ¿Cuáles son los tipos de investigación encontrados en los artículos?

Encontramos que, del total de los estudios primarios, 15 estudios (48%) tienen como propósito de investigación realizar una propuesta de solución, en su mayoría frameworks. Existen ocho artículos (26%) correspondientes a la clasificación, evaluación de la investigación. Se observó el mismo porcentaje de distribución de estudios para experiencia personal (4, 13 %) y para validación de la investigación (4, 13%).

5 Amenazas a la validez

Se analizaron las potenciales amenazas a la validez que podrían afectar al SMS, respecto a las cuatro categorías sugeridas por Wohlin et al. [12].

Validez del constructo. En este SMS, con el fin de mitigar estas amenazas, describimos el significado que le hemos dado al modelado conceptual y Juegos serios basados en literatura reconocida [1], [2], [3], [4], [5], [6].

Validez interna. Para mitigar las preocupaciones sobre la validez interna, los cuatro primeros autores crearon un protocolo de revisión como parte de la investigación de un trabajo de investigación del Seminario de modelado conceptual de la Maestría en Ingeniería en Sistemas de Información (UTN-FRBA) y éste fue revisado por los últimos dos autores (docentes del Seminario).

Validez externa. Se tomó la decisión de utilizar tres motores de búsqueda en nuestra búsqueda de las revistas y actas de congresos que son relevantes y recomendados para el campo de la Ingeniería de software. No se consideró la literatura gris, como los artículos disponibles solo en forma de resúmenes, presentaciones en PowerPoint, tesis doctorales o libros, porque incluirlos podría haber afectado la validez de nuestros resultados.

Fiabilidad. Se intentó mitigar el sesgo de las publicaciones definiendo cuidadosamente (a) los criterios de inclusión y exclusión para poder seleccionar estudios primarios y (b) los criterios de exclusión específicamente, con el fin de seleccionar reglas basadas en las preguntas de investigación predefinidas en el trabajo. Para aumentar la confiabilidad, paralelamente un grupo de dos alumnos y una docente aplicaron los criterios y otro grupo de dos alumnos con la otra docente los aplicaron por separado, realizaron la catalogación de los estudios; se discutieron las discrepancias

entre ellos, con el propósito de determinar si era apropiado incluir un artículo en particular o no, y de ese modo se obtuvo el listado final de estudios primarios. Además, se diseñó un formulario para la registración de los datos con Excel y se mapearon las preguntas de investigación de acuerdo con el esquema de clasificación definido para cumplir con los objetivos de este estudio. Se considera que el efecto potencial de este sesgo tiene menos importancia en estudios de mapeos sistemáticos que en las revisiones sistemáticas de literatura.

Para fortalecer la confiabilidad, luego de aplicar los criterios de inclusión y exclusión, se creó una matriz con las propiedades de los datos extraídos de los artículos y se los catalogó con las preguntas de investigación con el motivo de cumplir con el objetivo de este estudio.

6 Conclusiones

En este artículo se presentó un mapeo sistemático de la literatura para analizar el estado del arte respecto al modelado conceptual de Juegos serios. Se seleccionaron 31 estudios primarios de un conjunto inicial de 558 artículos resultantes de las búsquedas realizadas en SCOPUS, IEEE Xplore y ACM, en el período comprendido entre enero del año 2010 y junio del año 2021. Una vez analizados los estudios primarios, se concluye que:

- UML es el lenguaje de modelado predominante para el modelado de Juegos serios en los estudios primarios. Sin embargo, se detectan otros, como UP4EG, DSML, Deterministic Finite Automaton (DFA), Discrete Event System Specification (DEVS) y Fuzzy Inference Systems (FIS).
- No ha sido posible identificar características de modelado conceptual diferenciables de acuerdo con el ámbito del uso de los Juegos serios, aunque sí se reconocen estudios primarios con propuestas de frameworks especializados para ciertas disciplinas o problemáticas.
- Si bien predomina la utilización de diagramas UML en su mayoría los diagramas de clases y diagramas de actividad; también se emplea una representación heredada del desarrollo de videojuegos: Storyboards. Por otro lado, un 32 % de los estudios primarios no mencionan el uso de diagramas específicos.
- El 30% de los estudios primarios proponen un framework para el proceso de desarrollo de Juegos serios. El mismo porcentaje (30 %) de los artículos propone una metodología.
- El 48 % de las publicaciones realizan una propuesta de solución, el 26 % de los estudios presentan una evaluación de investigación y los artículos de validación de investigación y de experiencia personal tienen un 13% cada uno. No se identificaron estudios primarios filosóficos o que informen una opinión.
- Cabe destacar que la mayoría de los frameworks encontrados no especifican la manera para realizar el modelado conceptual.

En la sección introducción de este artículo, se mencionó que la motivación del desarrollo de este SMS ha sido acercar a los alumnos del Seminario de Modelado Conceptual de la Maestría en Ingeniería en Sistemas de Información (UTN-FRBA) a tópicos de interés propuestos en congresos internacionales relacionados al modelado

conceptual y adquirir los conocimientos sobre el tema de modelado conceptual analizado en el SMS.

Referencias

1. Mayr, H.C., Thalheim, B. The triptych of conceptual modeling. Software System Model 20, 7–24 (2021).
2. Larman, C. UML y PATRONES. Una introducción al análisis y diseño orientado a objetos y al proceso unificado (Begoña, M., Trad.). Madrid: Pearson Educación, SA (Original en inglés publicado en 2002) (2003).
3. Pons, C.F., Giandini, R.S. y Pérez, G.A. Desarrollo de software dirigido por modelos. Editorial de la Universidad Nacional de La Plata (EDULP) / McGraw-Hill Educación, (2010).
4. Michael, D. R., & Chen, S. L. Serious games: Games that educate, train, and inform (2005).
5. Delgado, J. C. S., & Sanz, C. V. (2020). Juegos serios para potenciar la adquisición de competencias digitales en la formación del profesorado/Serious Games to Enhance Digital Competencies Acquisition for Training Faculty. Educación, 44(1), NA-NA.
6. Petri, G., von Wangenheim, C. G., & Borgatto, A. F. (2017, May). Quality of games for teaching software engineering: an analysis of empirical evidences of digital and non-digital games. In 2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering Education and Training Track (ICSE-SEET) (pp. 150-159). IEEE.
7. ER 2021. 40 th International Conference on Conceptual Modeling. Disponible en: https://er2021.org/topics.html.
8. B. Kitchenham, D. Budgen, P. Brereton, Evidence-Based Software Engineering and Systematic Reviews, Chapman and Hall 1st.Editon. Chapman and Hall/CRC. USA, 2015.
9. K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, “Systematic mapping studies in software engineering”, in Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, pp. 68–77, 2008.
10. Wieringa R., Maiden N., Mead N., Rolland C. Requirements engineering paper classification and evaluation criteria: A proposal and a discussion. Requirements Engineering, 11, pp. 102–107 (2005).
11. Chimuris Gimenez, A., Miguel, J. C., Bassi, M., Garrido, N., Velazquez, G., Panizzi, M. Apéndice. Modelado Conceptual de Juegos Serios: Revisión sistemática de la literatura. Disponible en: https://doi.org/10.6084/m9.figshare.15183552.v1.
12. C. Wohlin, P. Runeson, M. Hst, M.C. Ohlsson, B. Regnell, A. Wessln, “Experimentation in Software Engineering”, Springer Publishing Company, 2012.

ISBN 978-987-633-574-4
ISBN 978-987-633-575-1

CACIC 2021

WORKSHOP ARQUITECTURA, REDES Y SISTEMAS OPERATIVOS

COORDINADORES

Jorge Ardenghi (UNS)
Carlos Buckle (UNPSJB)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978–987–633–574–4

.Análisis del comportamiento de variantes de TCP cuando se producen desconexiones de un nodo móvil de una red heterogénea

Diego R. Rodriguez Herlein[1], Carlos A. Talay[1] and Luis A. Marrone[2]

[1] Instituto de Tecnología Aplicada I.T.A
Universidad Nacional de la Patagonia Austral – U.A.R.G. – U.N.P.A.
Río Gallegos, Argentina
[2] L.I.N.T.I
Universidad Nacional de La Plata
La Plata, Argentina
{dherlein@uarg.unpa.edu.ar, ctalay@uarg.unpa.edu.ar and lmarrone@info.unlp.edu.ar}

Abstract. El presente trabajo está dedicado al análisis de desempeño de distintas variantes del protocolo TCP en redes heterogéneas, cuando se producen desconexiones del host móvil. El objetivo es analizar el comportamiento de variantes de TCP, influenciadas en su desarrollo por el escenario de implementación, en una red hibrida con enlaces cableados e inalámbricos y desconexiones de distinta duración. Para ello, se implementó un modelo simple de una red de acceso WLAN en el simulador NS-2 y se realizaron distintas pruebas que permitieron clasificar y cuantificar los efectos negativos para el rendimiento de TCP que producen las desconexiones frecuentes.

Keywords: TCP, Desconexiones, NS2, Rendimiento.

1 Introduction

El protocolo TCP se diseñó específicamente para proporcionar un flujo de bytes confiable de extremo a extremo, a través de una red no confiable que puede tener diferentes topologías, anchos de banda, retardos, tamaños de paquete y demás características, a través del camino [1]. Transporta la mayor parte del tráfico de internet, por lo que su rendimiento depende, en gran medida del propio rendimiento del este protocolo.

Para lograr esa confiabilidad extremo a extremo, asigna un número de secuencia a cada byte transmitido, y espera una confirmación positiva (ACK) de la capa de transporte del receptor. Para administrar la cantidad de datos que se envían a la vez, utiliza el denominado algoritmo de ventana deslizante [2]. El tamaño de la ventana establece el número máximo de bytes que pueden ser transmitidos sin haber sido aún reconocidos.

El protocolo TCP fue diseñado originalmente para redes cableadas donde las pérdidas y los retrasos de paquetes se deben casi con exclusividad a la congestión de la red.

La congestión es uno de los principales problemas que se afronta en la transmisión de datos, e implica que se saturan los recursos de la red, degradándose su utilización [3]. Cuando la congestión empieza a producirse, el tiempo de transmisión a través de la red aumenta. Conforme la congestión se hace más severa, los nodos de la red descartan paquetes, llegando al extremo en que esta puede colapsar. A partir de lo que se denominó el colapso por congestión, se incorporó un mecanismo de control, conformado por distintos algoritmos. Estos algoritmos de control de congestión, basados en ventanas, permiten adaptar la tasa de envío en forma dinámica, evitando saturar la capacidad de la red y, al mismo tiempo, utilizar eficientemente el ancho de banda disponible, proporcionando una parte justa del ancho de banda de la red a todas las conexiones [4].

De esta forma, el control de congestión primero debe detectar la congestión y después, tomar las acciones necesarias. En las redes cableadas, TCP supone que hay congestión cuando se pierde un paquete.

Las redes inalámbricas han experimentado un importante auge en los últimos años debido, principalmente al desarrollo tecnológico. Estas redes brindan flexibilidad y portabilidad al usuario, sin tener que sacrificar la conexión a Internet o a la red del lugar de trabajo.

El control de congestión estaba diseñado a medida de las redes cableadas, donde los datos normalmente llegan en orden y prácticamente sin errores. Este diseño está claramente influenciado por el escenario. De esta manera, se encuentra con nuevos desafíos en los enlaces inalámbricos, debido a su naturaleza más impredecible [5].

Una de las principales características del enlace inalámbrico es su alta tasa de error (BER). La tasa de error en tránsito es importante debido a que presentan características de transmisión diferentes. Esto tiene su origen en la propia naturaleza del medio físico, que depende de las condiciones del entorno, dado que las señales que se propagan sufren de atenuación, interferencia y ruido, por lo que los paquetes que se reciben pueden estar dañados y se descartan, produciendo la pérdida de paquetes en tránsito. Además, se le suman los efectos de la movilidad de los hosts y del entorno, que pueden modificar las condiciones de transmisión, produciendo otras pérdidas de paquetes en tránsito. Por lo tanto, el reordenamiento de paquetes es más frecuente. De esta manera, es habitual observar largos retardos y pérdida de paquetes que no se deben exclusivamente a la congestión y el control de congestión enfrenta nuevos desafíos en el entorno inalámbrico.

Históricamente la detección de la congestión es por perdida de paquetes. Esta suposición desdibuja en un escenario inalámbrico porque ya no es una medida confiable para detectar la congestión [6]. En este nuevo escenario hay un porcentaje elevado de pérdidas que no tiene que ver con la congestión. Esto se debe a que la mayoría de las técnicas de control de congestión están diseñadas para un escenario a nivel de red, pero estas pérdidas ocurren a nivel de enlace.

Por esta razón, TCP reacciona inadecuadamente ante las pérdidas de paquetes no relacionadas con la congestión, por ejemplo, si se pierde un paquete de datos debido a interferencias de radiofrecuencia de corta duración en el canal de trasmisión. A pesar de que no hay desbordamientos de buffer, TCP tomara la decisión de reducir la

ventana de congestión en forma innecesaria y, por ende, su tasa de transmisión y el rendimiento. En cambio, lo que debería haber sucedido es recuperarse de la pérdida y continuar la transmisión a la misma tasa de envío pues no había congestión en la subred.

Los problemas de rendimiento de TCP no se circunscriben exclusivamente a los enlaces inalámbricos. Otros escenarios requirieron modificaciones del control de congestión para mejorar su rendimiento en esos enlaces. Se puede observar que la búsqueda de mejora de rendimiento es en función de escenarios en particular [7]. Como en su diseño original, las variantes de TCP están fuertemente influenciadas por las características del escenario en particular para el que fueron desarrolladas.

Sin embargo, TCP debe poder desplegarse en cualquier tipo de red y debe ser flexible para lograr un buen rendimiento en redes donde del origen a destino puede transitar distintos tipos de medio de transmisión. Hoy es muy probable que la conexión comience en un enlace inalámbrico (cliente) y termine en un servidor conectado a una red cableada. Para lograr esa flexibilidad, ha tenido que renunciar a conseguir un rendimiento óptimo en distintos escenarios como son las redes de muy alta velocidad, las redes con valores altos de BDP (producto de ancho de banda – latencia), y en medios con valores de pérdidas altas como los entornos inalámbricos.

Durante los últimos años, se han realizado distintas propuestas para mejorar el protocolo y adecuarlo a estos nuevos ambientes. A pesar de todas las propuestas, que dieron lugar a tantas otras variantes del protocolo TCP, no han tenido una gran aceptación, debido a que debe transitar todo tipo de redes.

Analizando algunas de las distintas implementaciones de TCP [8], se pudo observar que están ligadas a escenarios particulares. Detrás de cada variante hay una búsqueda de mejora de la performance en función de un conjunto acotado de características de estos escenarios y se observa su fuerte influencia en el desarrollo de las variantes de TCP. Es decir, las mejoras fueron desarrolladas para mejorar la performance en escenarios de particulares características comunes.

Una mejora en un escenario es de esperar que mejore el rendimiento para esos requerimientos. A pesar de todas las propuestas que se pueden encontrar en la literatura, estas no han tenido una gran aceptación, dado que su aplicación será, en general, en una red donde un segmento TCP es inevitable que atraviese escenarios diferentes.

Como se está evaluando el rendimiento de una sesión TCP, esta se evalúa de extremo a extremo. Por esta razón, resulto relevante analizar qué es lo que sucede cuando estas variantes deben atravesar un escenario con distintos tipos de enlace en su camino. De esta forma, es interesante ensayar el rendimiento de distintas variantes de TCP desarrolladas para distintos tipos de enlaces y escenarios, como lo que se plantean en una red heterogénea de acceso WLAN.

Por esta razón se utilizaron en los ensayos, versiones o implementaciones de TCP pensados para mejorar el rendimiento en escenarios tanto fijos como inalámbricos.

2 Casos de Prueba

En una conexión que incluye enlaces inalámbricos, se producen frecuentemente desconexiones debido a la movilidad, crisis de energía o variaciones en el entorno. Este tipo de desconexiones puede tener distinto tiempo de duración, aunque suele durar más que un RTO y es más corto que la vida útil de una conexión TCP.

El emisor ignora los motivos de las pérdidas e invoca los algoritmos de control de congestión erróneamente, reduciendo así el rendimiento efectivo tras la reconexión.

Es así que resulta interesante analizar el comportamiento de una red WLAN de acceso ante la presencia de desconexiones de diferentes tiempos de duración. Se simula un modelo sencillo (Figura 1) de 3 nodos y un solo flujo para poder visualizar el efecto en el rendimiento que producen estas desconexiones. Las redes de acceso inalámbricas típicas comprenden un host móvil conectado de forma inalámbrica a una estación base, que a su vez está conectada a la red troncal cableada, posiblemente a Internet. En la Figura 1 se observa el modelo implementado en el simulador. La razón de recurrir a este modelo sencillo, responde a que permite introducir, en forma controlada, los efectos individuales pudiéndose observar sus respuestas en forma específica y, de esta manera, aislar el efecto de las desconexiones de las otras causas de perdida de rendimiento de TCP.

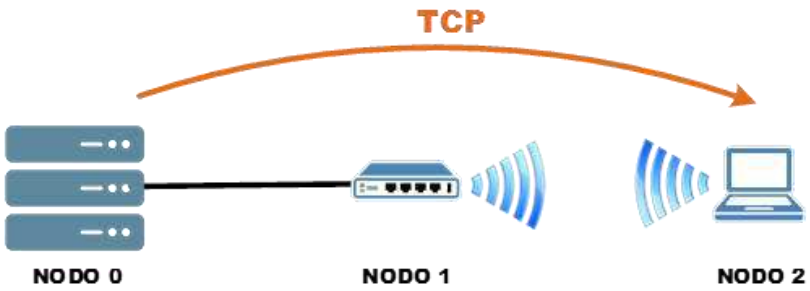


Fig. 1. Modelo Simple de tres nodos y un flujo TCP. (WLAN).

En cuanto a las simulaciones planteadas en este trabajo, no se tiene en cuenta el efecto de la movilidad, ya que en definitiva se traducirá a errores en la comunicación.

Para explorar el efecto de las desconexiones se implementó el modelo propuesto en el simulador de redes de eventos discretos (NS-2, Network Simulator 2), en su versión 2.35 (released nov. 4 2011) [9]. La selección de esta topología es una aproximación a un escenario inalámbrico con un nodo fijo, una estación base y un nodo móvil, con la simplificación práctica que el enlace inalámbrico no presenta errores y solo tiene desconexiones de distinta duración producidas sobre el enlace inalámbrico.

El nodo fijo y la estación base están vinculados por un enlace cableado que se configuro como dúplex, con un ancho de banda 10 Mb/s, retardo de propagación 2 ms. y política de servicio de las colas DropTail. El enlace entre la estación base y el nodo móvil es inalámbrico y se configuro como modo de propagación TwoRayGround, la capa física WirelessPhy, MAC 802.11, la antena OmniAntenna, 2Mb./s con MAC 802.11. El nodo inalámbrico no posee movimiento.

El flujo TCP se define para cada una de las simulaciones entre el nodo fijo (emisor) y el nodo móvil (receptor). El flujo comienza a transmitir y las desconexiones comienzan a partir del paquete número 1000. En cada simulación se transmitieron 3.000 segmentos de TCP de 1.000 Bytes cada uno. Las simulaciones fueron realizadas en forma independiente para cada variante del protocolo y para cada una de los tiempos de desconexión. Las desconexiones ensayadas tuvieron una duración de 0, 05, 0,1, 0,5, 1, 2, 5, 10 y 20 segundos.

El flujo TCP se define para cada una de las simulaciones entre el nodo fijo (emisor) y el nodo móvil (receptor). El flujo comienza a transmitir y las desconexiones comienzan a partir del paquete número 1000. En cada simulación se transmitieron 3.000 segmentos de TCP de 1.000 Bytes cada uno. Las simulaciones fueron realizadas en forma independiente para cada variante del protocolo y para cada una de los tiempos de desconexión. Las desconexiones ensayadas tuvieron una duración de 0,05, 0,1, 0,5, 1, 2, 5, 10 y 20 segundos.

En cuanto a las variantes ensayadas, se consideraron las variantes más utilizadas tales como Compound (Windows) [10], CUBIC (Linux) [11]. También se incorporó TCP Westwood [12] desarrollado para las redes con enlaces inalámbricos y TCP Reno, como referencia de control de congestión basado en la perdida de paquetes. Además, se incluyó TCP vegas [13] como un caso particular, tanto por sus respuestas en los ensayos, como su particular mecanismo de control de congestión que sentó las bases para desarrollar distintas técnicas que permitieran diferencias las causas de pérdidas por congestión de las que no lo son, utilizando solo la retroalimentación del receptor sin ayuda de la red.

3 Resultados Obtenidos

Para poder analizar los efectos de las desconexiones de distinta duración sobre variantes de TCP, a continuación, se presentan distintas métricas tales como Throughput vs. Tiempo, Throughput Promedio y el Tiempo Total de Transmisión en función de la duración de la desconexión, la evolución del tamaño de la ventana de congestión y del número de secuencia del segmento TCP en función del tiempo.

La Figura 2 representa el throughput instantáneo de TCP Reno para el ensayo con una desconexión de 5 segundos de duración.

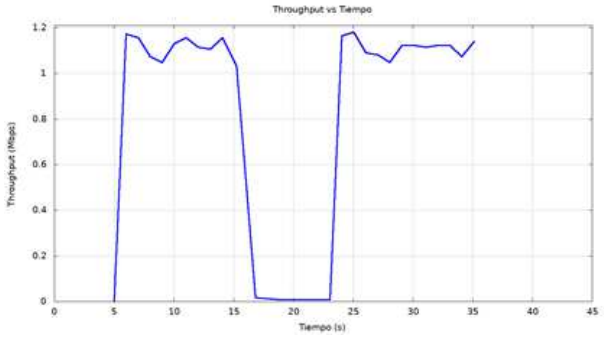


Fig. 2. Throughput vs Tiempo - TCP Reno (Desconexión 5s.)

Como puede observarse, la desconexión con una duración de 5 s., produce una caída abrupta del throughput y recupera su ritmo de transferencia de datos casi inmediatamente recuperada la conexión.

Una de las métricas de interés para análisis es la evolución del tamaño de la ventana de congestión en función del tiempo de la simulación, pues tiene directa relación con el rendimiento de TCP.

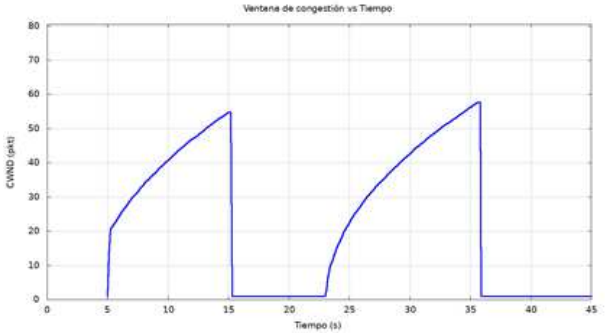


Fig. 3. CWND vs Tiempo - TCP Reno (Desconexión 5s.)

En la Figura 3 se observa la evolución de la ventana de congestión en función del tiempo para la variante TCP Reno para desconexión de 5 segundos. Se puede observar cómo actúan los algoritmos de control de congestión de TCP Reno, reduciendo el valor de la ventana a 1 durante la desconexión y disparando el algoritmo de Slow Start.

Así mismo, en la Figura 4 se puede observar los valores del throughput promedio en Mbit/s, para cada uno de los casos considerando.

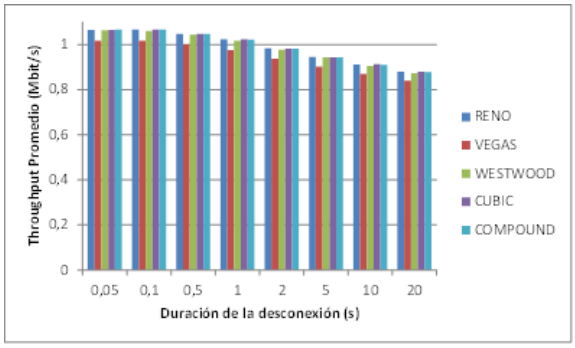


Fig. 4. Throughput promedio vs duración de las desconexiones

En esta figura se observa un patrón común para todas las variantes ensayadas: el throughput promedio disminuye a medida que el tiempo de desconexión aumenta. Más allá de alguna pequeña variación en la disminución relativa entre ellas, de los resultados se desprende que TCP Vegas es de las variantes ensayadas, la que presenta el menor valor de throughput promedio en todos los casos.

Dado que las simulaciones se realizan transmitiendo una cantidad de información constante, resulta interesante analizar el tiempo que demora cada una de las variantes estudiadas en transmitirlos. Esta medida puede tener directa relación con el rendimiento, lo que permitirá observar cómo afecta esas desconexiones a cada una de las variantes analizadas desde otra perspectiva.

En la siguiente figura se observa el Tiempo Total de Transmisión en función del tiempo para cada una de las variantes ensayadas y para cada uno de los tiempos de desconexión.

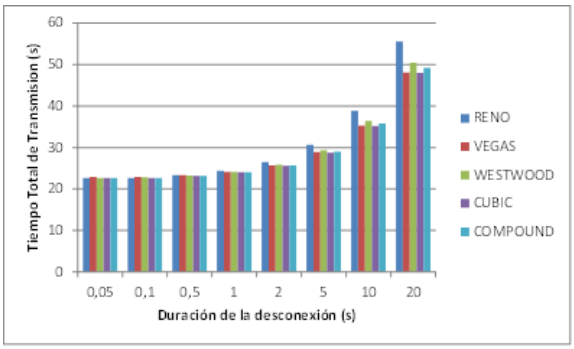


Fig. 5. Tiempo total de transmisión vs duración de las desconexiones

Ahora observamos que a medida que el tiempo de desconexión crece, el tiempo necesario para trasmitir todos los datos también lo hace, de una forma exponencial. Además, para desconexiones de baja duración el tiempo necesario para transmitir todos los paquetes es prácticamente del mismo orden. Sin embargo, a medida que aumenta el tiempo de desconexión, empiezan a diferenciarse distintos valores de tiempo total de transmisión, siendo la variante de TCP Reno la que más tiempo necesita para concluir la transmisión de los datos.

Como análisis complementario al tiempo total de transmisión, resulta interesante observar la evolución del número de secuencia del segmento TCP. Al ser un valor instantáneo permite observar la dinámica de la transmisión.

A continuación, se observan una serie de 3 figuras (Figura 6 a Figura 8) que representan la evolución del número de secuencia del segmento TCP en función del tiempo de simulación. La serie muestra a las variantes de Reno, Vegas y Westwood para la desconexión de 20 segundos de duración.

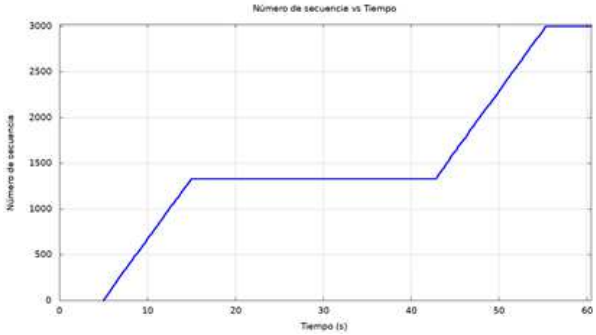


Fig. 6. Nº de Secuencia vs. Tiempo - TCP Westwood (Desconexión 20s.)

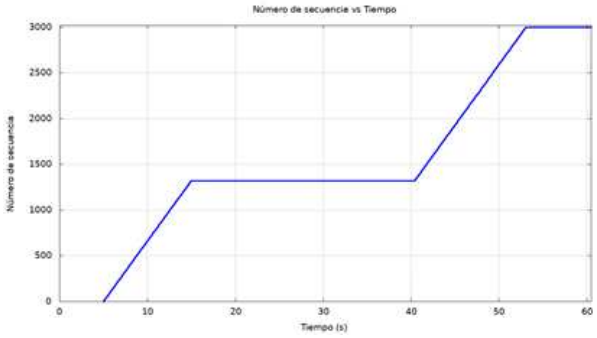


Fig. 7. Nº de Secuencia vs. Tiempo - TCP Vegas (Desconexión 20s.)

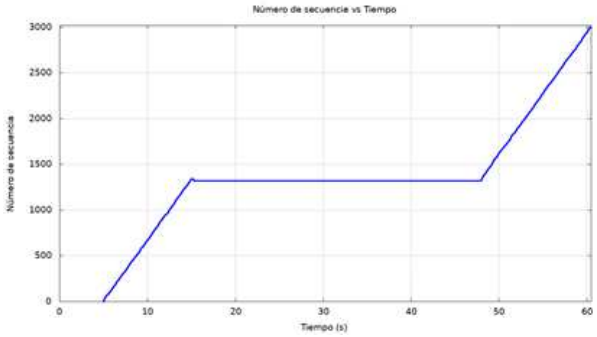


Fig. 8. Nº de Secuencia vs. Tiempo - TCP Reno (Desconexión 20s.)

Para el mismo tiempo de desconexión, comparando TCP Westwood (Figura 6) y TCP Vegas (Figura 7), con TCP Reno (Figura 8) es este último el que demora más tiempo en retomar el envío de los datos, lo que repercute directamente en el tiempo total de transmisión.

En la Figura 8 se puede determinar el tiempo necesario desde el momento de la desconexión para que TCP Reno continúe la transmisión de los datos. Si bien la desconexión es de 20 segundos, TCP Reno demora casi el doble de ese tiempo en retomar el envío.

Teniendo en cuenta las consideraciones anteriormente expuestas, se observa que la interrupción del envío de datos afecta en menor medida al protocolo Vegas y en mayor medida al protocolo Reno, y la variante Westwood, es la que presenta una respuesta intermedia.

4 Trabajos Futuros

Se prevé ampliar el número de variantes del protocolo TCP analizadas, como así incluir otras causas de pérdidas de paquetes en tránsito de las redes inalámbricas, y estudiar cómo afectan su rendimiento.

5 Conclusiones

En los resultados de las simulaciones se pudo observar que, si bien el throughput promedio disminuye a medida que el tiempo de desconexión aumenta, en forma similar para las variantes ensayadas. Sin embargo, al analizar el tiempo total de transmisión se observa allí las diferencias en la recuperación de cada variante, siendo TCP Reno la que requiere mayor tiempo para completar el envío de datos.

Durante la desconexión, se descartan tanto los paquetes de datos como los ACK, y cada intento de retransmisión conduce a una retransmisión fallida, Esta condición solo se detecta por la caducidad de los RTO. Estos intentos consecutivos fallidos harán que el emisor TCP aumente exponencialmente los tiempos de acuerdo al algoritmo de Karn, Esto dará como consecuencia que se espacien los reintentos y por tanto el canal permanezca inactivo desperdiciando recursos de red.

Además del tiempo sin actividad, el rendimiento de TCP también se degrada por el disparo de los algoritmos de control de congestión cuando no hay descarte de paquetes en los nodos intermedios. Esto produce que el TCP emisor reduzca su tasa de envío cuando no es necesario. Esto se observa en la gráfica de la evolución del tamaño de la ventana de congestión en función del tiempo (Figura 3)

En los ensayos se observó que TCP Reno requiere mayor tiempo para recuperar la tasa de envío, debido a los crecientes períodos de inactividad producidos por el backoff exponencial que hace crecer los valores de RTO. Esto se produce porque cada timeout en serie reduce su ventana de congestión, lo que implica iniciar la fase de Slow Start en cada ocasión. Dado que las desconexiones también impiden el arribo de los ACK, no se disparan los algoritmos de Fast Retransmit ni Fast Recovery, pues no puede recibir ningún ACK duplicado. La recuperación se inicia reduciendo el valor de la ventana de congestión (en 1) e invoca al algoritmo de Slow Start, causando un impacto muy negativo en el rendimiento de TCP.

En los casos más severos, al producirse una desconexión y posterior reconexión del host móvil, el emisor podría seguir a la espera hasta que ocurra un timeout para reiniciar la retransmisión.

Además, una vez que deja de estar en espera, requiere de un tiempo para llegar a las tasas de trasmisión anteriores a la desconexión del nodo. Esto se debe no solo al disparo del control de congestión, por más que no la hubiera, sino también debido a la continua reducción a la mitad del umbral (ssthresh) con los timeout en serie.

Agradecimientos

Agradecemos al alumno de la Lic. en Sistemas de la UNPA-UARG Franco A. Trinidad su colaboración en la realización de las pruebas con el simulador.

Referencias

1. J. Postel, RFC 793: Transmission Control Protocol, September 1981.
2. A. Afanasyev, N. Tilley, P. Reiher, and L. Kleinrock, Host-to-Host Congestion Control for TCP. IEEE Communications Surveys Tutorials, vol. 12, no. 3, 3rd quarter 2010, pp. 304- 340
3. Rodríguez Herlein, D. R.; Talay, C. A.; González, C. N.; Trinidad, F. A.; Almada, L.; Marrone, L. A., Un análisis de comportamiento entre distintos mecanismos de control de congestión ensayados sobre una topología mixta. CACIC 2018. Tandil, Argentina. http://sedici.unlp.edu.ar/handle/10915/73349
4. Rodríguez Herlein D. R., Talay C. A., González C. N., Trinidad F. A., Almada M .L., Marrone L. A. (2019) Contention Analysis of Congestion Control Mechanisms in a Wireless Access Scenario. In: Pesado P., Aciti C. (eds) Computer Science – CACIC 2018. CACIC 2018. Communications in Computer and Information Science, vol 995. Springer, Cham. https://doi.org/10.1007/978-3-030-20787-8_18
5. Teja F. R., Vidal, L, Alves, L., TCP sobre enlaces wireless – Problemas y algunas posibles soluciones existentes, Curso de posgrado y actualización, Instituto de Ingeniería Eléctrica, Facultad de la República, marzo 2004. https://studylib.es/doc/8541837/artículo-sobre-tcp-en-wireless
6. D. R. Bhadra, C. A. Joshi, P. R. Soni, N. P. Vyas and R. H. Jhaveri, Packet loss probability in wireless networks: A survey, 2015 International Conference on Communications and Signal Processing (ICCSP), 2015, pp. 1348-1354, doi: 10.1109/ICCSP.2015.7322729
7. D. R. Rodriguez Herlein, Análisis del rendimiento del protocolo TCP en redes de acceso wireless, Master Thesis, Fac. Informática., UNLP, La Plata, Bs.As., Argentina, 2020
8. Saleem-ullah Lar, Xiaofeng Liao, An initiative for a classified bibliography on TCP/IP congestion control, Journal of Network and Computer Applications, Volume 36, Issue 1, 2013, Pages 126-133, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2012.04.003
9. T. Issariakul, E. Hossain, Introduction to Network Simulator 2, Segunda Edición. New York: Editorial Springer. 2012
10. S. Mascolo, C. Casetti, M. Gerla, S. Lee, and M. Sanadidi, TCP Westwood: congestion control with faster recovery, Univ. California, Los Angeles, Tech. Rep. CSD TR, vol. 200017, pp. 1–14, 2000
11. Sangtae Ha; Injong Rhee; Lisong Xu (July 2008). CUBIC: A New TCP-Friendly High-Speed TCP Variant, ACM SIGOPS Operating Systems Review. 42 (5): 64–74
12. S. Mascolo, C. Casetti, M. Gerla, S. S. Lee and M Sanadid, TCP Westwood: Congestion control with faster recovery. Technical Report 200017, UCLA CSD 2000
13. Low, Steven; Peterson, Larry & Wang, Limin. Understanding TCP Vegas: Theory and Practice, https://www.cs.princeton.edu/research/techreps/TR-616-00

ISBN 978–987–633–574–4

Entorno de contenedores con emuladores de sistemas embebidos STM32

Esteban Carnuccio[1], Waldo Valiente[1], Mariano Volker[1], Raúl Villca[1], Matías Adagio[1]

[1] Universidad Nacional de La Matanza,
Departamento de Ingeniería e Investigaciones Tecnológicas
Florencio Varela 1903 - San Justo, Argentina
{ecarnuccio, wvaliente, mvolker, rvillca, maadagio}@unlam.edu.ar
www.unlam.edu.ar

Resumen. Ante la gran importancia que han tenido los sistemas embebidos, en los últimos años, debido al auge de Internet de las Cosas. Resulta de vital importancia conocer y probar el hardware antes de adquirirlo, para saber si cumple las necesidades de un proyecto determinado. En ese contexto, esta investigación se centró en la creación de un entorno automatizado de emulación, que permita probar rápida y fácilmente determinadas placas de desarrollo, sin necesidad de adquirir el hardware físico. Con esa premisa, se desarrolló un entorno dentro de un contenedor Docker, que permite realizar la emulación de determinadas placas de la familia STM32 a través del programa Qemu, listo para funcionar. De esta forma se podrá realizar distintas pruebas, sin la necesidad de realizar una tediosa configuración e instalación de los componentes y las dependencias.

Palabras Clave. Docker, Emulación, Internet de Las Cosas, Qemu, Sistemas Embebidos, STM32.

1 Introducción

En los últimos años creció de forma exponencial la tecnología de Internet de las Cosas. El término IoT toma relevancia cuando se superó la cantidad de dispositivos conectados a internet, que el número de personas que existían en el mundo en ese momento. Según las proyecciones de [1], se estima que en el año 2025 habrá aproximadamente cien mil millones de sistemas embebidos conectados, que transmitirán los datos de sus sensores para ser procesados en servidores externos, a través de internet. En este sentido, el tiempo y los costos que incurren para configurar un entorno de trabajo de software encargado de construir el sistema embebido, se vuelve un factor importante. Ya que para determinar si la placa de desarrollo puede cumplir con las necesidades, es necesario adquirir la placa físicamente, con el agregado de la curva de aprendizaje de su utilización. Muchas veces se trabaja sobre un producto que, en una etapa avanzada de un proyecto, se descubre que no es la indicada para cumplir con sus requerimientos. Como consecuencia, se debe adquirir un nuevo sistema embebido, que pueda cumplir con sus objetivos, produciendo así atrasos y aumento en los costos.

Para tratar de minimizar estos riesgos existen simuladores y emuladores de sistemas embebidos, como *Thinkercad* y *Proteus*. Pero estos presentan ciertas limitaciones de uso, que dificultan su utilización en los distintos proyectos IT [2]. Por ese motivo en esta investigación se desarrolló un entorno de integración automatizado a través de contenedores Docker, que permiten ejecutar programas en distintas placas STM32 emuladas a través de Qemu. De forma tal que permita realizar rápidamente distintas pruebas de aprendizaje en el entorno y el sistema embebido. Estos ejemplos a su vez funcionan en el embebido real.

En este documento, inicialmente se describe brevemente la comparación con otros entornos. Posteriormente se brinda una introducción al funcionamiento del contenedor de Docker y su registro de contenedores llamado Docker Hub. Luego, se detalla la forma en que se implementó y configuró el entorno de trabajo utilizando Qemu, dentro de Docker para los sistemas embebidos seleccionados. Finalmente, se detalla los ejemplos de código que se pueden ejecutar dentro del entorno.

2 Trabajos relacionados

Esta investigación se sustentó en diferentes trabajos relacionados sobre el emulador de Qemu para diferentes sistemas embebidos. A continuación, se realiza un repaso de los principales. Existen desarrollos como [3], donde se presenta una extensión de Qemu para integrarlo con la aplicación Eclipse. Esta debe ser instalado manualmente por el usuario y es un poco complicada su configuración. Además, solo se explica el ejemplo de encendido y apagado del led de testeo de la placa, mejor conocido como "*Blinky Led*". Sin embargo, este no ofrece explicación de cómo se deben emular otros sensores, actuadores y componentes del embebido.

Por otro lado, existe el proyecto realizado por *Beckus* [4]. El cual presenta su propia versión adaptada del código fuente de Qemu, que a su vez fue modificada por otros autores, como se describen en [5]. Pero para lograr utilizarlos, es difícil hacerlos funcionar correctamente, ya que se debe hacer la instalación y compilación de forma manual. No obstante, el proyecto de *Beckus* ofrece la forma de crear un contenedor Docker, pero no se logró hacerlo funcionar.

En [6] se menciona el proyecto de *Pebble*, que es una adaptación de *Beckus*, también es difícil su configuración e instalación. Pero es un proyecto en proceso, que no registra cambios presentes de actualización.

Finalmente, en el registro de contenedores de Docker[7], se puede descargar una imagen que emplea el proyecto *Beckus*, pero este no permite ejecutar los ejemplos. Además, no posee la documentación detallada con las formas en que se deben ejecutar dichos ejemplos. Tampoco tiene habilitado el ingreso por protocolo *SSH* dentro del contenedor, para su utilización.

Para subsanar las falencias descritas de los proyectos antes expuestos, en este trabajo de investigación se detalla cómo se armó el entorno contenido en una imagen Docker, en el que se modificó y adaptó el proyecto de *Beckus*. Obteniendo así un entorno de emulación bien documentado y funcional. Para ello se describen los distintos cambios realizados y se comentan brevemente los distintos tutoriales explicativos, para la

emulación de los sensores y actuadores en ciertas placas de desarrollo de la familia de microcontroladores STM32. Todo esto se hizo de forma automatizada, para que se pueda probar el mismo programa, que funciona en forma física, dentro del contenedor en forma emulada.

3 Desarrollo

3.1 Docker

Como se comenta en [8], Docker es una de las herramientas más populares en estos días en el mundo de la tecnología (IT). Básicamente, hay dos corrientes principales en el paradigma de Docker. El Primero, la plataforma Docker es de código abierto, esto permite que se equipe continuamente con nuevas características y funcionalidades relevantes. Siendo aprovechada no solo por programadores, sino también por equipos operativos de IT. La segunda tendencia es la adopción sin precedentes de la tecnología de contenedores, que es complementada por varios proveedores de soluciones y servicios de IT en todo el mundo. Estos dos permiten una mayor simplicidad en el desarrollo de aplicaciones, gracias a la implementación automatizada y acelerada de contenedores Docker. Siendo ampliamente promocionada como la clave diferenciadora del éxito sin precedentes de este paradigma.

3.2 Docker Hub

Docker Hub[1] es un registro de contenedores mantenido por Docker Inc. En el repositorio se encuentran las principales imágenes oficiales, como por ejemplo *Busybox*, Sistemas operativos como Ubuntu o Windows, incluso programas ya empaquetados como Apache, Node.js o WordPress entre otras. También permite a cualquier usuario crear una cuenta en forma gratuita y subir sus propias imágenes [9]. La protección de seguridad brindada es muy simple. Solo los propietarios de las imágenes publicadas y los usuarios habilitados, pueden realizar cambios en ellas. Además, tiene un sistema de puntuación por estrellas, similar al utilizado en repositorio GitHub o incluso el que utiliza Android en sus aplicaciones. Esto permite en el caso de imágenes con similares funcionalidades, seleccionar a la imagen mejor posicionada.

El repositorio de Docker permite probar nuevas versiones de las aplicaciones publicadas, o buscar nuevas aplicaciones que sirvan para un propósito dado. Las imágenes de Docker son una forma fácil de experimentar sin interferir la configuración actual de la computadora, ni aprovisionar una máquina virtual y no tener que preocuparse por los pasos de instalación. Además, permite la centralización en un único canal, facilitando compartir públicamente la imagen entre diferentes integrantes [10].

3.3 Imagen de Docker Creada

Acorde a los objetivos antes mencionados, se creó una imagen Docker que ya dispone el entorno utilizable, evitando el proceso de instalación y configuración de las herramientas necesarias para su funcionamiento. De manera tal, que el desarrollador

[1] URL Docker Hub: http://hub.docker.com

pueda de forma rápida y sencilla emular sus programas para STM32, en un entorno a través de Qemu. Para esto, este trabajo se basó en el proyecto de *Beckus* [4], adaptándolo a los objetivos planteados. En este sentido, se modificó el código fuente de Qemu, para que pueda emular funcionalmente las placas *Stm32-p103*, *Stm32-Maple* y *Stm32-f103c8t6* (conocida como *BluePill*). Siendo publicado en un repositorio propio de Github, junto con los códigos modificados de los ejemplos de las placas antes mencionadas. Por esa razón, la siguiente es la dirección web de dicho repositorio.

https://github.com/soaunlam2021/soa-entorno-integracion **(1)**

Este se encuentra estructurado en tres directorios, de la siguiente manera:

- **Qemu_stm32:** Este directorio contiene el código fuente del emulador Qemu, que fue adaptado para esta investigación.
- **Stm32_demos:** Contiene diversos ejemplos de código fuente que fueron adaptados, para poder ser ejecutados en la emulación y en el hardware real de las placas mencionadas. A su vez contiene información acerca del hardware de dichos SE.
- **Workflows:** En donde se encuentra la configuración de la generación automática de la imagen en Docker Hub.

El método que se utiliza para construir la imagen de Docker de forma automática es a través del archivo *docker-image.yml.* Este archivo se configuró, para que se ejecuten automáticamente ciertos comandos, que se encuentran indicados dentro un archivo llamado *Dockerfile.* Esta secuencia de comandos, son ejecutados dentro de una máquina virtual (VM) que se encuentra en los servidores de Github, y son accionados cada vez que se actualiza el repositorio. Una vez que dentro de la VM se genera la imagen de Docker, automáticamente el archivo *yml* la carga y la publica dentro del registro de contenedores de Docker Hub. Para que de esta forma cualquier persona pueda descargarla y utilizarla fácilmente. Por consiguiente, la imagen publicada en esta investigación se puede encontrar desde línea de comando de la siguiente manera, mediante el nombre *soaunlam/qemu_stm2*:

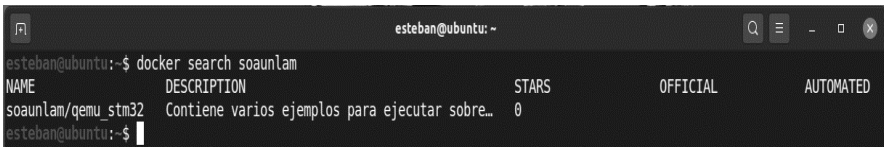


Fig. 1 Búsqueda de la Imagen Docker generada

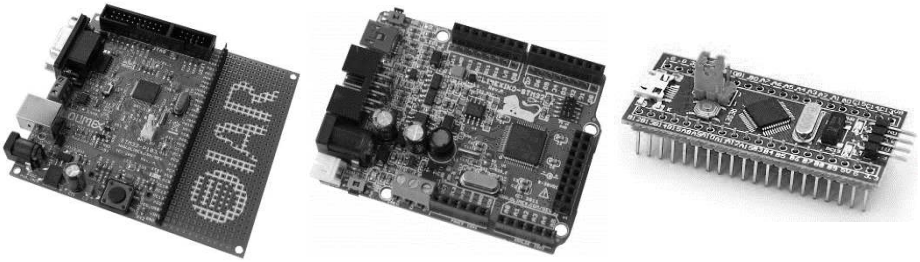
La ventaja de haber utilizado los *workflows* de Github, a través del archivo *yml,* es que la creación de la imagen se realiza en los servidores de Github y no en la máquina local. Con lo cual, esto nos evita tener que realizar el trabajo manualmente y el tiempo de procesarlo localmente.

3.4 Contenido del Archivo Dockerfile

Los archivos *Dockerfiles* permiten generar imágenes personalizadas. En estos archivos se especifican los comandos, en forma de meta instrucciones, que Docker interpretará para construir la imagen deseada. El archivo Dockerfile correspondiente a este trabajo se encuentra publicado en el repositorio de esta investigación. El contenido de dicho archivo fue organizado, para usar la menor cantidad de capas resultantes que conforman la imagen, como es explicado en la sección de mejores prácticas [10]. Para ello en el Dockerfile, se especifica como capa base a la versión adaptada del Sistema Operativo Ubuntu 20.04. Se consideró conveniente, utilizar una versión determinada y no la última existente, debido a que entre distintas versiones pueden variar sus dependencias, generando una imagen de Docker defectuosa. Luego se instalaron las dependencias y bibliotecas de Linux que necesita Qemu, junto con programas adicionales. Algunos de ellos son *apt-utils, gcc-arm-none-eabi, gcc, python2.7, pkgconf, git, make, libglib2.0-dev, libpixman-1-dev,* entre otros. Es importante mencionar que, para el correcto funcionamiento de algunos ejemplos, fue necesario instalar las dependencias *open-ssh* y *net-tools,* dado que estos deben ser ejecutados en parte a través de terminales ssh. Una vez instaladas todas las dependencias, se borra la cache utilizada para su instalación, de manera de poder liberar espacio en la imagen generada. Luego se descarga el repositorio de esta investigación desde Github (1). Posteriormente, dentro se configura y compila el código de Qemu, que se encuentra dentro del directorio *Qemu_stm32*. A su vez también se compilan y se generan los archivos binarios de los distintos ejemplos. Seguidamente se configura la imagen para que el usuario pueda acceder al contenido de este a través de protocolo SSH. Para ello se modifican los archivos *sshd_config* y *.bashrc* para ello. De esta forma se genera la imagen de Docker con todo configurado e instalado, para que el desarrollador pueda ejecutar ejemplos de código emulados en Qemu rápidamente. También para que tenga la posibilidad de modificar su código accediendo a estos de forma externa, a través de SSH. De tal manera que pueda acceder a los archivos contenidos dentro de la imagen, a través de distintos programas, como por ejemplo editores de código.

3.5 Diferencias entre las placas soportadas por la imagen de Docker

Las tres placas *Stm32-p103*, *Stm32-Maple* y *Bluepill,* Fig. 2 (a, b y c), pertenecen a la misma serie denominada “convencional” de microcontroladores de STM32F1 con 32 bits. Esta arquitectura se encuentra bien equilibrada y se adapta a las necesidades básicas que se esperan en los mercados de consumo, donde las limitaciones de costos y el tiempo de comercialización son esenciales. Además, están diseñadas para responder a los requisitos de forma simple, robusta y con larga vida de utilidad.



(a) Stm32-p103 (b) Stm32-Maple (c) BluePill

Fig. 2 Modelos de placas soportados por esta investigación

Estas placas de desarrollo tienen de similitud, el voltaje de trabajo (2.0V a 3.6V), su rango de temperaturas (-40 °C a 105 °C) y su velocidad de funcionamiento (hasta 72MHz). No obstante, sus diferencias radican en su tamaño y sus interfaces de entrada y salida. La placa Stm32-p103 [11], tiene las dimensiones de 100 x 90 mm. Además, internamente posee un botón y dos LEDs, uno de estado y otro de encendido. Como interfaces externas posee un conector JTAG, 3 puertos USART, 2 conectores SPI, 2 del tipo I2C, un puerto CAN, un conector USB y 2 conectores ADC. Mientras que la placa Stm32-Maple [12],posee las dimensiones de 69 x 54 mm. A su vez, internamente posee 2 botones, uno de reset y otro programable. Además posee 3 LEDs, dos de estado y uno de encendido. Como interfaces externas tiene un pin de conexión SWD, un conector de UEXT, otro de PWR JACK y distintos puertos CON1-POWER, CON2-ANALOG, ON3_DIGITAL y CON4 -DIGITAL. Además, posee un conector LI BAT, USB, un conector para tarjeta SD/MMC y un puerto CAN. Por otro lado, la placa BluePill [13]es la más pequeña de todas, con tan solo 23 x 53 mm. Además, internamente posee un botón de reset y contiene dos LED, uno de estado y otro de encendido. Adicionalmente, como conexiones externas posee un puerto USB, un pin de conexión SWD, 2 conectores del tipo I2C, 2 de SPI, un puerto CAN, un conector JTAG y 2 ADC.

3.6 Código fuente del contenido de la Imagen

Como se mencionó anteriormente, este trabajo se basó en el proyecto de *Beckus* y se lo adaptó para que pueda funcionar correctamente, en las emulaciones de las placas previamente mencionadas. En este apartado se describen brevemente las modificaciones que se debieron realizar en dicha adaptación. Estos ajustes fueron tanto en el código fuente base de Qemu y en los ejemplos. Dado que estos fueron creados para funcionar emulando solamente la placa *stm32-p103*, por lo que no estaban preparados para funcionar en su totalidad en las placas *stm32-Maple* y *BluePill*. Esto se debió principalmente a que los microcontroladores de esas placas poseen diferentes configuraciones internas. Por ese motivo en el repositorio de git (1), se crearon tres subdirectorios diferentes. Los cuales contienen los ejemplos de código fuente de programas, para ejecutar en cada una de las placas mencionadas. Ya que entre ellos presentan pequeñas modificaciones que lo hacen funcional. En este sentido, una de las adaptaciones que se debió

realizar, consistió en la asignación de la USART[2], la cual es diferente dependiendo del hardware que se esté utilizando. Los puertos de las USART en los sistemas embebidos resultan de vital importancia, dado que a través de ellos el desarrollador puede realizar una depuración indirecta de los programas que se ejecute en dicho hardware. La gran mayoría del código fuente de los ejemplos que se encuentran en la imagen de Docker creada, utilizan este mecanismo de depuración. Otra de las modificaciones que fue necesario realizar, fue la adaptación del manejador de interrupciones, dado que se mapea diferente dependiendo de la placa emulada. Además, se adaptó la conexión interna en determinados pines, como por ejemplo el LED de testeo. Por otra parte, se modificó el código fuente de Qemu, para poder generar la emulación de los eventos de pulsadores externos, en las placas *stm32-Maple* y *BluePill*. Estos eventos se crearon, de forma tal que se generen cuando el usuario envié un comando *sendkey,* desde la consola de Qemu, al ejecutar cualquier programa en él. De esta forma se podrá emular la acción de un actuador del tipo pulsador. Adicionalmente se adaptó el código fuente, para que la utilización de los registros de hardware sea a través de la utilización de funciones de bibliotecas. Las cuales corresponden al funcionamiento del hardware pertinente. Al mismo tiempo, se generó un mecanismo de compilación que permite generar los binarios de todos los ejemplos.

3.7 Ejecución del emulador Qemu dentro de la Imagen Docker

Cuando el usuario descargue la imagen de Docker de este trabajo, del repositorio Docker Hub, deberá asociarle un contenedor para poder trabajar con ella. Para ello una de las formas de uso, es a través del comando "`docker run -it`". El cual crea un contenedor, asociado a una pseudo-terminal interactiva, la cual permite interactuar por medio de línea de comandos. Esto se puede visualizar en la Fig. 3. Dependiendo de lo que desee hacer, desde dicha terminal se podrá ejecutar cualquiera de los ejemplos, que se encuentran dentro de alguno de los subdirectorios: *Stm32-p103, Stm32-Maple y BluePill*. Los ejemplos de código fuente que se encuentran en estos directorios, permiten entre otras cosas: emular el encendido y apagado de un led de testeo, emular un programa que trabaje con un pulsador, emular un sensor que trabaje con valores analógicos, hacer programas que emulen interrupciones por software y hardware, emular el trabajo del temporizador que posee cada placa, poder emular el trabajo de los puertos USART para la depuración remota indirecta y permitir trabajar con programas que funcionen con el Sistema Operativo de Tiempo Real (*FreeRTOS*). En esta última opción, se permite ejecutar, en los sistemas embebidos emulados, programas que funcionan en un único o múltiples hilos de ejecución.

[2] USART: Dispositivo que controla los puertos y dispositivos serie. Se encuentra integrado en la placa base o en la tarjeta adaptadora del dispositivo.

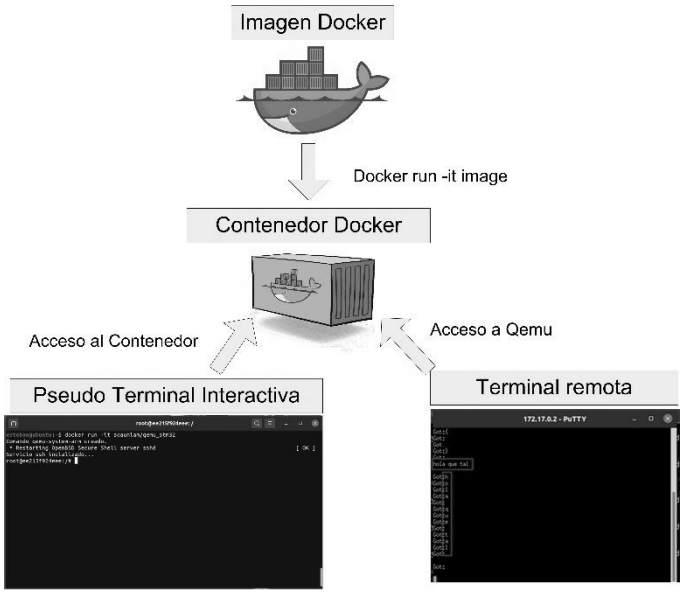


Fig. 3 Forma de ejecución de programas emulados en Qemu dentro del contenedor Docker

Para poder usar los binarios de los ejemplos en Qemu, que son generados cuando se compilan en el momento en que se genera la imagen, se debe ejecutar el emulador de una forma específica. Esta puede variar dependiendo del ejemplo que se quiera ejecutar, pero en la mayoría de los casos se debe seguir una forma base estándar de secuencia de pasos. Primero, el usuario deberá ejecutar el programa Qemu de la siguiente manera, desde la línea de comandos de la pseudo-terminal.

```
qemu-system-arm -M <nombre de la placa> -kernel <ar-
chivo.bin> -serial tcp::7777,server -nographic
```

El comando **qemu-system-arm,** es el archivo ejecutable del programa Qemu. Con el parámetro **-M,** se indica el modelo de placa que se desea emular. En nuestro caso, se puede emular las placas *stm32-p103, stm32-maple* y *stm32-f103c8.* Luego con la opción **-kernel,** se le indica la ubicación del programa binario que se desea ejecutar dentro del sistema embebido emulado. Adicionalmente con la opción **-serial,** se le dice al emulador que todas las entradas y salidas que se envíen al puerto serial de las placas sean redirigidas al puerto **7777** usando el protocolo TCP. Finalmente, con la opción **-nographic,** se deshabilita las salidas gráficas que genera el emulador.

Cuando se inicia la ejecución de Qemu, el emulador se quedará esperando una conexión externa a través del puerto 7777 del contenedor. Para ello el usuario deberá utilizar otra terminal remota y conectarse al puerto anteriormente indicado, utilizando la dirección IP que tenga asignado el contenedor de Docker, mediante el protocolo Telnet. Esto se muestra en la Fig. 4. Una vez que se establece la conexión, el programa emulado continuará su ejecución normalmente. Como ya se mencionó, la forma de ejecución puede variar, dependiendo del ejemplo que se desea ejecutar. Por ese motivo,

en esta investigación se generó un instructivo en forma de tutorial. En donde se detallan los pasos que se deben seguir, para poder ejecutar cada uno de los ejemplos disponibles dentro del entorno del contenedor. Este tutorial se encuentra dentro del repositorio generado (1), y por consiguiente, también se encuentra dentro de la imagen Docker generada.

En muchos de los ejemplos que se encuentran dentro de la imagen de Docker, se realiza depuración indirecta en forma remota a través de los puertos seriales. Como consecuencia de que el puerto serial es utilizado, tanto para mostrar datos en una terminal remota, como para ingresarlos a través de ella. Esto se puede visualizar en la siguiente figura, donde se muestra un caso de datos de entrada y salida a través de una terminal de este tipo.

Fig. 4 Terminal remota con datos de entrada y salida a través del puerto serial

4 Conclusiones

En este trabajo se presentó una alternativa, para que los desarrolladores de soluciones de sistemas embebidos, que pueden ser utilizados para IoT, realicen pruebas en placas STM32 emuladas a través del programa Qemu. Este trabajo les puede ayudar a los desarrolladores, establecer si determinado hardware le es o no de utilidad para sus proyectos, sin necesidad de adquirir el hardware físico. De manera que lo pueda realizar de forma rápida y sencilla, sin tener que realizar tediosas instalaciones y configuraciones de programas. Debido a que el emulador Qemu se encuentra empaquetado, configurado y automatizado dentro de una imagen Docker, fácil de emplear. Si bien este trabajo se centró en la emulación de tres placas: *stm32-p103, stm32-Maple* y *stm32-f103c8,* se planea en un futuro realizar el mismo trabajo de automatización, configuración y emulación mediante contenedores Docker para otros tipos de placas de desarrollo.

5 Referencias

1. Rose, K., Eldridge, S., Chapin, L.: La Internet De Las Cosas - Una Breve Reseña. , Reston, United State (2015).
2. Valiente, W., Carnuccio, E., Volker, M., de Luca, G., Villca, R., Adagio Matías: Entorno de contenedores de emuladores que contienen sistemas embebidos. In: XXIII Workshop de Investigadores en Ciencias de la Computación. pp. 12–17 (2021).
3. Eclipse Foundation: https://eclipse-embed-cdt.github.io/debug/qemu/.
4. Beckus: http://beckus.github.io/qemu_stm32/.
5. Lovric, D., Olsson, C.: Virtual Controllers (Tesis de Maestría). Department of Automatic Control, Lund University, Sweden (2016).
6. Muñoz, J.F., Goenaga, I.M.: Ofera Project: Open Framework for Embedded Robot Applications, European Union's Horizon 2020, Unión Europea (2019).
7. Amamory: https://hub.docker.com/r/amamory/qemu-stm32.
8. Chelladhurai, J.S., Singh, V., Raj, P.: Learning Docker - Second Edition: Build, ship, and scale faster. Packt Publishing, Birmingham, Reino Unido (2017).
9. Miell, I., Sayers, A.H.: Docker in practice. Manning Publications Co., Shelter Island, NY (2019).
10. Goasguen, S.: Docker Cookbook: Solutions and Examples for Building Distributed Applications. (2015).
11. Olimex: STM-P103 development board - User's manual., Plovdiv, Bulgaria (2016).
12. Olimex: OLIMEXINO-STM32 development board - Users Manual. , Plovdiv, Bulgaria (2011).
13. STMicroelectronics: STM32F103x8 DataSheet. (2015).

ISBN 978–987–633–574–4

Service Proxy with Load Balancing and Autoscaling for a Distributed Virtualization System

Pablo Pessolani, Marcelo Taborda and Franco Perino

Department of Information Systems Engineering
Universidad Tecnológica Nacional – Facultad Regional Santa Fe
Santa Fe, Argentina
{ppessolani, mtaborda, fperino}@frsf.utn.edu.ar

Abstract. Cloud applications are usually composed by a set of components (microservices) that may be located in different virtual and/or physical computers. To achieve the desired level of performance, availability, scalability, and robustness in this kind of system is necessary to describe and maintain a complex set of infrastructure configurations.

Another approach would be to use a Distributed Virtualization System (DVS) that provides a transparent mechanism that each component could use to communicate with others, regardless of their location and thus, avoiding the potential problems and complexity added by their distributed execution. This communication mechanism already has useful features for developing distributed applications with replication support for high availability and performance requirements.

When a cluster of backend servers runs the same set of services for a lot of clients, it needs to present a single entry-point for them. In general, an application proxy is used to meet this requirement with auto-scaling and load balancing features added. Autoscaling is the mechanism that dynamically monitors the load of the cluster nodes and creates new server instances when the load is greater than the threshold of highest CPU usage or it removes server instances when the load is less than the threshold of lowest CPU usage. Load balancing is another related mechanism that distributes the load among server instances to avoid that some instances are saturated and others unloaded. Both mechanisms help to provide better performance and availability of critical services.

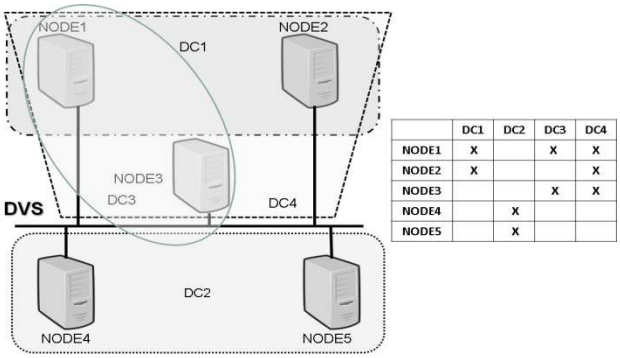
This article describes the design, implementation, and testing of a service proxy with auto-scaling and load balancing features in a DVS.

Keywords: Autoscaling, Load Balancing, Distributed Systems.

1 Introduction

Nowadays, applications developed for the cloud demand more and more resources, which cannot be provided by a single computer. To increase their computing and storage power, as well as to provide high availability and robustness they run in a

distributed environment. Using a distributed system, the computing and storage capabilities could be extended to several different physical machines (nodes). Although there are various distributed processing technologies, those that offer simpler ways of implementation, operation, and maintenance are highly valued. Also, technologies that provide a Single System Image (SSI) are really useful because they abstract the users and programmers from issues such as the location of processes, the use of internal IP addresses, TCP/UDP ports, etc., and more importantly, because they hide failures by using replication mechanisms. A Distributed Virtualization System (DVS) is an SSI technology that has all these features [1]. A DVS offers distributed virtual runtime environments in which multiple isolated applications can be executed. The resources available to the DVS are scattered in several nodes of a cluster, but it offers aggregation capabilities (allows multiple nodes of a cluster to be used by the same application), and partitioning (allows multiple components of different applications to be executed in the same node) simultaneously. Each distributed application runs within an isolated domain or execution context called a Distributed Container (DC). Fig. 1 shows an example of a topological diagram of a DVS cluster.



	DC1	DC2	DC3	DC4
NODE1	X		X	X
NODE2	X			X
NODE3			X	X
NODE4		X		
NODE5		X		

Fig. 1. Illustration of a DVS topology

A problem that must be considered when using a distributed application refers to the location of a certain service used by an external or internal client, or by another component of the distributed application itself. One way to solve this problem would be to use existing Internet protocols. With the DNS protocol, the IP address of the server can be located in the IP network, and with ARP the MAC address of the server can be located within a LAN. However, one issue that must be taken into account when working with a cluster is that the network and its nodes may fail, preventing continuity in the delivery of a given service.

When a cluster of backend servers runs the same set of applications for a lot of clients, it needs to present a single entry-point for their services. In general, an Application Proxy (AP), also known as Reverse Proxy, is used to meet this requirement with auto-scaling and load balancing features added. Autoscaling is the mechanism that dynamically monitors the load of the backend servers and creates new server instances when the load is greater than a threshold of highest CPU usage (*high_CPU*) or it removes server instances when the load is lower than a threshold of lowest CPU us-

ISBN 978–987–633–574–4

age (*low_CPU*). Load-balancing is another related mechanism that distributes the load among backend server instances to avoid that some instances are saturated and others unloaded. Both mechanisms help to provide better performance and availability of critical services. This technology is widely used in certain scenarios such as web applications, where end-users send requests from their devices as clients, and the SP is the component that establishes sessions with backend servers, thus distributing, balancing, and orchestrates services between internal services and microservices [2]. To distinguish between these two usages scenarios, in this article Application Proxy (AP) refers to the former, and Service Proxy (SP) refers to the latter.

This article presents the design, implementation, and experimentally proves the capabilities of an SP with autoscaling and load balancing features for a DVS. Therefore, the project focused on building an operational prototype of an SP for the DVS (not a commercial-class one), relegating performance and high availability improvements for future works.

The rest of the article is organized as follows: Section 2 refers to related works. Section 3 provides an overview of background technologies and Section 4 describes the design and implementation of the SP for the DVS (referred to as DVS-SP). Section 5 presents the tests for the SP performance evaluation and finally, the conclusions and future works are summarized in Section 6.

2 Related Works

APs are not unknown by the scientific community, so a lot of research and development works have previously been carried out, but for IP environments. Therefore, only those with the most important features and more popular [3,4] are presented here for space reasons.

2.1 NGINX

A very popular HTTP server and reverse proxy is NGINX [5]. It is free, open-source, and well known for its high performance, stability, rich feature set, and low resource consumption.

NGINX can handle tens of thousands of concurrent connections and provides caching when using the *ngx_http_proxy_module* module and supports load balancing and fault tolerance. The *ngx_http_upstream_module* module allows for de *nginx groups* of backend servers to distribute the requests coming from clients.

2.2 HAproxy

HAproxy [6] (stands for High Availability Proxy) is an HTTP reverse-proxy. It is a free, open-source, reliable, high-performance load balancer and proxying software for TCP and HTTP-based applications. It is also an SSL/TLS tunnel terminator, initiator, and off-loader, and provides HTTP compression and protection against DDoS. It can

handle tens of thousands of concurrent connections by its event-driven, non-blocking engine.
HAproxy was designed for high availability, load balancing and provides redirection, server protection, logging, statistics, and other important features for large-scale distributed computing systems.

3 Background Technologies

This section presents the products and tools that have been studied and analyzed as technological support for the design and implementation of the DVS-SP prototype.

3.1 M3-IPC

The DVS provides programmers with an advanced IPC mechanism named M3-IPC [7] in its Distributed Virtualization Kernel (DVK) which is available at all nodes of the DVS cluster. M3-IPC provides tools to carry out transparent communication between processes located at the same (local) node or in other (remote) nodes. To send messages and data between processes of different nodes, M3-IPC uses Communications Proxies (CPs) processes. CPs act as communication pipes between pairs of nodes.

M3-IPC processes are identified by *endpoints* that are not related to the location of each process, and then it does not change after a process migration. This feature becomes an important property that facilitates application programming, deployment, and operation. An *endpoint* can be allocated by a process or by a thread and must be unique in each DC.

M3-IPC supports message transfers (which have a fixed size) and blocks of data between endpoints. If the sender and receiver endpoints are located in the same node, the kernel copies the messages/data between the processes/threads which own the endpoints. If the sender and receiver are located in different nodes, CPs are used to transfer messages and data between nodes, and the DVK of both nodes copies those messages/data between the CPs and the processes/threads.

3.2 Group Communication System (GCS)

To exchange information between a group of processes that run on several nodes or in the cloud, communication mechanisms with characteristics such as reliability, fault tolerance, and high performance are required. Several tools offer these features such as Zookeeper [8], Raft [9] or the Spread Toolkit [10].
The Spread Toolkit was chosen for the DVS-SP development because it is a well-known GCS used by the authors' research group in other projects. On Spread Toolkit, two kinds of messages are distinguished. Regular messages: sent by a group member, and; Membership messages: sent by the Spread agent running on each node.

Regular messages can be sent by members using the provided APIs for broadcast (multicast) them to a group. However, unicast messages could be sent to a particular

ISBN 978–987–633–574–4

member. Membership messages are sent by Spread to notify members about a membership change, such as the joint of a new member, the disconnection of a member, or a network change. Network changes can be the crash of a node (or a set of nodes), a network partition, or a network merge after a partition.

Spread provides reliable delivery of messages (even in the event of network or group member failures) and the detection of failures of members, or the network. It also supports different types of ordering in message delivery such as FIFO, Causal, Atomic, etc. making it an extremely flexible tool for the development of reliable distributed systems.

Spread Toolkit is based on the group membership model called Extended Virtual Synchrony (EVS) [11], tolerating network partition failures and network merge, node failures, process failures and restart.

4 Design and Implementation of the DVS-SP

The design of the DVS-SP started proposing its architecture, describing its components and the relations between them. The active components are (Fig. 2):

- The Main Service Proxy (MSP) reads a configuration file, initializes all data structures, and starts the other components threads.
- For each Frontend Client Node (specified in the configuration file), a pair threads are started for the CPs. The Client Sender Proxy (CSP) thread sends messages from the DVS-SP to the Client node. The Client Receiver Proxy (CRP) thread receives messages from the Client node.
- Similarly, for each Backend Server Node, a pair threads are started for the CPs. The Server Sender Proxy (SSP) thread sends messages from the DVS-SP to the Server node. The Server Receiver Proxy (SRP) thread receives messages from the Server node.
- A Load Balancer Monitor (LBM) thread receives notifications about changes in the load state from the Load Balancer Agents (LBA) running on each backend server node.
- Each Client and Server node uses the Node Sender Proxy (NSP) and the Node Receiver Proxy (NRP) to communicate with the DVS-SP proxies.

Proxy messages differ from application messages. Proxy messages are the transport of single application messages (like a tunnel), a batch of application messages, a block of raw data, an acknowledgment message, or a proxy HELLO message. The reader should consider that this architecture was not designed to serve user applications such as web browsers as clients. It should be used among application services, i.e. web servers (Clients) which need to read/write files from/to network filesystems (Servers).

4.1 Main Service Proxy (MSP)

As was mentioned earlier, the MSP reads the configuration file which describes the cluster, initializes all data structures, and starts the other components threads.
In the configuration file four types of items are specified:

a. *MSP*: specifies the node name where the MSP runs, the node ID, and the high-water and low-water load levels.
b. *Server*: specifies the server node name and its node ID.
c. *Client*: specifies the client node name and its node ID.
d. *Services*: describes the name of the service (i.e. fileserver), the external endpoint (*ext_ep*) in which the SP will receive requests from clients, the lowest (*low_ep*) and highest (*high_ep*) endpoint numbers which servers could use to serve the requests, and eventually the pathname of a server program to run on a server node.

Services could be running on server nodes (persistent service) or they could be started when the MSP receives a new request from a client (ephemeral service). Therefore, the MSP creates a *Session* for each pair of client-server processes. Once the MSP detects that a new client requests the same service on the same node using the same pair of endpoints, it removes the old *Session* from its database and terminates the old server process. Afterward, it creates a new *Session* for the new pair of processes. This behavior is a piece of the auto-scaling mechanism of the DVS-SP. A session is defined by the following tuple: {*dcid, clt_ep; clt_node, clt_PID, svr_ep, svr_node, svr_PID*}.

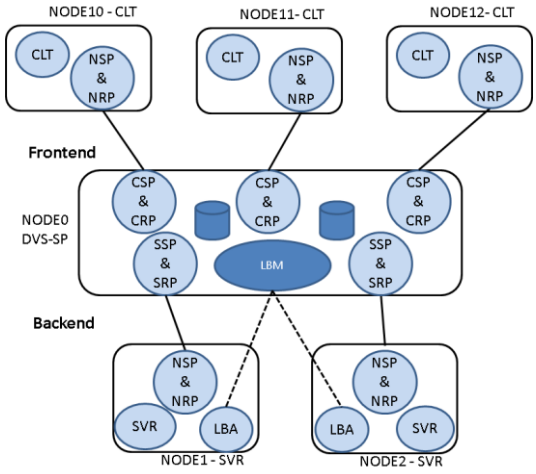


Fig. 2. DVS-SP Architecture.

As the LBM manages the load database of all Servers, the DVS-SP scheduling module can decide when it is time to allocate another Server node for new sessions or when it is time to dismiss it.

4.2 Client Receiver Proxy (CRP) and Server Sender Proxy (SSP)

When a new session starts, the client sends a message to the external endpoint (*ext_ep*) of the DVS-SP through its NSP to the CRP. The CRP compares several fields of the sessions to find an active session that matches. If it does not exist, it searches the server's database for the first non-saturated server node (server load < *high-water*) and allocates it for that session. The reader might ask: why not choose the server with the least load?. That policy would go against Autoscaling because an unloaded server quickly could acquire more work and could not be removed from the cluster (scaling-down).

If a program is specified for that service, the CRP sends a remote command to the server's node to execute the server program. Then, the proxy message is forwarded to the server NRP.

4.3 Server Receiver Proxy (SRP) and Client Sender Proxy (CSP)

When an SRP receives a message from its server's NSP it checks if a session exists. If all the session's parameters match but the server's PID, it removes it as an expired session. If a program was specified for that service, it sends a remote command to the server's node to terminate the server process. If the server's PID also matches, it gets the client endpoint field of the session and queues the proxy message into the CSP message queue. As the CSP is waiting for messages in its queue, it forwards the message to its Client NRP.

Several endpoint conversions are done into the header of proxy messages on CRP and SRP to hide the real architecture from clients and servers. Clients only request the DVS-SP as their single server, and servers only reply to the DVS-SP as their single client (service proxy behavior).

4.4 Load Balancer Monitor (LBM)

The LBM collects information about the load level of server nodes. The Load Balancer Agents (LBA), report their node load levels when they change. The load levels are defined as *LVL_UNLOADED, LVL_LOADED, and LVL_SATURATED.*
The LBM manages and keeps updated the node status database used by CRPs to allocate servers for new sessions. When a server node fails (reported by the GCS), the LBM deletes all the sessions with that node. When the load level of all active servers is *LVL_SATURATED* during a specified *START_PERIOD*, the LBM commands the hypervisor to start a new node (scaling-up). If a server node has no active sessions during a specified *SHUTDOWN_PERIOD*, it will be shut down (scaling-down).

4.5 Load Balancer Agents (LBA)

LBA periodically evaluates the load of its node. Currently, the load of a node is defined as the mean of CPU usage (reported by the pseudo-file */proc/stat*) in the specified *LBA_PERIOD* period. Although there are additional metrics that could be con-

sidered to describe the load of a node [12], such as memory usage, network traffic, disk I/O, etc., only CPU usage was considered to simplicity the prototype implementation.

Each server node keeps a *load_lvl* variable that stores its load level. In each period, if the new load level value differs from *load_lvl*, the LBA reports this new level to the LBM using the GCS, and updates *load_lvl*. Therefore, the dissemination of load information is event-driven. This mechanism consumes lower network bandwidth than a periodic one [13]. In the case that the LBM is doesn´t alive or is unreachable, the LBA doesn't report any message. Then, when LBM comes back, the LBA starts to report the load level again.

5 Evaluation

This section describes the tests and micro-benchmarks used to verify the correct operation of the DVS-SP in a DVS virtual cluster. It should be considered that the tests should have been carried out in a home virtualized environment and not a physical environment as a consequence of the inability to access the laboratories during 2020 and 2021 due to the regulations established by the national government in relation to COVID-19. This fact does not imply important consequences to demonstrate the correct behavior of the DVS-SP, but for performance measurements. It's known that CPU, memory, disk, network virtualization could distort the results.

The hardware used to perform the tests was a PC with a 6-core/12-threads AMD Ryzen 5 5600X CPU, 16 GBytes of RAM, and SATA disks. The virtualization was carried out using VMware Workstation version 15.5.0 running on Windows 10 and a cluster of 6 nodes was configured, each node in a VM: NODE{0-5}. Each VM was assigned a vCPU and 1 GB of RAM. The VMs were clones of each other running Linux kernel 4.9.88 modified with the DVK module. The DVS-SP runs on NODE0; servers run on NODE{1,2}, and clients run on NODE{3-5}. This virtual cluster only was used to test the correct behavior of the DVS-SP on allocating new sessions to new servers when the other servers are saturated and, exchange load information among the LBAs and the LBM and to test fault-tolerance on server crashes, node failures, or network partitions.

To evaluate the DVS-SP performance (taking into account the previously mentioned test environment) a minimal cluster of 3 nodes was used: DVS-SP run in NODE0, NODE1 was the server node, and NODE2 was client node. Two main metrics were measured:

- *Latency*: Two programs were used, a latency client and a latency server. The server program waits for a request and, when it receives one, it replies to the client. The client sends a request and then it waits for the reply, measuring the elapsed time among these two events. Another derived metric of these tests was the message transfer throughput.
- *Data transfer throughput*: A file transfer pair of programs was used. The client can request a GET operation to transfer a file from server to client, or a PUT operation to transfer a file from client to server. The server measures the time between the

first request received from the client and the last message sent to it then, it calculates the throughput.

In Fig. 3(A), the relative latency to a Local Ping is presented, where:

- *Local Ping:* Ping to the *localhost* interface address (average 0.036 ms).
- *Local Ping:* HTTPing to a web server on the same node.
- *Local Latency*: The client and server programs were executed on the same node.
- *Remote Ping*: Ping to Ethernet interface address of another node.
- *Remote HTTPping*: HTTPing to a web server on another node.
- *Remote Latency:* The client latency program was run on one node and the server latency program was executed on another node.
- *DVS-SP Latency*: The communications between the client and the server traverses the DVS-SP.

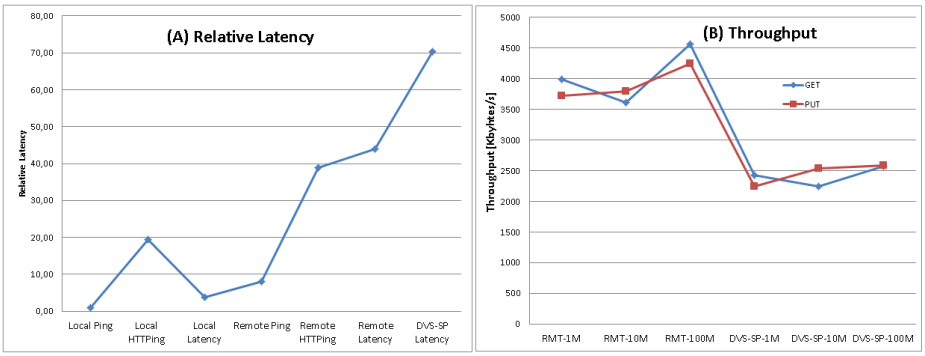


Fig. 3. (A) Relative Latency and (B) Data Transfer Throughput.

The use of remote DVS communications against a client/server HTTP communication imposes a latency penalty of 13%. Client/server using the DVS-SP adds 50% to the communication latency. This indicates that the DVS-SP is not suitable for use in high latency networks such as WANs or the Internet.

In Fig. 3(B), the data transfer throughput is presented for two scenarios: Client on one node and Server on another node (RMT) and then, with the DVS-SP between them. Three files sizes were used for transfers: 1, 10, and 100 Mbytes. Two well-known tools were used to compare file transfer performance; ***scp*** reports measurements from 13 to 20 [Mbytes/s] and ***wget*** from 10 to 44 [Mbytes/s] highlighting how resource virtualization affects performance metrics.

In the same way as latency, throughput is also affected by the use of both DVS and DVS-SP, however, there are still improvements to be made using data compression in data transfers planned for future works. Developers should consider these overheads as a tradeoff against the advantages and features provided by the DVS and the DVS-SP.

6 Conclusions and Future Works

A DVS provides scalability, reliability, and availability, it is simple to deploy and configure, and lightweight in terms of requirements which reduces the OPEX.

The contribution of this article is to present a Service Proxy with Load Balancing and Autoscaling features for a DVS as a proof of concept. Several tests were performed to demonstrate de SP capabilities as one of the several features that DVS architecture has. This DVS-SP uses a centralized and probabilistic approach to balance the computational loads and avoid backend server saturation.

The use of an AP is a common practice deploying Cloud Applications. However, a configuration with a single AP that centralizes all communications between clients and servers has in the AP a single point of failure and a performance bottleneck; therefore, reducing service availability and scalability. A future project will be to design and implement a DVS-SP cluster with replication support that can handle nodes and network failures and can tolerate higher performance demands.

For those architectures with clients and servers in the same network (in a trusted environment), a distributed Load Balancer with Autoscaling could be developed without an SP in the middle. The communications latency should be reduced and the single point of failure and performance bottleneck should be eliminated because each client will communicate with its allocated server directly.

References

1. P. Pessolani, T. Cortes, F. Tinetti, S. Gonnet: "*An Architecture Model for a Distributed Virtualization System*"; Cloud Computing 2018; Barcelona, España.2018.
2. M. Fowler. https://martinfowler.com/articles/microservices.html. Last accessed July 2021.
3. Usage statistics of Nginx. https://w3techs.com/technologies/details/ws-nginx. Last accessed July 2021.
4. Companies using HAproxy. https://enlyft.com/tech/products/haproxy. Last accessed July 2021.
5. Nginx. https://www.nginx.com/. Last accessed July 2021.
6. HAproxy. http://www.haproxy.org/. Last accessed July 2021.
7. P. Pessolani, T. Cortes, F. G. Tinetti, and S. Gonnet: "*An IPC Software Layer for Building a Distributed Virtualization System*", CACIC 2017, La Plata, Argentina, 2017.
8. Zookeeper. https://zookeeper.apache.org/. Last accessed July 2021.
9. Diego Ongaro, John Ousterhout. "*In Search of an Understandable Consensus Algorithm*", 2014 USENIX Annual Technical Conference. ISBN 978-1-931971-10-2. 2014.
10. The Spread Toolkit. http://www.spread.org. Last accessed July 2021.
11. L. E.Moser, et al., "*Extended Virtual Synchrony*", in Proceedings of the IEEE 14th International Conference on Distributed Computing Systems, Poznan, Poland, June 1994.
12. Siavash Ghiasvand, et al. "*An Analysis of MOSIX Load Balancing Capabilities*", International Conference on Advanced Engineering Computing and Applications in Sciences, November 20-25, 2011 - Lisbon, Portugal.
13. M. Beltrán and A. Guzmán, "*How to Balance the Load on Heterogeneous Clusters*," International Journal of High Performance Computing Applications , vol. 23, no. 1, pp. 99-118, Feb. 2009.

ISBN 978–987–633–574–4

Algoritmos para determinar cantidad y responsabilidad de hilos en sistemas embebidos modelados con Redes de Petri S^{3}PR

Ing. Luis Orlando Ventre[1], Dr. Ing. Orlando Micolini[1]

[1]Laboratorio de Arquitectura de Computadoras, FCEFyN-Universidad Nacional de Córdoba
Av. Velez Sarfield 1601,CP-5000, Córdoba, Argentina
{luis.ventre, orlando.micolini}@unc.edu.ar

Abstract. La evolución de la tecnología, el uso del IoT, y los requerimientos reglamentarios de la industria impactan en el diseño de sistemas embebidos convirtiéndolo en complejo y desafiante e imponiendo métodos formales para su desarrollo. Más aun considerando el reducido time-to-market, es determinante minimizar los tiempos de desarrollo. En este escenario los sistemas deberán ser concurrentes y seguros para aprovechar el rendimiento de las modernas arquitecturas multicore. Las Redes de Petri extendidas, son un reconocido y adecuado lenguaje de modelado, análisis y ejecución de sistemas reactivos, paralelos y concurrentes. Para potenciar los esfuerzos del modelado, se utiliza el modelo para obtener automáticamente parte de la implementación del sistema. En este trabajo se presenta un conjunto de algoritmos, a partir de un sistema modelado con Redes de Petri, para determinar automáticamente los hilos y responsabilidades de ejecución, esto tiene por objetivo mitigar los tiempos de desarrollo y reducir los errores de programación.

Keywords: Determinación automática de hilos, Redes de Petri, Generación de código, Sistemas embebidos, IoT.

1 Introducción

Actualmente los estándares de la Industria 4.0 [1] demandan la utilización de técnicas formales en el diseño de sistemas embebidos críticos, reactivos (RS) y dirigidos por eventos (EDA) [2]. Este escenario impone que el diseño cumpla con complejos requerimientos no funcionales ya que son sistemas multi-hilos, concurrentes e interactúan con variables y eventos del propio sistema y del mundo exterior, donde los datos y eventos son heterogéneos y no deterministas [3].

Determinar la secuencia de ejecución de estados consecutivos en los sistemas multi-hilos considerando sus combinaciones de ejecución con otros hilos, aun si estos son de estado finito, presentan un problema importante e intrínsecamente difícil de resolver. Así como la complejidad de determinar el número de hilos, las variables locales asociadas a los mismos, y las variables compartidas por estos, son también problemas poco explorados y determinantes de resolver en el diseño de RS y EDA.

Las fases del diseño de un sistema embebido incluyen el desarrollo del modelo basado en un conjunto de requerimientos [4]. Este modelo es el fundamento para otras etapas, incluida la etapa de codificación y testing de la aplicación [5]. Las redes de Petri (RdP) extendidas y no autónomas son un lenguaje de modelado de propósito general que admite el modelado de sistemas reactivos, concurrentes y paralelos independientemente de la plataforma.

En el diseño de RS y EDA, la transformación del modelo en software implica un trabajo de traducción, gestión del uso de recursos y determinación de la cantidad y responsabilidad de los hilos. Esto se logra a través de sucesivas iteraciones con el fin de solucionar errores potenciales de interpretación e implementación.

En este trabajo y con la finalidad de cerrar esta brecha se propone una metodología que parte de un sistema modelado con RdP del cual se obtienen sus invariantes de transiciones. Con estos componentes los algoritmos aquí presentados son responsables de analizar la RdP y determinar automáticamente el número de subprocesos activos máximos simultáneos, así como de determinar el número de subprocesos máximos requeridos y la responsabilidad de los mismos. Los algoritmos propuestos como metodología tienen la finalidad de reducir los tiempos de desarrollo, mitigar errores de codificación y contribuir en la generación automática de código; como así también, evaluar y gestionar la asignación de recursos en el sistema. Esta metodología es aplicable a sistemas modelados con una clase de RdP denominada Sistema de Procesos Secuenciales Simples con Recursos (S^3PR) la cual es adecuada para sistemas que comparten recursos y sincronización.

De acuerdo con nuestra investigación de documentos, no se han encontrado algoritmos con las funcionalidades para la determinación automática de la cantidad y responsabilidad de los hilos a partir del modelo del sistema realizado con una RdP. Como antecedente a este trabajo, y como aporte de este grupo de investigación, se puede encontrar el desarrollo de un Procesador de Petri (PP) modular que ejecuta RdP ordinarias en [6], una metodología para el desarrollo de sistemas embebidos basada en RdP [7] y el desarrollo de un PP extendido modular con un algoritmo para determinar la cantidad de hilos en ese procesador [8]. A diferencia de este último, en el actual trabajo se presentan tres algoritmos los cuales son de aplicación general para ser ejecutados sin la necesidad de un PP. Asimismo, en [9] se realizó un estudio sobre más de 70 referencias que hacen uso de la RdP para la solución de RS y EDA y no se encontraron metodologías similares a las aquí presentadas.

La siguiente sección presenta los objetivos de este trabajo; mientras que en la sección 3 se exponen la metodología y herramientas. Luego, en la sección 4, los algoritmos propuestos, mientras que en el apartado 5 se expone un caso de aplicación y los resultados, y finalmente en el apartado 6 las conclusiones y trabajos futuros.

2 Objetivos

En el modelo del sistema, realizado con una RdP, se encuentra explícita la lógica del sistema. La etapa de codificación transforma este modelo en software, lo que implica esfuerzos iterativos para interpretar, transcribir y refinar el modelo. Esto conlleva una carga de esfuerzo y tiempo en las etapas de desarrollo, que se pretenden mitigar.

El objetivo principal e innovador de esta propuesta es contribuir a la generación automática de código con la determinación de la cantidad y responsabilidad de los hilos en la ejecución del modelo, en concordancia con el paralelismo intrínseco del mismo; el cual tiene la capacidad de expresión de una máquina de Turing. Como así también, evaluar y gestionar el uso y la asignación de recursos en el sistema. Esta propuesta, mantiene todas las propiedades verificadas en el modelo, ya que se ejecuta la ecuación de estado extendida [10] del modelo. Para esto se utiliza un monitor como mecanismo de control de concurrencia. La importancia del uso de este mecanismo radica en desacoplar la lógica de la política y las acciones, lo que resulta en un sistema modular, simple, mantenible y verificable. Estas metodologías tienen como finalidad garantizar un diseño seguro, correcto, eficiente y robusto de los sistemas embebidos y sus aplicaciones.

3 Metodología y herramientas.

Existen varias herramientas de modelado, entre las que se encuentran: diagramas UML [11] y RdP [12, 13]. Los diagramas UML brindan las características necesarias, en parte, pero esencialmente carecen de mecanismos de verificación formal que garanticen estrictamente el cumplimiento de requisitos críticos, los cuales son aspectos fundamentales a ser implementados en la RS y EDA.

Dado que las RdP extendidas y no autónomas [14] permiten modelar la concurrencia, sincronización, el estado local y global y el paralelismo, es posible verificarlas formalmente, son ejecutables [9] y escalables cuando se expresan con la ecuación de estado extendida [10]; se ha considerado el formalismo más conveniente y se ha seleccionado como herramienta de modelado y lenguaje de ejecución.

3.1 Redes de Petri

Una RdP, denotada como *PN*, es una quíntupla [5] definida por:

$$PN = (P, T, I^+, I^-, M_0) \tag{1}$$

Dónde:

- $P = \{p_1, p_2, \dots, p_n\}$ es un conjunto finito, no vacío, de plazas.
- $T = \{t_1, t_2, \dots, t_m\}$ es un conjunto finito, no vacío, de transiciones.
- I^+, I^- son las relaciones de incidencia de salida y entrada de las plazas, la Matriz de Incidencia es:

$$I = I^+ - I^- \tag{2}$$

- $M_0 = [m_0(p_1), m_0(p_2) \dots, m_0(p_n)]$ es el marcado inicial de la red.

RdP sincronizadas o no autónomas. Las RdP Sincronizadas introducen eventos, y son una extensión de las RdP [14, 15].

Ecuación de estado. La ecuación de estado de una RdP, con n plazas y m transiciones con brazos con peso mayor o igual a uno y marca inicial M_0 es:

$$M_{j+1} = M_j + I * \sigma \quad (3)$$

Siendo: σ el vector disparo, con dimensión $m \times 1$. Cuando se dispara una transición sensibilizada, se puede calcular el siguiente estado usando la ecuación (3).

Conflicto entre transiciones. Los conflictos entre transiciones de una RdP sincronizada [14] ocurren cuando dos o más transiciones se encuentran sensibilizadas, sus eventos asociados suceden simultáneamente y el disparo de una de ellas desensibiliza la otra transición. Estos conflictos son resueltos con una política de prioridades.

S^{3}PR. Para definir una RdP S^{3}PR es necesario primero definir los conceptos de RdP S^{2}P y S^{2}PR. Se define una RdP S^{2}P como una red que modela procesos secuenciales simples; las RdP S^{2}PR, modelan procesos secuenciales simples con recursos; y finalmente las RdP S^{3}PR son la composición neta de RdP S^{2}PR a través de un conjunto de plazas comunes (recursos). Una explicación específica de las distintas subclases de estas RdP se encuentra en [16].

3.2 Monitor de Concurrencia

En los sistemas RS/EDA se necesitan mecanismos para organizar el acceso exclusivo a los recursos y para sincronizar y comunicar entre las tareas. Uno de los mecanismos más naturales, elegantes y eficientes para la sincronización y la comunicación, especialmente para los sistemas multicore, es el monitor [17].

En este desarrollo, el monitor, gestiona los eventos en exclusión mutua con el fin de determinar cuál acción ejecutar y cuando; para lo cual se basa en dos componentes, que son: la lógica y la política. Es importante destacar que no es su responsabilidad la ejecución de las acciones, dado que estas son ejecutadas por los hilos.

3.3 Arquitectura del sistema de la solución

Los componentes de la arquitectura del sistema (RS/EDA) que da soporte a los objetivos de este trabajo se observan en la Fig. 1 y son: eventos de software, eventos físicos, monitor de concurrencia, manejador de eventos, RdP (lógica), política, hilos y sus acciones asociadas. Esta arquitectura es modular, sus componentes están desacoplados y sus interfaces claramente definidas. De esta manera se simplifica su diseño, gestión y control, lo que la hace mantenible, refactorizable y asegura que el formalismo de la RdP se mantiene.

En la Fig. 1 se muestran las interacciones entre los componentes de la arquitectura y a continuación se describirán las responsabilidades de los principales componentes.

Manejador de eventos. Es el módulo encargado de recibir los eventos/estímulos del sistema y del exterior. Dado que los eventos son responsables de desencadenar una acción que depende del estado del sistema, es necesario que este módulo redireccione el evento recibido al hilo correspondiente.

Monitor. Gestiona el acceso de los hilos en exclusión mutua con el fin de determinar cuál acción ejecutar y cuando; para lo cual utiliza la lógica (RdP) y la política. Es importante destacar que no es su responsabilidad la ejecución de las acciones.

Red de Petri. Modela y representa la lógica del sistema. Es el mecanismo que utiliza el monitor para determinar a partir de los eventos y del estado del sistema, las

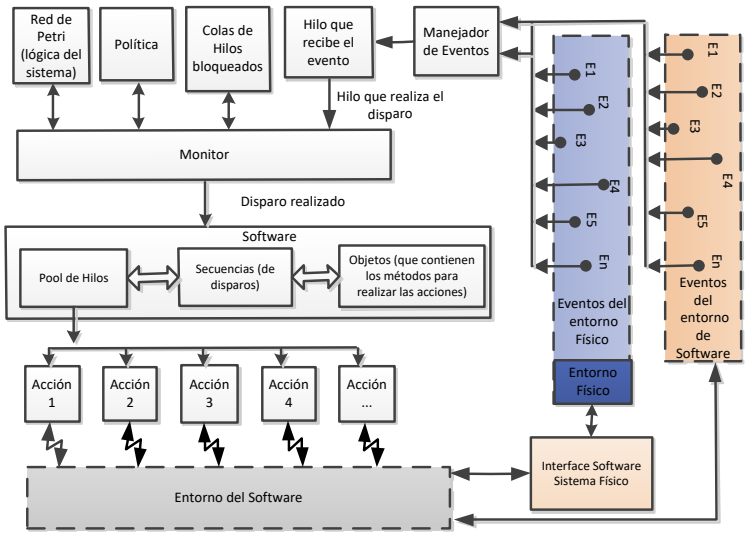


Fig. 1 Componentes de la arquitectura del sistema de la solución

acciones posibles a ejecutar por los hilos.

Política. Es el mecanismo que utiliza el monitor para solucionar los conflictos del sistema con un propósito específico. Decide entre los posibles acciones ejecutables, cual es la próxima a ejecutar.

Colas de hilos bloqueados. Su responsabilidad es bloquear los hilos, con las solicitudes de disparo de una transición requerida. Es necesario que cada cola sea asignada a una transición, puesto que cada transición opera como una variable de condición [17]

Acciones. Son las tareas a realizar por cada hilo de acuerdo al estado en que se encuentra el sistema.

Secuencia de disparos. Representa a las transiciones que se corresponden con los segmentos de ejecución determinados por los algoritmos propuestos en éste trabajo.

4 Algoritmos de la solución.

La hipótesis de este trabajo se basa en el hecho de que un modelo elaborado con una RdP no autónoma es un conjunto de instrucciones, ecuaciones y restricciones o reglas para generar el comportamiento de E/S de un sistema. Es decir, el modelo se describe

como estado transiciones y mecanismos para aceptar trayectorias de entrada y generar trayectorias de salida en función de su estado.

La definición, en términos de especificaciones del sistema, tiene la ventaja de una base matemática sólida y una semántica inequívocamente definida. Para especificar un comportamiento, el modelo necesita agentes. Se trata básicamente de un sistema informático capaz de ejecutar el modelo. El mismo modelo, expresado en formalismo, puede ser ejecutado por diferentes agentes, permitiendo así la portabilidad e interoperabilidad a un alto nivel de abstracción.

En este proyecto, los hilos son los agentes encargados de ejecutar el modelo haciendo uso de [10], para generar el comportamiento deseado.

En el modelo descripto con una RdP S^3PR los invariantes de transición se corresponden con los procesos para llevar a cabo un ciclo en el sistema por lo cual los algoritmos propuestos utilizan están propiedades estructurales para la determinación de la cantidad y responsabilidad de hilos de ejecución. Estos hilos ejecutan las transiciones consecutivas de cada invariante y adquieren los estados entre estas transiciones. Estos estados están asociados a las acciones que el sistema debe ejecutar.

4.1 Algoritmo para la determinación de hilos activos simultáneos.

1) Obtener los invariantes de transición (IT) de la RdP y para cada IT realizar:
2) Obtener el conjunto de plazas asociadas al IT en análisis.

$$PI_i = \bigcup_{\forall t \in Inv} \bullet t \ \cup \bigcup_{\forall t \in Inv} t \bullet$$

Dónde: PI_i representa el conjunto de plazas asociadas al iésimo IT.

3) Determinar las plazas relacionadas a acciones de cada IT. Para esto es necesario eliminar del conjunto PI_i, las plazas que son restricciones, recursos e idle:

$$PA_i = PI_i - \left\{ PI_{\text{restricciones}_i} \cup PI_{\text{recursos}_i} \cup PI_{idle_i} \right\}$$

Dónde: PA_i representa el conjunto de plazas de acciones asociadas al iésimo IT.

4) Del árbol de alcanzabilidad de la RdP, se debe obtener MA, el cual es el conjunto de todos los marcados posibles de todos los conjuntos de plazas PA_i.
5) De cada marcado posible (estado) del conjunto MA, se debe realizar la suma de las marcas. De todas estas sumas, se debe buscar la de mayor valor (marcado máximo). Esta será la cantidad máxima de hilos activos simultáneos en el sistema.

4.2 Algoritmo para determinar la responsabilidad de los hilos.

El algoritmo propuesto comienza con el análisis de la estructura de los IT de la red. La asignación de responsabilidad de ejecución de los invariantes varía de acuerdo a si estos son estrictamente lineales (secuenciales) o presentan forks (conflictos) y/o joins (uniones). En los casos donde el IT presente fork/join la responsabilidad de ejecución del invariante se fracciona en diferentes segmentos. Al igual que en el algoritmo anterior, los segmentos están compuestos por subconjuntos de plazas las cuales no incluyen recursos, restricciones ni plazas idle.

A continuación se analizaran cada uno de estos casos:

1) En el caso que la estructura de la red presente un IT lineal, es decir no comparte transiciones con ningún otro IT, la responsabilidad de ejecución de este IT es asignada a un segmento de ejecución.
2) En el caso de que dos (o más) invariantes compartan transiciones en conflicto estructural (fork). La responsabilidad de ejecución de los invariantes es segmentada. Esto tiene como ventaja que se elimina la lógica interna del hilo para decidir frente a un conflicto, por lo que la decisión del conflicto es tomada por solo por el componente responsable, el cual es la política. La responsabilidad de ejecución de este IT es asignada a distintos segmentos de ejecución; un segmento antes del conflicto (fork) y dos (o más) segmentos posteriores.
3) En el caso de que dos (o más) invariantes tengan en sus estructuras una plaza de union (join). La responsabilidad de ejecución de los invariantes es segmentada. Hasta el punto de unión (join) en dos o más segmentos, uno por IT, y después de la unión (join) solo un segmento de ejecución extra. Esto tiene como ventaja que mejora el paralelismo de la ejecución, dado que permite la ejecución de los segmentos posteriores y anteriores simultáneamente.

4.3 Algoritmo para la determinación de hilos máximos por segmento.

Para el cálculo de la cantidad de hilos necesarios por segmento:

1) Obtener los segmentos de los IT de la RdP con el algoritmo de la sección 4.1.
2) Determinar el conjunto de plazas de cada segmento denominado PS_i, con el árbol de alcanzabilidad de la Rdp, se debe obtener MS_i, el cual es el conjunto de todos los marcados posibles del iésimo segmento.
3) Del conjunto de marcados MS_i, se debe seleccionar el marcado máximo. Esta será la cantidad máxima de hilos necesarios de ese segmento.
4) Para obtener el número máximo de hilos necesarios el sistema, es necesario sumar los hilos máximos necesarios de todos los segmentos de ejecución.

5 Caso de aplicación y resultados

A continuación se utilizará como ejemplo la red S^3PR del autor Huang [18], para detallar los pasos de aplicación de cada algoritmo. Esta red ha sido modificada ya que en su versión original posee deadlock, es decir existe un estado o marcado a partir del cual la red no puede continuar evolucionando. Para evitar esto, se agregó la plaza de restricción P14 y los arcos correspondientes que se observa en la Fig. 2.
A continuación se realiza la aplicación de los algoritmos propuestos a la red. Se comienza con el algoritmo de la sección 4.1:
Las plazas idle de esta red S^3PR son: idle={P1,P8} y los recursos={P6,P7;P12,P13}

1) Los IT de esta red son:
   IT1={T1,T2,T4,T6}, IT2={T1,T3,T5,T6}, IT3={T7,T8,T9,T10}
2) Obtenemos el conjunto PI de cada IT:

   PI_1={P1,P2,P3,P5,P6,P13,P14}, PI_2={P1,P2,P4,P5,P7,P13,P14},

PI_3={P8,P9,P10,P11,P12,P13,P14}

3) a- Obtenemos el conjunto PA de cada IT:

PA_1={P2,P3,P5}, PA_2={P2,P4,P5}, PA_3={P9,P10,P11}

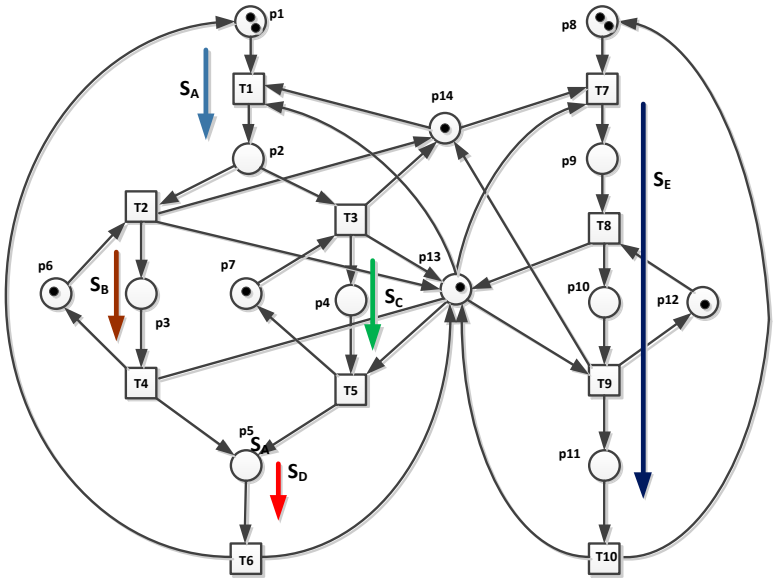


Fig. 2 RdP de ejemplo para aplicación de algoritmos (autor: Huang)

4) Obtenemos el conjunto de estados MA del conjunto de plazas PA donde PA = {P2,P3,P4,P5,P9,P10,P11}, y MA se observa en la Tabla 1.

Tabla 1. La tabla MA debe enumerar todos los marcados posibles, debido a la extensión de la misma a modo de ejemplo solamente se muestran los 4 primeros estados con su suma.

P2	P3	P4	P5	P9	P10	P11	SUMA
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1
0	0	0	0	1	0	0	1
0	1	0	0	0	0	0	1
…	…	…	…	…	…	…	…

De todos los marcados posibles, se busca el marcado máximo, este valor determina la máxima cantidad de hilos activos simultáneos, para el caso de la Fig. 2 son 3 hilos.

A continuación se aplica el algoritmo descripto en la sección 4.2 para determinar los segmentos de ejecución y responsabilidades de cada hilo; éstos se encuentran etiquetados en la Fig. 2 como S_A,S_B,S_C, S_D y S_E.

1) Se observa en la RdP de la Fig. 2 que el IT3 del sistema cumple la condición caso 1 del algoritmo, por lo cual se define el segmento de ejecución S_E y el subconjunto de plazas de acción asociadas al segmento $PS_E = \{P9, P10, P11\}$.
2) Se observan en la RdP que los IT1 e IT2 del sistema cumplen la condición caso 2 del algoritmo, por lo cual se determinan los siguientes segmentos de ejecución y plazas de acción asociadas:
   -Segmento previo al conflicto (fork) como segmento S_A y $PS_A = \{P2\}$.
   -Segmento Izquierdo como segmento S_B y $PS_B = \{P3\}$.
   -Segmento Derecho como segmento S_C y $PS_C = \{P4\}$.
3) Se observan en la RdP que los dos IT mencionados en el punto 2, cumplen además el caso 3 del algoritmo por lo cual se define un último segmento de ejecución luego de la unión (join) como segmento S_D y $PS_D = \{P5\}$.

Una vez determinados los segmentos de ejecución, se determinan los hilos máximos necesarios por segmento de ejecución. Para ello se aplica el algoritmo de la sección 4.3, el cual implica determinar los marcados máximos de los conjuntos de marcados MS_i para las plazas de acción de cada segmento de ejecución PS_i.

Para el caso de la RdP planteado en la Fig. 2 es: $Max(MS_A)$=1, $Max(MS_B)$=1, $Max(MS_c)$=1, $Max(MS_D)$ =1 y $Max(MS_E)$ =1.

Una vez determinados los hilos máximos por segmento, la suma de todos estos determina la máxima cantidad de hilos necesarios del sistema. En el caso de la red de la Fig. 2 es igual a cinco hilos.

Estos resultados han sido validados aplicando ésta metodología en las redes presentadas en [16, 19, 20] entre otras.

6 Conclusiones y trabajos futuros

En este trabajo se diseñaron un conjunto de algoritmos los cuales a partir del modelo de un sistema realizado con una RdP S^3PR no autónoma, determinan la cantidad de hilos necesarios en el software, la responsabilidad de los mismos y la máxima cantidad de hilos activos simultáneos. El diseño de la arquitectura propuesta, conserva las propiedades formales del modelo del sistema y mantiene la capacidad de expresión de una máquina de Turing. Asimismo los hilos y responsabilidades determinadas por este conjunto de algoritmos preservan estas propiedades.

En este trabajo se han validado los algoritmos propuestos con 16 redes distintas del tipo S^3PR, arrojando resultados similares a los obtenidos en el caso presentado. De esta validación es posible aseverar que estos algoritmos son eficientes debido a que la complejidad computacional es equivalente a la resolución de un sistema de ecuaciones lineales de dimensión igual a la matriz de incidencia, esta es necesaria para determinar los IT de la RdP mientras que la complejidad del árbol de alcanzabilidad de una RdP S^3PR corresponde a una combinación de procesos secuenciales simples. Además son precisos ya que definen sin ambigüedad un proceso para determinar el objetivo de cada uno de ellos; son determinísticos puesto que responden del mismo modo frente a las mismas condiciones y son finitos debido a que se garantiza su finalización. Estos algoritmos son importantes en las primeras fases del diseño de un sistema, dado que el número de hilos y sus responsabilidades

permiten establecer a priori parámetros del hardware, el grado de paralelismo y tiempos de respuesta para la implementación en sistemas embebidos. Esta metodología propuesta automatiza parte del proceso de diseño y generación del código mitigando así los tiempos de desarrollo y los errores de codificación.

Como trabajo futuro se mencionan las principales líneas de investigación para extender la presente metodología: incluir en los algoritmos las métricas de paralelismo y consumo de memoria entre diferentes modelos de un mismo sistema. Asimismo se está trabajando en un conjunto de algoritmos para determinar en forma automática el control sobre RdP que presentan un estado de deadlock y automatizar la determinación de políticas para el manejo de los conflictos en las RdP.

Referencias

1. Schwab, K., The fourth industrial revolution. 2017: Currency.
2. Halbwachs, N., Synchronous programming of reactive systems. 2013: Springer Science.
3. Munir, A., A. Gordon-Ross, and S. Ranka, Modeling and optimization of parallel and distributed embedded systems2015: John Wiley & Sons.
4. Zeigler, B.P., A. Muzy, and E. Kofman, Theory of modeling and simulation: discrete event & iterative system computational foundations2018: Academic press.
5. Diaz, M., Petri nets:fundamental models, verification and applications2013:John Wiley & Sons.
6. Phd. Micolini O., L.O. Eng. Ventre, and E.N. Eng. Daniele. Modular Petri Net Processor for Embedded Systems. (CACIC 2017) Revised Selected Papers. 2018. Springer CCIS.
7. Phd. Micolini, O., L.O. Eng. Ventre, and M. Eng. Ludemann. Methodology for design and development of Embedded and Reactive Systems Based on Petri Nets. in 2018 IEEE Biennial Congress of Argentina (ARGENCON). 2018. IEEE.
8. Eng. Ventre, L.O. and O. Phd. Micolini, Extended Petri Net Processor and Threads Quantity Determination Algorithm for Embedded System, in Communications in Computer and Information Science CCIS, E.J. Pesado P., Editor 2021, Springer, Cham: CACIC 2020.
9. Micolini, O., ARQUITECTURA ASIMÉTRICA MULTI CORE CON PROCESADOR DE PETRI, in Informatica2015, UNLP: UNLP La Plata, Argentina.
10. PhD. Micolini, O., et al. Ecuación de estado generalizada para redes de Petri no autónomas y con distintos tipos de arcos. in XXII (CACIC 2016).
11. Selic, B. and S. Gérard, Modeling and Analysis of Real-Time and Embedded Systems with UML and MARTE: Developing Cyber-Physical Systems2013: Elsevier.
12. Zhou, M. and N. Wu, System modeling and control with resource-oriented Petri nets. Vol. 35. 2018: Crc Press.
13. Siewert, S., Real time embedded components and systems2016: Cengage Learning.
14. David, R. and H. Alla, Discrete, continuous, and hybrid Petri nets 2010, Springer Science.
15. Micolini, O., Phd Thesis:Arquitectura asimétrica multicore con procesador de Petri, 2015: La Plata.
16. Liu, G. and K. Barkaoui, A survey of siphons in Petri nets. Information Sciences, 2016.
17. Peter A. Buhr, M.F., Monitor Classification. ACM Computing Surveys 1995. 27(1):
18. Huang, Y.S., Y. Pan, and P. Su, Transition-based deadlock detection and recovery policy for FMSs using graph technique. ACM Transactions on Embedded Computing Systems, 2013.
19. Timotei, A. and J.M. Colom. A New Approach to Prevent Deadlock in S3PR Nets with Unreplicable Resources. in ICORES. 2013.
20. Zhong, C.F. and Z.W. Li, Design of liveness-enforcing supervisors via transforming plant petri net models of FMS. Asian Journal of Control, 2010. 12(3): p. 240-252.

ISBN 978–987–633–574–4

Sistema domótico de control de iluminación y procesamiento de datos mediante mqtt centralizado en la nube

Carlos Binker[1], Hugo Tantignone[1], Guillermo Buranits[1], Eliseo Zurdo[1], Diego Romero[1], Maximiliano Frattini[1], Lautaro Lasorsa[1]

[1]Universidad Nacional de La Matanza, Florencio Varela 1903 (B1754JEC) -- San Justo, Buenos Aires, Argentina
{cbinker, htantignone, gburanits, djromero}@unlam.edu.ar;
{ezurdo, mfrattini, llasorsa}@alumno.unlam.edu.ar

Resumen. Este trabajo aborda la implementación de un sistema de control de dispositivos a nivel hogareño, empleando componentes IOT. Dicha implementación le brindará al usuario un tablero de control (dashboard) a través del cual se logrará la interacción para operar sobre sus dispositivos, ya sea emitiéndoles órdenes a los mismos, o bien recibiendo datos provenientes de dichos elementos. Este procedimiento de envío y recepción de mensajes hacia o desde los dispositivos IOT, se llevará a cabo a través de un protocolo de manejo de cola de mensajes diseñado para funciones de telemetría como lo es el MQTT, basado en tópicos que permite publicación y suscripción. Por otro lado el dispositivo IOT utilizado (placa ESP-01 Relay y el ESP01s) incorpora como valor agregado el hecho que se puede conectar su relé de salida a una llave de combinación, permitiendo así un control independiente de la iluminación, es decir a través de la aplicación web (dashboard) o bien desde la tecla de combinación tradicional y además la carga a controlar no tiene por qué ser un dispositivo inteligente, tal como exigen otros sistemas comerciales.

Palabras claves: IOT, MQTT, ESP01s, web socket, dashboard

1 Introducción

Se presentará un caso de estudio en donde se propone un sistema domótico de control de iluminación hogareño por medio de switches a través de una aplicación web (*dashboard.php*). Para ello se desarrolló una maqueta como prototipo que simula una instalación eléctrica básica hogareña (ver Figura 1). El módulo de control (la placa ESP-01 Relay en la cual va insertado el microcontrolador ESP01s) se inserta junto con la fuente de alimentación de 220 V AC a 5 V DC en una pequeña caja plástica (construida con una impresora 3D), que tiene la ventaja que puede ser montada dentro de un bastidor de 10 cm x 5 cm (ver Figura 2). El dashboard posee también un espacio en donde pueden mostrarse los datos provenientes desde los dispositivos IOT. En el caso del estudio planteado, estos datos vendrán como valores generados desde el propio ESP01s [1] (datos enviados por cada cliente, de acuerdo al estudio). Estos

datos podrían ser por ejemplo el envío de variables provenientes de sensores, tales como la tensión y la corriente alterna, y a partir de estos valores se podría determinar por ejemplo la potencia aparente, la potencia activa, el coseno fi, etc. El acceso al dashboard se da a través de una pantalla de autenticación del usuario *(login.php)*, el cual previamente debió registrarse (*register.php*). El dashboard además permite al cliente registrar y borrar los dispositivos pertenecientes al usuario logueado (*devices.php*). Estas interfaces web fueron construidas a partir de la instalación de un panel de desarrollo web (*Flatkit)* [2], el cual fue instalado en nuestro VPS [3] en la nube (ver Figura 3). La IP de nuestro VPS es una IP pública y el dominio asociado es *c2ing076.tk*, también se instaló en el VPS MySQL [4] para generar las siguientes tablas: *users* y *devices* (ver Figura 4). Los requerimientos del VPS son: 1 GB de RAM, 10 GB de SSD y un núcleo. El SO instalado es Ubuntu 18.04 LTS [5]. El broker mqtt [6] utilizado para el estudio fue EMQX [7], dado su excelente desempeño profesional y además posee una licencia de uso gratuito que permite el control de hasta cien mil dispositivos. El estudio se realizó para dos clientes con un único dispositivo ESP01s por cada cliente, el cual envía y recibe datos desde el broker mqtt bajo diferentes tópicos.

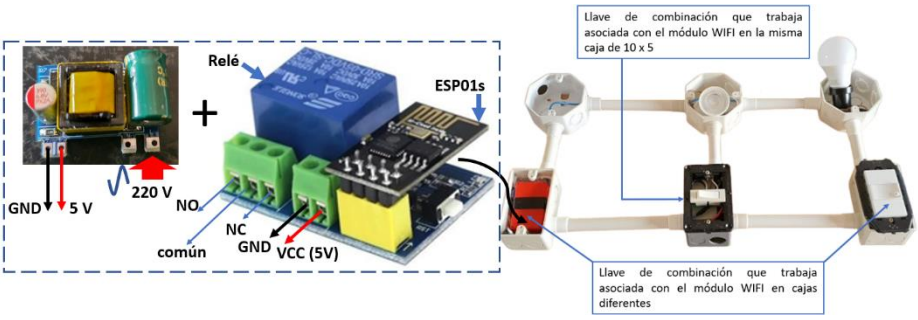


Figura 1. Placa ESP-01 Relay con ESP01s y fuente (izquierda) y maqueta prototipo (derecha)

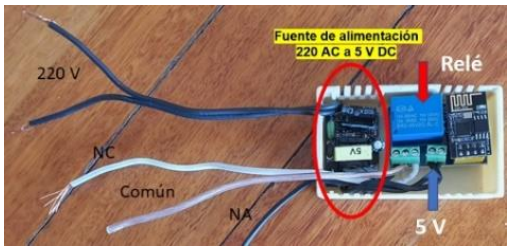


Figura 2. Caja contenedora del módulo Relay con el ESP01s más la fuente de alimentación

1.1 Hardware ESP01s

Se utilizó el microcontrolador ESP01 serie S que incorpora el chip ESP8266 de ESPRESSIF [8]. Para programar el dispositivo se utilizó Platformio [9] instalado en Vscode desde donde se confeccionó el proyecto (tanto la parte de PHP, JavaScript,

como la programación del 8266 misma). Se instaló SFTP [10] para la comunicación con el VPS. Esta placa de desarrollo con el ESP01s tiene conexión a WiFi, 2 puertos GPIO (I00 e I02). El WiFi es compatible con el protocolo 802.11 b/g/n y soporta autenticación WEP y WPA/WPA2 (Figura 5).

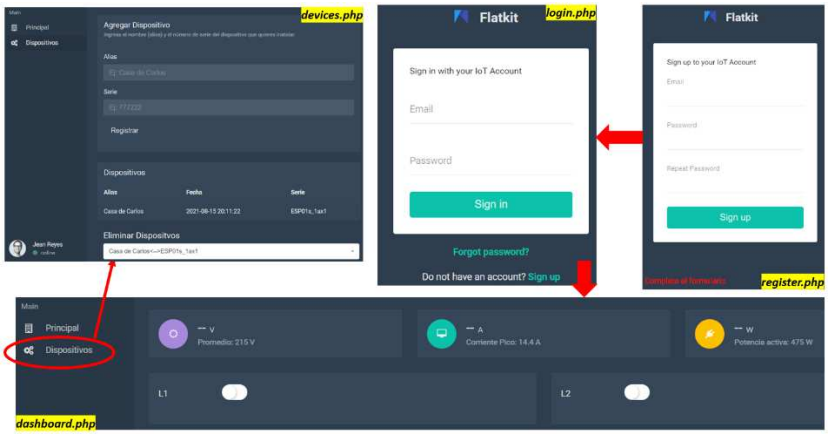


Figura 3. Vistas de los módulos register.php, login.php, dashboard.php y devices.php

Figura 4. Tablas users y devices de la base de datos admin_c2ing076 en MySQL

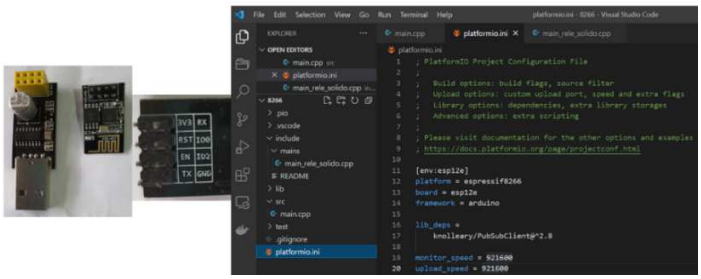


Figura 5. Programador ESP01s (izquierda), diagrama de pines (centro) y software Platformio

1.2 Placa ESP-01 Relay

El ESP01s fue montado sobre una placa que contiene un relé que permite controlar una carga para 220 V, en este caso una lámpara LED (podría ser cualquier dispositivo

hasta un máximo de 10 A). El relé está optoacoplado con la parte del microcontrolador, lo que permite una aislación galvánica entre los 220 V y la parte de control que funciona con sólo 3.3 V. El Relé se activa a través del pin IOO del ESP01s con un nivel bajo (LOW). El terminal común del relé está conectado normalmente al borne NC. Cuando se energiza la bobina se conecta al borne NA (Figura 6). Por otro lado recordemos que el relé está conectado en combinación con una llave inversora de tecla convencional, permitiendo así encender o apagar la luminaria tanto desde el dashboard como desde la llave física. Se hace énfasis en este punto, *ya que no es algo menor*, dado que la mayoría de estos módulos que pueden encontrarse de manera comercial son sólo un mero switch electrónico activado por wifi que sólo funciona como una llave de un punto y la carga a controlar es por lo general un dispositivo inteligente; en este caso de estudio podemos usar cualquier luminaria, que sólo cumpla con la máxima especificación de carga ya mencionada.

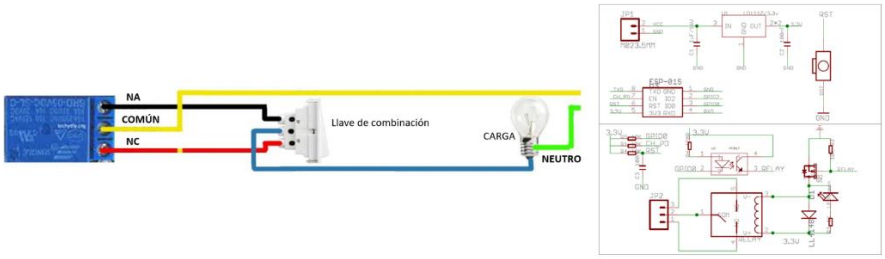


Figura 6. Diagrama circuital del módulo ESP-01 Relay y circuito de combinación con tecla

2 Descripción del broker MQTT EMQX

El broker mqtt, que residirá en el VPS, es el responsable de la administración de los mensajes provenientes tanto de los dispositivos IOT, como de las aplicaciones web. Estos mensajes se organizan bajo estructuras temáticas denominadas *tópicos* [11], [12], [13]. Cuando un dispositivo (por lo general a través de un elemento sensor) o una aplicación web envían información bajo determinado tópico, otros dispositivos u otras aplicaciones web podrán suscribirse a dichos tópicos. De esta manera, sólo aquellos dispositivos o aplicaciones web que se suscriban a determinados tópicos recibirán información de los mismos a través del broker mqtt. Por ende el mqtt es un protocolo denominado de *Publicación/Suscripción* y es el encargado de filtrar y administrar la distribución de mensajes entre dispositivos y aplicaciones (Figura 7). El broker elegido para esta investigación es el EMQX. Como se ha expresado en la introducción, la elección se debe al alto desempeño profesional que tiene frente a otros brokers (como por ejemplo Mosquitto, CloudMQTT, etc.) y por tener una licencia de uso gratuito capaz de soportar hasta 100K dispositivos (Figura 8). El EMQX se comunica con los dispositivos IOT por medio del puerto TCP 1883, en cambio con las aplicaciones web lo hace por medio de *web sockets*, empleando el puerto 8093 para *ws*, mientras que para *wss* (Websockets [14] sobre SSL/TLS) se emplea el puerto 8094. WebSocket es una tecnología que proporciona un canal de

ISBN 978–987–633–574–4

comunicación bidireccional y full-duplex sobre un único socket TCP. Está diseñada para ser implementada en navegadores y servidores web, pero puede utilizarse por cualquier aplicación cliente/servidor. En nuestro caso de estudio se emplearán los puertos 1883 (TCP) para comunicar el ESP01s con EMQX, mientras que la comunicación entre el *dashboard.php* y el EMQX se realizará a través del puerto 8094 por websocket seguro. También se ha instalado un certificado de seguridad SSL a través de Lets Encrypt [15] en nuestro VPS para el dominio utilizado en el estudio *c2ing076.tk*. Lógicamente que para que esto funcione se han habilitado las respectivas reglas en el firewall de nuestro VPS en cuanto a la habilitación de los puertos correspondientes. Hay otros puertos que también deberán configurarse para el correcto funcionamiento, como el 8883 para SSL y el 8090 para management. Por otro lado el puerto 18083 se emplea para conectarse al dashboard del EMQX (no confundir con nuestro panel *dashboard.php*). El dashboard del EMQX es un cliente web que permite administrar gráficamente el broker de una forma mucho más amigable y es provisto por EMQX y se accede por *http://c2ing076.tk:18083*. Los listeners pueden observarse en la Figura 9.

Figura 7. Ejemplo de tópicos y subtópicos. Operador mono nivel (+) y multinivel (#)

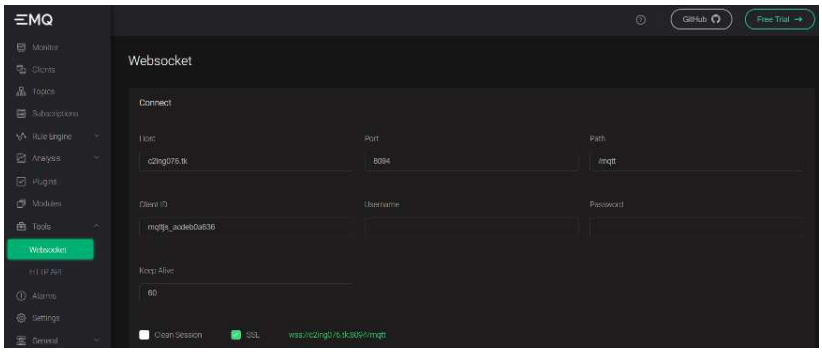


Figura 8. Dashboard (panel de control) correspondiente al EMQX. http://c2ing076.tk:18083

Tomando el ejemplo de tópicos de la Figura 7, el mismo puede interpretarse de la siguiente manera: existe un tópico principal denominado UNIVERSIDAD, la

universidad posee dos edificios (A y B), con tres pisos por edificio. A su vez en cada piso se han instalado sensores de humedad y de temperatura y en el Piso_PB se han agregado también módulos de medición de tensión eléctrica por su proximidad a los tableros de energía. Por ende, los edificios, como así también los pisos son subtópicos del tópico principal que como dijimos es UNIVERSIDAD. En el ejemplo, se está emitiendo un mensaje bajo el tópico UNIVERSIDAD/Edificio_B/Piso_2/temp enviando el valor de 25 °C. Por lo tanto el dispositivo que se suscriba a ese tópico recibirá todos los valores de temperatura sólo del Piso_2 del Edificio_B. También puede emplearse el símbolo monovalente **"+"** o el multivalente **"#"**. Como ejemplo del operador monovalente podemos indicar que si un dispositivo se suscribiera a UNIVERSIDAD/**+**/Piso_2/temp recibiría todos los mensajes de temperatura enviados de ambos edificios, pero sólo del Piso_2. En cambio como ejemplo de multivalente podríamos citar: UNIVERSIDAD/Edificio_A/**#**, en este caso se recibirían todas las variables sensadas de todos los pisos, pero solamente del Edificio_A.

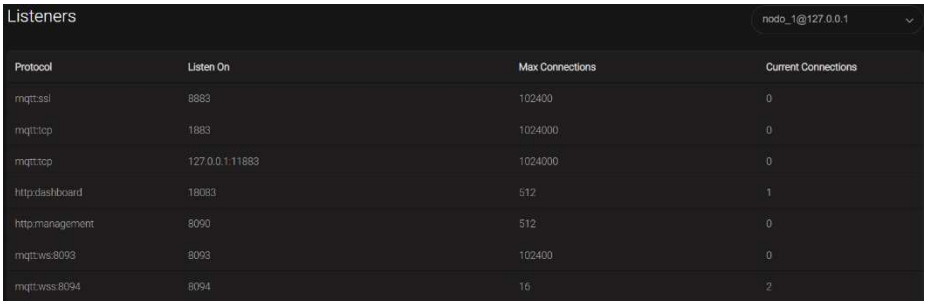


Figura 9. Configuración de los puertos de EMQX (listeners)

Otra característica muy importante de EMQX (en realidad de mqtt), es lo que se conoce como QOS (Quality of Service) [16], el cual puede tomar 3 valores posibles: 0, 1 ó 2. Un QOS igual a 0 significa que se confía en TCP para el envío de los mensajes, esto no significa que el mensaje va a llegar necesariamente al usuario, sí llegará al broker, pero éste lo enviará según la disponibilidad de la red, es decir si hubiera un fallo o el servidor se reiniciara por ejemplo, el mensaje no llegaría al cliente suscripto a un determinado tópico. De todas maneras es poco probable que suceda si la red es de buena calidad. En cambio si QOS es igual a 1, el broker intentará "*al menos 1 vez"* entregar el mensaje al usuario suscripto. Esto da la garantía que el mensaje será entregado al suscriptor del tópico, pero en el afán de cumplir con la meta de entrega, podría haber mensajes duplicados, es decir los suscriptores podrían recibir los mensajes en múltiples ocasiones. Esto también va a implicar mayor carga de tráfico y de procesamiento para el broker. Finalmente una QOS igual a 2 implica que sólo le llegará al suscriptor "*una sola vez"* el mensaje. Este último caso es importante en la situación en que el suscriptor no pueda recibir mensajes en múltiples ocasiones, pero implica un mayor nivel de procesamiento ya que el broker debe asegurarse que efectivamente el mensaje al suscriptor sólo se entregue en una oportunidad. Para finalizar debemos decir que el EMQX posee numerosos *plugins*,

que le añaden enorme funcionalidad. Por ejemplo puede autenticar clientes a través de una tabla de usuarios (mqtt_user) o bien ejecutar ACL (mqtt_ACL). Estas dos tablas vienen con una configuración por default y son provistas por EMQX.

3 Topología del sistema de control y análisis de tópicos

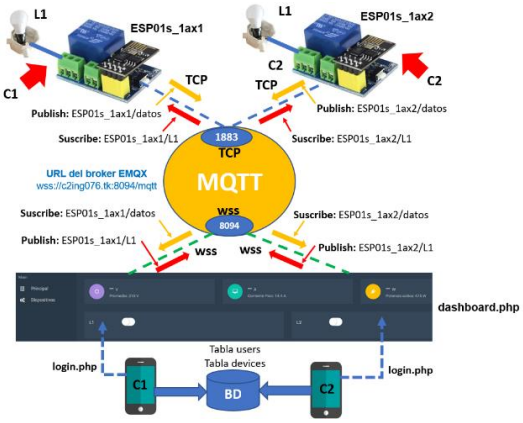


Figura 10. Topología de conexión para dos clientes (C1 y C2) con sus respectivos dispositivos

La topología de la figura 10 muestra un análisis para dos clientes con sus respectivos dispositivos ESP01s. Según puede observarse en el diagrama de la figura precedente, se utiliza la serie del dispositivo (almacenada en la tabla de la BD MySQL como *devices*) como parte principal del tópico. Cada dispositivo publica datos bajo el tópico **N°_de serie**/datos. Para el cliente 1 (C1) el número de serie es ESP01s_1ax1, mientras que para el cliente 2 (C2), la serie es ESP01s_1ax2. Esto permite fácilmente desde el código JavaScript del dashboard.php extraer mediante un método de Split (cuyo separador es el **"/"**) el tópico principal haciendo que los datos se transmitan únicamente al cliente logueado y a su dispositivo asociado.

3.1 Análisis de los procesos en el ESP01s y en la aplicación web (dashboard)

En el ESP01s se utilizan las siguientes funciones principales provistas a través de la librería *PubSubClient.h* [17]:

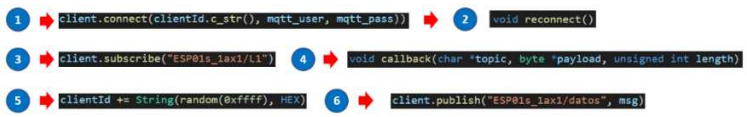


Figura 11. Funciones y métodos empleados en el ESP01s

Observando la figura 11, en **1** tenemos el método "*connect",* al cual debe pasársele el ClientId (ver **5** cómo se genera), el usuario y el password para conectarse al mqtt.

Esta función se incorpora dentro de **2** (función *reconnect*), y es en esta parte del código en donde se produce la suscripción del dispositivo a los diferentes tópicos. En este caso se hace con la función descripta en **3** bajo el tópico indicado "ESP01s_1ax1/L1" para C1 (cliente 1) y lo propio para C2 con el tópico "ESP01s_1ax2/L1". Finalmente en **4** se describe el método "*callback*", el cual recibirá los mensajes bajo los tópicos suscriptos y en función de esos mensajes procederá a excitar o no al relé que controla cada luminaria (ver figuras 12 y 13 en resultados obtenidos). La publicación por parte del ESP01s se realiza con un método dentro del loop (ver **6**). Un análisis similar se emplea desde el lado de JavaScript en los procesos y en la conexión a mqtt por parte del *dashboard.php*. Todas las funciones y métodos se basan en la librería *mqtt.js* [18].

4 Resultados obtenidos

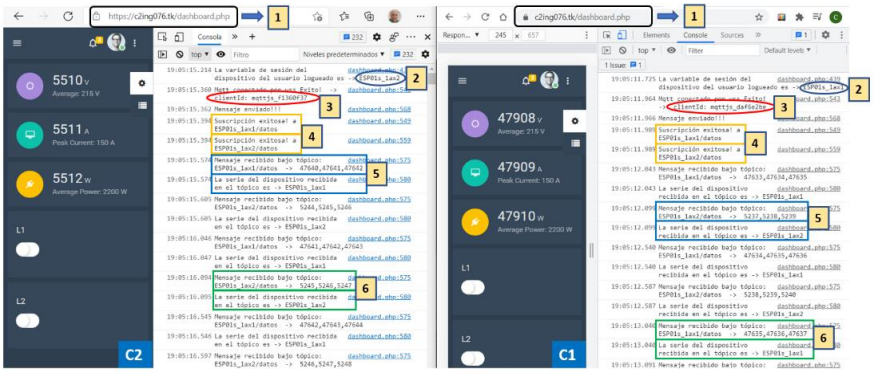


Figura 12. Captura de datos bajo tópicos ESP01s_1ax1 (C1) y ESP01s_1ax2 (C2)

En **1** observamos como ambos clientes (C1 y C2), efectivamente se loguean al mismo *dashboard.php*. En **2** observamos como es capturada la serie de cada dispositivo cuando el cliente se loguea, para ello se apela a las variables de sesión de PHP, dichas variables se obtienen desde la tabla *devices* de la base de datos tras corroborarse el logueo correcto del cliente en el archivo *login.php*. En **3** observamos como se genera un Id de cliente aleatorio en cada cliente logueado. Esto debe ser así porque si al cliente se le cae la conexión y se lograra de nuevo con el mismo ClientId, sería rechazado por el broker mqtt. Idéntica situación se da para el dispositivo ESP01s, en donde también apela a la generación de un ClientId aleatorio cada vez que se conecta al broker mqtt. En **4** se observa la suscripción que realiza el dashboard en todos los dispositivos de los clientes bajo el subtópico *datos*, en este caso para el estudio en cuestión se trata sólo de dos clientes, pero esto puede por supuesto extenderse a n clientes. En **5** se observa para cada uno de los clientes como al recepcionar datos en un tópico cuyo número de serie no se corresponde con el dispositivo del cliente, los mismos son ignorados. Finalmente en **6**, se observa como los valores recibidos bajo el tópico correcto impactan en el tablero html, indicando así los tres valores indicados.

ISBN 978–987–633–574–4

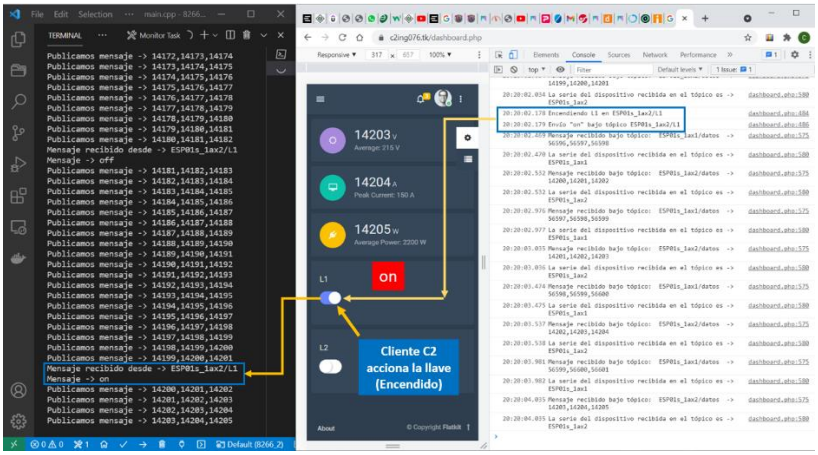


Figura 13. Encendiendo la luminaria bajo tópico ESP01s_1ax2/L1 (C2) bajo el mensaje ‘on’

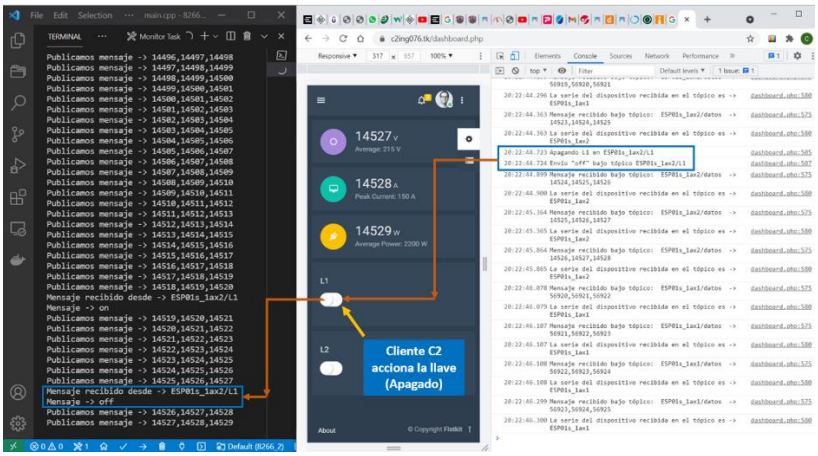


Figura 14. Apagando la luminaria bajo tópico ESP01s_1ax2/L1 (C2) ajo el mensaje ‘off’

5 Conclusiones y trabajo futuro

1. Este estudio contempla dos escenarios posibles en cuanto al tema de la instalación eléctrica se refiere. Por un lado un escenario que es el desarrollado en este estudio, en donde cada caja con su relé accionado por el microcontrolador puede ser ubicada en un bastidor de 10 cm x 5 cm, no alterando en absoluto la instalación eléctrica existente. Esto implica que cada usuario registre varios dispositivos.
2. El segundo escenario prevé la posibilidad de que con un único dispositivo por usuario (por ejemplo un ESP-32) que posee múltiples pines GPIO se exciten a varios relés, en donde cada luminaria a controlar se enmarcaría bajo un tópico. La complejidad acá radica en construir la placa, en alojar la misma en alguna caja (que podrá ir embutida o no, etc.); en síntesis esta es una buena alternativa pero quedaría

circunscripta a nuevas instalaciones eléctricas, o bien al deseo del usuario de modificar la instalación existente, cosa que no siempre es técnicamente factible.

Entre los trabajos futuros se puede mencionar:

- Incorporación de un motor de reglas implementado en Node.js, que esté a la espera de eventos preconfigurados, por ejemplo "*que se active un ventilador cuando la temperatura del recinto a controlar supere determinado umbral*".
- Configuración dinámica del dashboard por medio de los usuarios del sistema, de manera tal que se vayan incorporando los nuevos dispositivos de cada usuario al tablero de control.
- Suscripción dinámica de tópicos de todos los dispositivos de los usuarios (tarea a ejecutarse en el dashboard.php), esto se hace imprescindible conforme aumente la cantidad de usuarios del sistema.

6 Referencias

1. NodeMCU Connect Things EASY, http://www.nodemcu.com/index_en.html
2. https://themeforest.net/item/flatkit-app-ui-kit/13231484
3. https://www.ambit-bst.com/blog/definici%C3%B3n-de-iaas-paas-y-saas-en-qu%C3%A9-se-diferencian
4. https://es.wikipedia.org/wiki/MySQL. Fuente Wikipedia.
5. https://ubuntu.com/
6. Protocolo MQTT (Message Queue Telemetry Transport), http://mqtt.org.
7. EMQX. https://www.emqx.io/
8. Espressif System ESP8266 Series, https://www.espressif.com/. Ceja, J., Renteira, R., Ruelas, R., & Ochoa, G. (2017). Módulo ESP8266 y sus aplicaciones en el internet de las cosas. Revista de Ingeniería Eléctrica, 24-36.
9. https://platformio.org/install/integration
10. https://github.com/liximomo/vscode-sftp
11. Hands-On Internet of Things with MQTT: Build connected IoT devices with Arduino and MQ.
12. Implementación de middlewarepublicador/subscriptor para aplicaciones web de monitoreo. In XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires).
13. Naik, N. (2017, October). Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP. In 2017 IEEE international systems engineering symposium (ISSE) (pp. 1-7). IEEE.
14. Arquitectura de software con websocket para aplicaciones web multiplataforma. VI Workshop Innovación en Sistemas de Software (WISS). CACIC 2014.
15. https://letsencrypt.org/es/
16. https://www.hivemq.com/blog/mqtt-essentials-part-6-mqtt-quality-of-service-levels
17. https://github.com/knolleary/pubsubclient
18. https://github.com/mqttjs/MQTT.js

ISBN 978–987–633–574–4

Open R.A.N. y Fallas en una red de Telecomunicaciones

Carlos Peliza[1], Fernando Dufour[2], Ariel Serra[3], Gustavo Micieli[4], Darío Machaca[5]

Universidad Nacional de La Matanza
Florencio Varela 1903 (B1754JEC) - San Justo, Buenos Aires, Argentina
cpeliza@unlam.edu.ar, fdufour@unlam.edu.ar, aserra@unlam.edu.ar

Abstract. Este trabajo, pretende recordar las dificultades que atraviesa una implementación de telecomunicaciones con tecnología novedosa, según la visión de los propios especialistas en el tema.

Con posterioridad, introducir los conceptos fundamentales de una arquitectura abierta que conforma parte de la generación 5G de redes móviles para el acceso a la red (llamada Open R.A.N.), revisando para ello la bibliografía disponible. A continuación, luego de enunciar las posibilidades de desarrollo e implantación de la arquitectura revelaremos la elegida para una prueba de concepto en Argentina y como corolario de los puntos anteriores, exponer las problemáticas encontradas en la PoC realizada en Puerto Madryn.

El análisis de problemáticas en la implantación de nuevas tecnologías dentro de las redes móviles es inherente al desarrollo de estas, sin embargo, parece reñido cada vez más con las políticas económicas de las compañías, por eso cabe interrogarse sobre cuál será el límite entre enfrentar al mercado con un producto en desarrollo o con uno verdaderamente asentado en la red.

Keywords: Open RAN, PoC, Redes Móviles, 5G.

1 Introducción: El pasaje de problema propietario a problema de plataformas abiertas.

En el universo de las comunicaciones, el pasaje de un mundo de redes de conmutación telefónica a redes de paquetes que transportan voz, como una parte más de la información no puede definirse como sencillo.

Ha sido un trabajo de aprendizaje donde se involucran, sistemas de complejidad creciente y expertos con la ductilidad para adaptarse a un cambiante esquema de trabajo. En resumen, el pasaje de hardware de comunicaciones propietario hacia la industria del software de telecomunicaciones encontró y encontrará escollos de diferente dificultad.

Atrás en el tiempo han quedado los problemas para configurar el servicio de llamada en espera de una red NGN que requería de análisis y actuación propias de un mundo telefónico y, en la actualidad, podemos hallarnos frente a la problemática ge-

nerada por el exceso de retransmisiones de segmentos TCP en una red de telecomunicaciones.

Con esta breve introducción hemos pretendido clarificar la situación actual del mundo de las Telecomunicaciones y de la Industria del software para ese mundo. Ambos ecosistemas se hallan en plena fusión y el devenir de las complicaciones, ya no arroja la asignación de soluciones a uno u otro lado (informático o telefónico), sino que espera la comunión de estos para avanzar en más y mejores servicios al usuario final (Roca, Biga, & Del Giorgio, 2012).

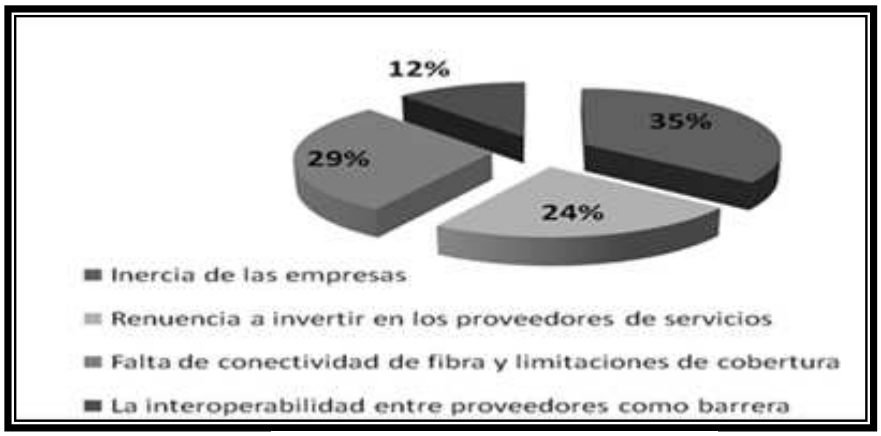


Ilustración 1 La barrera entre proveedores

2 Objetivos

El presente trabajo tiene como objetivo hacer una descripción general de la arquitectura Open R.A.N. y distinguir en ella los desarrollos que fueron elegidos en Argentina, por ello el estilo de este trabajo de investigación es comparativo y se basa en el análisis de fuentes bibliográficas y documentación existente, junto a la realización de pruebas de concepto y funcionamiento del servicio Open R.A.N.

Como objetivo primordial, se pretende exhibir la dificultad de coordinación ante problemas con la red en funcionamiento, para brindar una solución en la velocidad esperada.

3 Algunos antecedentes

En Argentina, el pasaje de 3G a 4G ocurrió casi en simultaneo con el cambio de tecnología en las redes fijas desde conmutación tradicional con hardware propietario a plataformas abiertas basadas en IP, por esa razón resulta pertinente recordar la opi-

nión de los especialistas con relación a los cambios, dado que ya en 2014 se expresaban por la complicación de trabajar con múltiples proveedores.

Mas cercano en el tiempo, una nueva consulta, pero esta vez con relación a la virtualización de funciones de red, una de las tecnologías en las que se basa la 5° generación de redes móviles (Dufour Fernando Javier;Micieli Gustavo;Serra Ariel, 2015) o 5G mostraba los siguientes resultados:

Ilustración 2 Integración entre diferentes proveedores

En ajustada síntesis podemos afirmar que las experiencias de los especialistas en telecomunicaciones consultados se han visto atravesadas tanto por el avance hacia el mundo de la informática partiendo del universo de la conmutación como por incorporar las formas de trabajo propias de ese mundo informático.

La forma de trabajo del mundo informático, enunciada bajo el paraguas de sistema abierto con ejemplo distintivo en el software Open Source, han llegado para buscar imponerse en el mundo de las comunicaciones con sistemas propietarios. Así las cosas, el especialista en sistemas de conmutación NEC, debió mutar a especializarse en sistemas de redes convergentes de paquetes y en la actualidad, especializarse en protocolos y lenguajes de programación.

4 El universo del Open R.A.N.

Este universo surge en 2018 con la conformación de la O-RAN Alliance que según podemos recolectar de su página web:

La O-RAN ALLIANCE fue fundada en febrero de 2018 por AT&T, China Mobile, Deutsche Telekom, NTT DOCOMO y Orange. Se estableció como entidad alemana en agosto de 2018.Desde entonces, O-RAN ALLIANCE se ha convertido en una

comunidad mundial de operadores de redes móviles, proveedores e instituciones académicas y de investigación que operan en la industria de la red de acceso por radio (RAN).

La misión de O-RAN ALLIANCE es remodelar la industria de RAN hacia redes móviles más inteligentes, abiertas, virtualizadas y totalmente interoperables. Los nuevos estándares O-RAN permitirán un ecosistema de proveedores de RAN más competitivo y vibrante con una innovación más rápida para mejorar la experiencia del usuario. Las redes móviles basadas en O-RAN mejorarán al mismo tiempo la eficiencia de las implementaciones de RAN, así como las operaciones de los operadores móviles. como el compromiso de ciertos operadores (O-RAN ALLIANCE, 2018).

Entre las compañías más destacadas que conforman la Oran Alliance se pueden encontrar operadores dentro de los servicios de telecomunicaciones como: AT&T, China Mobile, Deutsche Telekom, NTT DOCOMO, Orange, Bharti Airtel, DISH Network, KDDI, Rakuten Mobile, Reliance Jio, Singtel, TIM, Telefónica, Verizon, Vodafone junto a proveedores de servicios para las telcos como ser: Accelleran, Accenture, Altiostar, AMD, Amdocs, Anritsu, Broadcom, Ciena, Cisco, Commscope, Dell, Facebook, Kyocera, Mavenir e Intel entre otros.

En suma, podemos explicitar que, pese a las dificultades manifestadas, la industria de las Telecomunicaciones ha volcado su energía hacia formas de integración entre diferentes proveedores y puesto especial foco en operar y desarrollar arquitecturas abiertas.

Open RAN significa Open Radio Access Network (RAN), consiste en transformar la red de acceso móvil de radio en una arquitectura abierta que conecta un dispositivo con el núcleo de la red. En los sistemas móviles tradicionales, estaba conformado un sistema donde cada proveedor de tecnología tenía la incumbencia, de esta forma la solución de acceso de una zona geográfica se asignaba a un proveedor / fabricante de hardware/software de acceso por radio.

La Ilustración 3 nos muestra esquemáticamente la transformación desde el universo tradicional hacia la propuesta de O-RAN Alliance. Allí podemos distinguir la tradicional torre con antenas en la parte superior y debajo de ella la radiobase o BBU, cumpliendo la función de conectar y asignar recursos de radio a los usuarios según los mandatos del core de red frente al sistema Open RAN con componentes distribuidos o no, de acuerdo con el diseño tecnológico elegido.

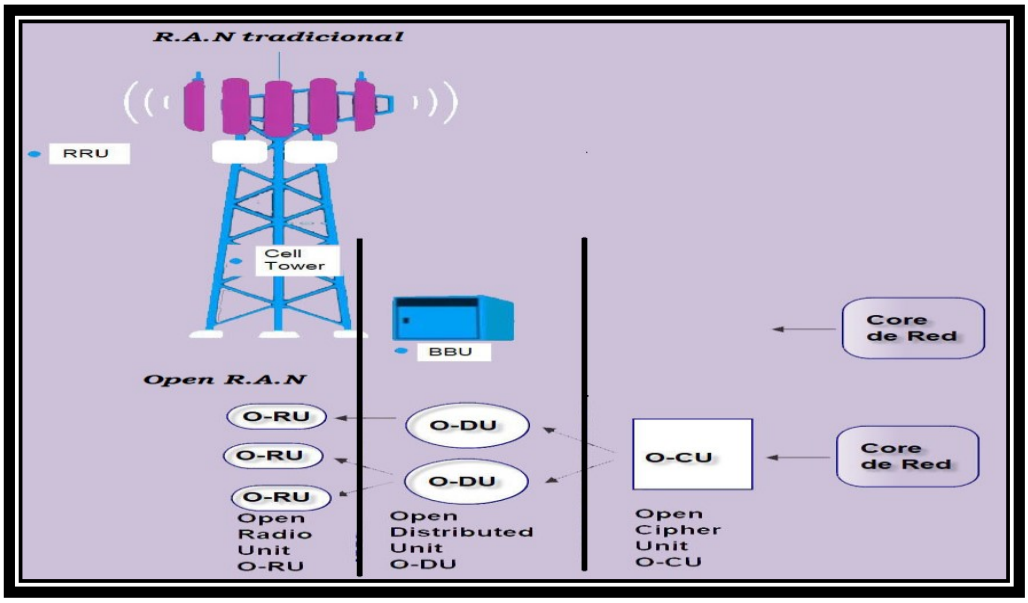


Ilustración 3 Esquema de RAN tradicional vs Open RAN

Es necesario destacar que, tanto las unidades remotas O-RU como las unidades distribuidas O-DU y las unidades de cifrado O-CU, pueden ubicarse en diferentes lugares físicos, ya sea a pie de la torre o en lugares remotos con mayor seguridad.

4.1 Open RAN en Argentina

Como parte del grupo de investigación en redes de 5 generación hemos dedicado energía al análisis del funcionamiento de la arquitectura Open RAN en Argentina, en particular a la prueba de concepto desarrollada por una de las compañías miembros de la O-RAN ALLIANCE en una ciudad de la provincia de Chubut.

Las características demográficas de la ciudad en cuestión la hacen única para una prueba de concepto debido a lo que explica Sergio Kaminker (Cannizzaro, 2014), hallarnos frente a una ciudad portuaria y fabril con una importante cantidad poblacional.

Para la prueba de concepto se han elegido dos modalidades de conexión entre las zonas geográficas a brindar cobertura y el core virtualizado de la red móvil. Una de las modalidades tiene la unidad de distribución (ODU) física y la otra la tiene virtualizada (vODU), en ambos casos la unidad de cifrado es virtual (vCU)

La ilustración 4 permite conocer que se ha elegido usar un escenario donde la unidad de cifrado (vCU) es única y brinda servicio a dos ODU (una física y una virtual) que se ubican en lugares diferentes dentro del esquema, cerca de las ORU y lejos de la ORU, lo que conforma un escenario de pruebas complejo y completo, dado que se cubren muchas variantes de conexión (de ahí su completitud) al mismo tiempo que resulta compleja la detección de fallas por la cantidad de equipamiento involucrado.

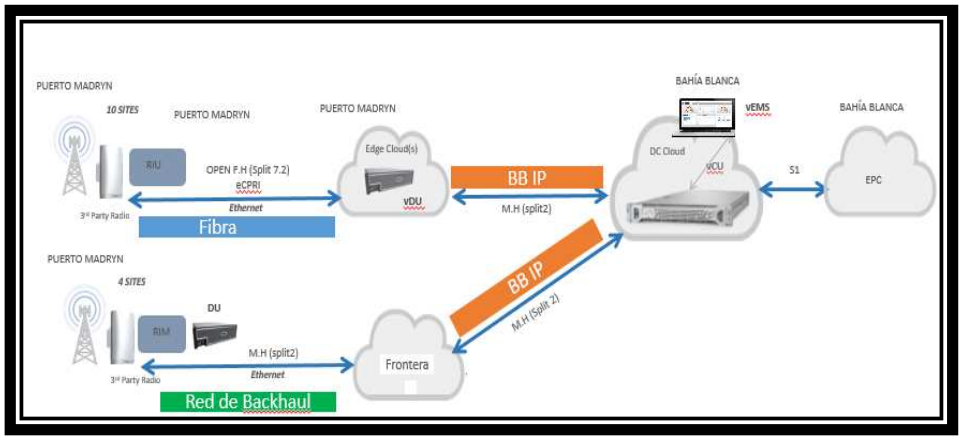


Ilustración 4 Esquema de conexión elegido

La inserción de un ecosistema Open RAN en una red tradicional funcionando, implica la participación de diferentes actores. Podemos describir entre estos actores, los equipos de facturación, los sistemas de gestión y los de monitoreo de alarmas junto a los sectores de ingeniería de red y mantenimiento en sitio.

El sistema de monitoreo de calidad de servicio es en sí mismo un punto necesario de análisis a profundizar, dado que será quien nos brinde los primeros detalles de la complejidad del abordaje. Recordemos que se han seleccionado radiobases en servicio activo y para evitar la falta de cobertura en toda la zona, de las radiobases se han modificado sectores de atención.

Así, ante un corte de energía eléctrica, se obtiene un registro de evento en la red tradicional y el la Open RAN, permitiendo por comparación, el análisis de la respuesta de ambos sistemas, mientras que los reclamos por fallas en zonas de cobertura permitirán asociar el evento a uno de los dos sistemas.

5 Las complicaciones de un sistema abierto

Una vez concluidos los trabajos de puesta en marcha del equipamiento donde se debe destacar la necesidad de un trabajo conjunto de ingeniería de asignación de recursos y de registro en todos los sistemas, a la vez que la accesibilidad a la información de todos los participantes de la prueba se debe avanzar en analizar los indicadores de la calidad de servicio.

El primer punto de negociación ha sido que mirar, donde mirarlo y quien lo mira, porque al producirse degradación en los valores de calidad de la red se presentan cuellos de botella de difícil solución. Los indicadores de calidad surgen de la interacción de diferentes situaciones de la red, como ser interferencias en el acceso por radio, demoras en el procesamiento de la información, latencia de la red, criterios subjetivos de opinión del usuario con relación a un servicio, problemas de cobertura de la radiobase etc.

En resumen, debe acordarse cuales son los parámetros objetivos de medición para considerar adecuada la respuesta del ecosistema y cuáles son los puntos de interconexión entre las partes intervinientes o cuáles serán los instrumentos o sistemas para la verificación de los problemas hallados.

En particular, en la prueba descrita se descubrió una degradación del servicio en los casos de uso de streaming de video por TCP. Las pruebas de puesta a punto no habían cubierto/previsto esta situación particular porque como resulta natural, en una prueba de puesta a punto no puede considerarse lo mismo que cuando la solución se halla en funcionamiento.

En otras palabras, cuando se definen las pruebas de puesta en servicio su hacen verificaciones completas pero que nunca pueden reemplazar las situaciones a las que el usuario o la red se hallan sometidas y por la contraria, también es posible y necesario hacer pruebas que descarten factores de incidencia, con la razón de detectar la causa raíz de un problema.

6 Retransmisiones en TCP

La RFC 0793 explica el mecanismo de funcionamiento en TCP de la siguiente manera

> La transmisión es fiable gracias al uso de números de secuencia y de acuses de recibo. Básicamente, se le asigna un número de secuencia a cada octeto de datos. El número de secuencia del primer octeto de datos en un segmento se transmite con ese segmento y se le denomina el número de secuencia del segmento. Los segmentos también llevan un número de acuse de recibo que es el número de secuencia del siguiente octeto de datos esperado en la transmisión en el sentido inverso. Cuando el módulo de TCP transmite un segmento conteniendo datos, pone una copia en una cola de retransmisión e inicia un contador de tiempo; si llega el acuse de recibo para esos de datos, el segmento se borra de la cola. Si no se recibe el acuse de recibo dentro de un plazo de expiración, el segmento se retransmite (Postel, 1981).

Donde, en particular pondremos foco en la situación de las retransmisiones que se pueden producir por diferentes causas, a saber: demora en la recepción del acuse de recibo, perdida de dicho acuse de recibo. Ambas razones, incrementaran el tráfico en la red y por consiguiente producirán una demora en el procesamiento de la información, lo que potencialmente generaría más retransmisiones.

Sin embargo, queda claro que un problema de retransmisiones en TCP se genera a causa de la red y repercute en ella, una de las mayores dificultades surge cuando en un sistema abierto las demoras que produce la red pueden ser atribuidas a ilimitadas causas externas.

Particularmente, en esta prueba de concepto de Open RAN, se ha notado un elevado porcentaje de retransmisiones TCP en las reproducciones de archivos de video en formato streaming.

7 Análisis, Pruebas y Tiempo.

Avanzando con el tema de la dificultad para la integración de proveedores, de Open RAN en conjunto con la prueba de concepto, se ha comprobado que los indicadores de calidad de servicio del proveedor tradicional de red se veían degradados cuando los clientes que formaban parte de la operación de Open RAN hacían consumos de streaming de video en TCP.

Esta condición particular de uso de TCP depende de la OTT que ofrece el servicio (Amazon, Disney+, MovistarPlay, etc.) y no de los proveedores involucrados, por lo que era una condición que no se podía modificar.

Se pudo comprobar un aumento de retransmisiones en la zona de pruebas, que afectaban la medición de calidad del servicio brindado en la zona. El proveedor incumbente junto con el personal de operaciones de la red móvil y la empresa que se hallaba bajo prueba, de manera cooperativa comenzaron una serie de reuniones bajo la forma de mesas de trabajo para mitigar y/o solucionar el problema detectado.

En primer lugar, se decidió realizar una comparativa con un sistema de pruebas que no se hallara bajo influencia de radiación electromagnética pero que al mismo tiempo utilizara parte de la solución diseñada, con la finalidad de poder aislar el origen de la falla, usando una configuración en la cámara de compatibilidad electromagnética y de esta manera se pudo descartar la influencia de elementos externos.

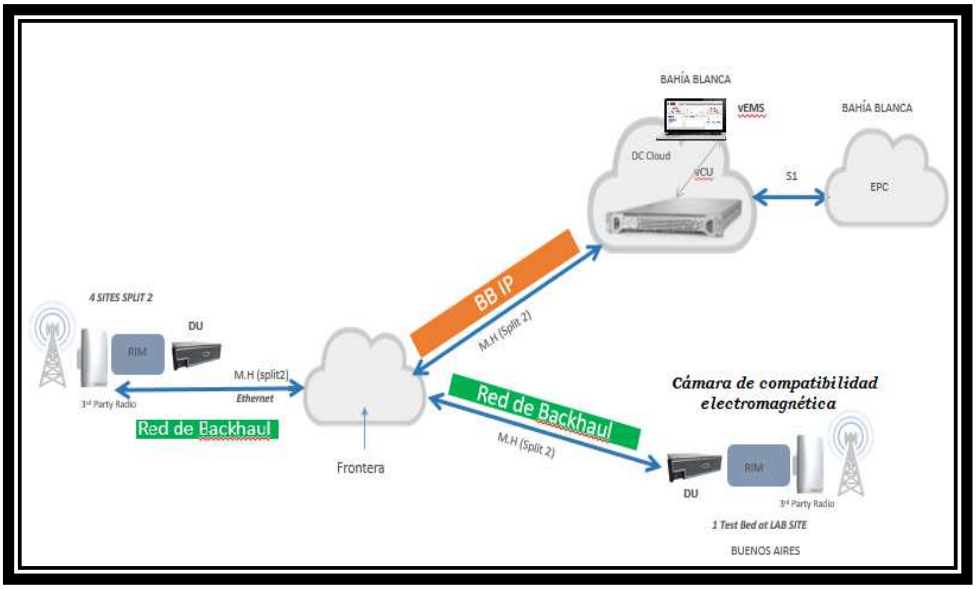


Ilustración 5 Propuesta de análisis por secciones

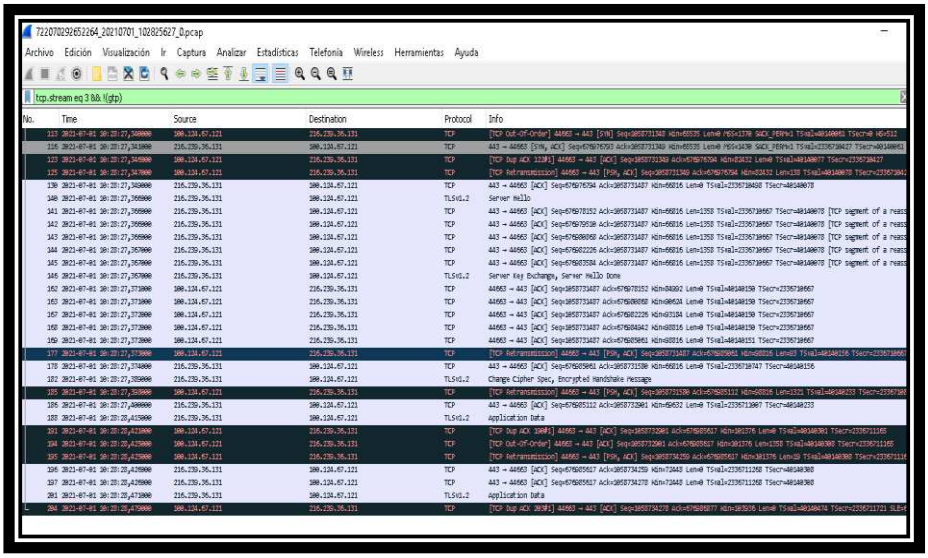


Ilustración 6 Retransmisiones medidas en cámara electromágnetica

Las mediciones de retransmisiones arrojaron los resultados de la Ilustración 6 que demuestran que las retransmisiones no tienen origen en radiaciones electromagnéticas externas a la red de acceso tanto Open RAN como tradicional.

Una vez descartado el posible ítem electromagnético, se debió recurrir a analizar la posibilidad de cambiar los contadores de espera de TCP, para verificar que no estuvieran mal programados o no se hubiera tenido en cuenta la latencia propia de las distancias desde el core de la red, hasta el lugar de la prueba.

En resumen, las pruebas demostraron que la solución se debe buscar en la red y la interconexión entre la sección Open RAN y fuera del acceso propiamente dicho (cabezales de radio frecuencia, polución electromagnética y unidades ODU) y se optó por analizar los contadores de TCP.

Con posterioridad al cambio en los tiempos de espera en los contadores, la situación no presentó variaciones en cuanto a las retransmisiones y su presencia.

La mesa de trabajo en su última reunión coincidió, tras el análisis de varias mediciones, usando generadores de descargas de diferente tamaño en bytes, en que existía una relación directa entre el tamaño de la descarga y el porcentaje de retransmisiones observado.

Tras poco más de dos meses de trabajos y mediciones no se ha podido determinar el origen de las retransmisiones de TCP que cuando son producidas por archivos de tamaño reducido afectan el funcionamiento de la zona de atención. Esto, si bien no es una situación común en la red es un posible foco de ataque por negación de servicio que debe ser considerado.

La ilustración siguiente muestra la afectación de una celda de la zona cuando se dan las condiciones descritas y cuando no ocurre el streaming de video por TCP de archivos de pequeño tamaño.

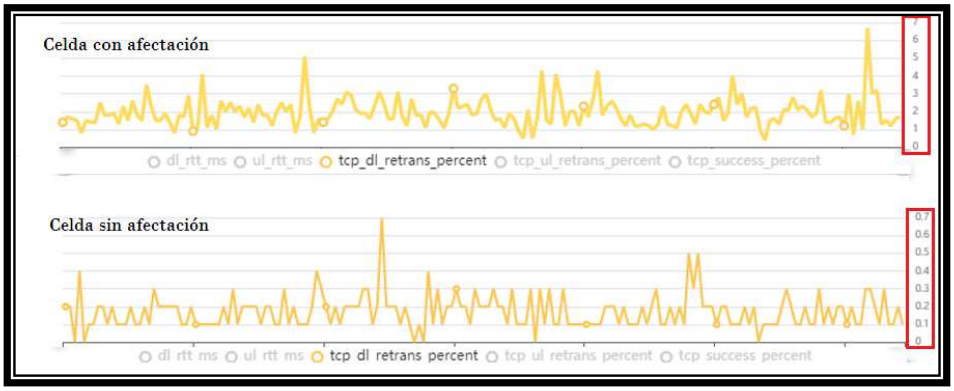


Ilustración 7 Afectación de una celda

En conclusión, a la etapa de pruebas, análisis y mediciones, la solución para reducir los tiempos de espera de los contadores de TCP ha sido beneficiosa para el funcionamiento general de la red móvil, no ha influido de manera notable en solucionar la cantidad de retransmisiones en la zona Open RAN. Esto no representa un problema, porque forma parte del aprendizaje de ambos sectores, y abre un camino de experiencia y análisis a futuros problemas.

8 Conclusiones

El punto de inicio de este desarrollo teórico ha sido la experiencia de campo de los expertos en Telecomunicaciones, que han manifestado cierta renuencia de las empresas a invertir sin un tiempo de retorno de las inversiones acotado, y la dificultad que se presenta al trabajar con múltiples proveedores.

La experiencia de las pruebas de concepto viene a subsanar en alguna medida la mencionada demora en los retornos de inversión, mientras que se presenta como barrera infranqueable para los casos en que los proveedores no tienen una sólida experiencia en puesta en marcha y mantenimiento de los sistemas que comercializan. La complejidad de los nuevos sistemas parece ser tanto su fortaleza como su talón de Aquiles y debe tenerse consideración de los tiempos involucrados en la plena funcionalidad del sistema general.

A modo de síntesis, podemos concluir que la búsqueda de rápidos retornos de inversión, no puede ser basada en la integración de operadores diferentes porque ello eleva los tiempos de respuesta y curvas de aprendizaje si no se trabaja con proveedores de sólida experiencia en el rubro.

Las nuevas tecnologías debido a ser nuevas, no siempre se hallan en una etapa de maduración importante, por el contrario, la competencia entre proveedores genera la búsqueda del sistema novedoso que permita mantener las ventas y debe salir al mercado con rapidez, por lo tanto, si bien el retorno rápido de inversiones puede tomar dependencia directa con un ahorro de costos debido la competencia de proveedores no puede relacionarse con la velocidad de funcionamiento de un sistema.

Sobre la experiencia particular de insertar un entorno Open RAN en una red tradicional funcionando, un punto a mejorar ha sido la dificultad para coordinar las mediciones en todos los puntos de la red y la comparación de esas mediciones, tómese como ejemplo que la prueba de descarga de un archivo usando celulares, incluye el trabajo de personal en zona, personal conectado de manera remota a los equipamientos intervinientes en el servicio y personal de los proveedores (generalmente con husos horarios diferentes), dicha diferencia horaria, puede resultar una barrera elevada de franquear.

9 Referencias

Roca, J. L., Biga, D., & Del Giorgio, H. (1 de Diciembre de 2012). *Secretaria de Ciencia y Tecnología UNLaM.* Obtenido de https://cyt.unlam.edu.ar/index.php?seccion=17&idArticulo=672

Dufour Fernando Javier;Micieli Gustavo;Serra Ariel. (1 de 11 de 2015). C189_Redes LTE. San Justo, Bs.As, Argentina.

Cannizzaro, A. (15 de 10 de 2014). *www.conicet.gov.ar.* Obtenido de Puerto Madryn aumentó su población 14 veces desde 1970 hasta la actualidad: https://www.conicet.gov.ar/puerto-madryn-aumento-su-poblacion-14-veces-desde-1970-hasta-la-actualidad/

O-RAN ALLIANCE. (2018). Acerca de O-RAN ALLIANCE.

Postel, J. (Septiembre de 1981). *IETF.* Obtenido de Transmission Control Protocol: https://datatracker.ietf.org/doc/html/rfc6093

ISBN 978–987–633–574–4

Estrategias de Pre-procesamiento de Datos para el Análisis de Tráfico de Redes como Problema Big Data

Mercedes Barrionuevo[1], María Fabiana Piccoli[1]
[1] Universidad Nacional de San Luis
Ejército de los Andes 950, San Luis, Argentina
{mdbarrio, mfpiccoli}@unsl.edu.ar

Abstract. Detectar posibles ataques a una red de computadoras requiere contar con métodos o estrategias trabajando en conjunto para la clasificación del tráfico. El área constituye un problema básico de amplio interés sobre todo en conceptos emergentes como Big Data, con sus nuevas tecnologías para almacenar, procesar y obtener información a partir de grandes cantidades de datos.

El reconocimiento del tráfico malicioso en una red depende, en primera instancia, de la eficiencia en la recolección de datos y su correcto pre-procesamiento a fin de ser lo más representativo al aplicar el modelo de análisis de datos elegido. Este tema es el abordado en este trabajo, formando parte de un proyecto integral de detección de ataques a redes de computadoras aplicando Computación de Alto Desempeño en GPU, Inteligencia Artificial y Procesamiento de Imágenes

Keywords: Big Data. Tráfico de redes. Normalización y limpieza de datos. Ataques.

1 Introducción

En la actualidad, la información, los sistemas y las redes informáticas brindan un gran apoyo a diversas empresas y organizaciones convirtiéndose en importantes recursos para las mismas. La confidencialidad, integridad y disponibilidad de la información resultan esenciales para mantener la ventaja competitiva, la rentabilidad, el cumplimiento de las leyes y la imagen institucional. Sin embargo, las organizaciones, sus redes y sistemas de información se enfrentan en forma creciente y constante a amenazas, las cuales buscan afectar la seguridad informática [1].

Un procedimiento de detección de anomalías en una red debe ser capaz de hacer frente al constante incremento en el número de ataques y al gran volumen de datos transferidos, buscando dar soporte a las tareas de monitoreo de red y de identificación de comportamiento anómalo o ataques a las redes [2]. Es por ello que el problema se lo considera un problema Big Data o de Datos Masivos.

La tarea de aprender a detectar ataques implica construir un modelo predictivo, un clasificador, capaz de distinguir entre conexiones “malas”, llamadas intrusiones o ataques, y conexiones normales o “buenas” [3].

Cuando hablamos de conexiones hacemos referencia a una secuencia de paquetes [3] que comienzan y terminan en momentos bien definidos, donde los datos fluyen desde una dirección IP origen a una dirección IP destino según un protocolo bien definido. Algunos expertos en intrusiones creen que la mayoría de los ataques novedosos son variantes de ataques conocidos y la "*firma*" de éstos puede ser

suficiente para detectar variantes novedosas [2]. Por lo tanto, tiene sentido seguir analizando sus variaciones.

Para realizar un buen análisis de las conexiones es necesario una adecuada "preparación de los datos" en la que se eliminen o corrijan aquellos incorrectos y se decida la estrategia a seguir con los incompletos o faltantes. Además, se proyectan los datos para considerar únicamente aquellas variables o atributos que van a ser relevantes, con el objetivo de hacer más fácil la tarea propia de análisis de datos. Esta etapa incluye la selección, limpieza y transformación de los datos; y, a su vez, consta de cuatro subfases: selección de datos, limpieza de datos, construcción de datos (atributos derivados, registros generados), y formateo de datos [4].

En [5, 6] se presentaron resultados satisfactorios de un sistema para la detección de ataques usando algunos específicos. Como la base de datos utilizada para la evaluación de los algoritmos de clasificación eran de prueba, estas ya tenían sus datos en el formato requerido. Es por ello que es de interés en este trabajo hacer frente a la Etapa 2 mostrada en la Figura 2.2: Etapa de Pre-procesamiento de Datos.

Por lo tanto, el objetivo planteado en este trabajo es mostrar el procesamiento realizado a los datos recolectados directamente del tráfico de red, previo a la aplicación de los algoritmos paralelos de minería de datos y de visualización mostrados en [5,6] para la detección de ataques o posibles anomalías.

Este documento está organizado como sigue: la siguiente sección describe los conceptos teóricos involucrados en el desarrollo de este trabajo. La sección 3 detalla el preprocesamiento, transformación y limpieza realizado a los datos. Finalmente se muestran los resultados experimentales obtenidos, y se detallan las conclusiones y líneas futuras de trabajo.

2 Marco Teórico

En esta sección se analizan brevemente diferentes conceptos, entre los cuales se destacan distintos aspectos referidos a los datos en contextos de Big Data, su normalización y correlación. Todo relacionado con el problema que nos interesa: el análisis de tráfico de redes para la detección de ataques. Cada uno de estos temas se aborda en las siguientes secciones.

2.1 Big Data

El volumen de datos circulante en la red de redes ha alcanzado niveles inimaginables en la última década y, al mismo tiempo, los dispositivos de almacenamiento han reducido de forma significativa sus precios. Las empresas privadas e instituciones de investigación capturan terabytes de datos de la interacción de los usuarios, redes sociales y de diversos sensores.

La clave en la era de Big data son los datos, los cuales se pueden utilizar para responder a muchas preguntas, pero no a todas. En la actualidad el trabajo con datos presenta ciertos retos a afrontar:

- El aumento masivo del volumen de datos puede implicar una disminución en la calidad del análisis.

■ En los grandes volúmenes de datos no siempre hay contexto, por lo cual se deberá contar con expertos del tema, por ejemplo, con ingenieros en redes de telecomunicaciones.
■ Los datos cambian cada cierto tiempo, por lo cual pueden llegar a generar inconsistencias, si se ha tenido en cuenta sólo un tipo de dato de entrada.
■ Los datos que comprueban las hipótesis planteadas pueden ser difíciles de obtener.

Por ello, el desafío de esta era es darle sentido a este gran conjunto de datos. El análisis de datos en Big Data involucra recolectar datos de diferentes fuentes, unirlos y/o mezclarlos para ser tratados por los analistas, para finalmente, entregar resultados de utilidad a la organización de interés.

El proceso de convertir grandes cantidades de datos no estructurados para ser datos útiles a las organizaciones no es una tarea trivial. Por lo tanto, se deben combinar diferentes estrategias y metodologías para lograr una mejor respuesta.

Al trabajar con datos, uno de los primeros pasos necesarios para hacer un análisis de datos es determinar qué tipo de estrategia usar: descriptiva, exploratoria, inferencial, predictiva, causal o mecánica. Las respuestas a las preguntas mostradas en la Figura 2.1 determinarán el enfoque a usar en la resolución del problema.

En nuestro caso, buscamos predecir si una conexión cumple o no con ser un ataque. Parece un buen inicio utilizar un análisis predictivo, siempre y cuando, los datos del tráfico de una red hayan podido ser puestos en un estado homogéneo. En consecuencia, se debe preparar los datos para un enfoque predictivo.

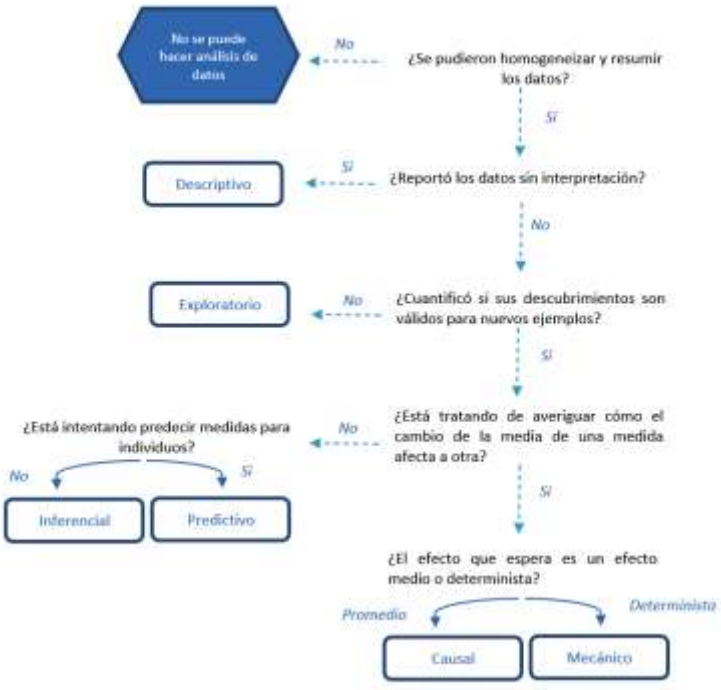


Fig 2.1 Diagrama de flujo del tipo de análisis de datos según la pregunta

En las siguientes secciones detallamos el ciclo de vida de los datos en entornos Big Data, la metodología a implementar y las estrategias empleadas para realizar el pre-procesamiento de los datos.

2.1.1 Ciclo de vida de Big Data

Existen metodologías como SEMMA [11] que son incompletas dado que ignoran las etapas de recopilación de datos. Estas etapas generalmente constituyen la mayor parte del trabajo en un proyecto exitoso de Big Data.

En la figura 2.2 se muestran las etapas por las que tienen que pasar los datos en un proceso de Big Data, siendo la Etapa 2 de Pre-procesamiento de los datos nuestro punto de interés. Este ciclo comienza luego de definir el problema y evaluar correctamente cuánto potencial de ganancia tiene para una organización, siendo buena estrategia investigar y/o analizar lo que otras organizaciones han implementado en la misma situación estudiando soluciones que sean razonables para su compañía.

Lo primero a considerar en un ciclo de análisis de Big Data es “adquirir y recopilar los datos” definiendo cuáles datos serán de relevancia: Etapa 1.

Fig 2.2: Etapas de Big Data

Una vez que los datos son recuperados, es necesario almacenarlos en un formato fácil y apto de usar. Luego es necesario almacenarlos en una base de datos siendo una de las alternativas más comunes a utilizar el sistema de archivos Hadoop [7], Spark [8], entre otros; “preprocesar los datos”, implica remodelar los datos limpios recuperados previamente y usar estadísticas para la imputación de valores perdidos, detección de valores atípicos, normalización, extracción de características y selección de funciones: Etapa 2.

“Modelar y analizar datos” implica probar diferentes modelos y esperar resolver el problema empresarial en cuestión. En la práctica, normalmente se desea que el modelo proporcione algunos conocimientos del negocio seleccionando el mejor modelo o combinación de modelos evaluando su rendimiento en un conjunto de datos excluido: Etapa 3; y finalmente, “implementar y evaluar” esta etapa implicaría aplicar el modelo a nuevos datos y una vez que la respuesta esté disponible, evaluar el modelo: Etapa 4.

2.1.2 Metodología

En términos de metodología, el análisis de Big Data difiere significativamente del enfoque estadístico tradicional. El análisis comienza con los datos, luego se los modela para obtener una respuesta.

En aplicaciones de análisis a gran escala, se necesita una gran cantidad de trabajo (normalmente el 80% del esfuerzo) sólo en la limpieza de los datos, para ser utilizado posteriormente por modelos de aprendizaje automático.

Una vez que los datos se pre-procesan están disponibles para el modelado, los resultados de las evaluaciones en los diferentes modelos deben ser los razonables y/o esperados. Finalmente, una vez implementado el modelo, se deben informar tales evaluaciones y resultados adicionales al experto en el problema, para que analice si la información obtenida le aporta conocimiento o no.

Si bien no se tiene una metodología única a seguir en aplicaciones reales a gran escala, normalmente una vez definido el problema se aplican estas pautas generales en la mayoría de los problemas.

2.2 Aspectos Generales del Pre-procesamiento de Datos

La etapa de preparación de los datos consiste en aplicar limpieza, normalización de los datos y la selección de características. Cada una de estas fases se describen brevemente en las siguientes secciones.

2.2.1 Limpieza de datos

Una vez que los datos son recolectados, pueden existir diversas fuentes de datos con diferentes cantidades de atributos. Es importante preguntarse si es práctico homogeneizarlos.

Si las fuentes de datos son completamente diferentes, la pérdida de información puede resultar muy grande al homogeneizarlas. En este caso, podemos pensar en alternativas. ¿Puede una fuente de datos ayudarnos a construir un modelo de regresión y la otra un modelo de clasificación? ¿Es posible trabajar con la heterogeneidad a nuestro favor en lugar de simplemente perder información? Tomar estas decisiones es lo que hace que la analítica sea interesante y desafiante.

En este punto necesitamos limpiar los datos no estructurados, convertirlos en una matriz de datos para aplicar algún algoritmo como así también eliminar o reemplazar datos con valores nulos. Particularmente, para el problema que nos convoca los datos pueden ser recolectados por los administradores de red usando distintas herramientas, las cuales generan datos con distintos formatos y variados atributos.

2.2.2 Normalización de Datos

En muchos algoritmos basados en distancias es necesario escalar los datos, es decir normalizar el rango de valores numéricos, las distancias debidas a diferencias de un

atributo que van entre 0 y 1000 serán mucho mayores que aquellas debidas a diferencias de un atributo variando entre 0 y 10.

Como consecuencia, es necesario aplicar alguna función de normalización a los datos. Para ésto existen muchos métodos, siendo la técnica *z-score* la más utilizada por su sencillez en el cálculo. Este método conserva el rango (máximo y mínimo) e introduce la dispersión de la serie (desviación estándar/varianza), transformando linealmente los valores de tal manera que el valor medio de los datos transformados es igual a 0 mientras que su desviación estándar es igual a 1. La fórmula de transformación es la correspondiente a la ecuación (1).

$$x = (x_1 - \mu) / \sigma \; . \qquad (1)$$

Donde x es la muestra actual, x_1 es la muestra transformada, μ denota la media de los datos y σ representa la desviación estándar.

2.2.3 Selección de Características

La selección de características es fundamental para la detección de ataques o anomalías. Este proceso consiste en dar un peso a cada característica para determinar cuál de ellas es la que tiene mayor impacto.

La ponderación de características mejora la precisión, logrando un mayor rendimiento. Las métricas comúnmente conocidas para la selección de características son *chi-cuadrado* (CHI), *ganancia de información*, *coeficiente de correlación* y *razón de probabilidades* (OR)[9].

Por tratarse de valores numéricos y por su simplicidad en el cálculo, la métrica utilizada en este trabajo es el *Coeficiente de Correlación* entre variables.

La correlación, también conocida como *Coeficiente de Correlación Lineal* (de Pearson), es una medida de regresión que pretende cuantificar el grado de variación conjunta entre dos variables. Es una medida estadística que cuantifica la dependencia lineal entre dos variables, es decir, si se representan en un diagrama de dispersión los valores que toman dos variables, señalará lo bien o lo mal que el conjunto de puntos representados se aproxima a una recta. Formalmente, la podemos definir como el número que mide el grado de intensidad y el sentido de la relación entre dos variables, ver ecuación **(2)** [10].

$$\rho(x,y) = \mathrm{cov}(x,y) \; / \; \sigma x * \sigma y \; . \qquad (2)$$

Siendo la covarianza entre dos variables definida como:

$$\mathrm{cov}(x,y) = (\Sigma \; (x_i - x)(y_i - y))/n \; . \text{ para } i=1...n$$

Los valores que puede tomar la correlación son: $\rho = -1$ para la correlación perfecta negativa, $\rho = 0$ cuando no existe correlación y $\rho = +1$ para la correlación perfecta positiva.

La Limpieza, Normalización y Selección de Características forman parte de la Etapa 2 mostrada en la Figura 2.2 referida al preprocesamiento de los datos. Llevar adelante estas etapas dan origen a la propuesta de este trabajo.

3 Pre-procesamiento del Tráfico de Redes

Este trabajo forma parte de un proyecto integral, el cual aplica un modelo de aprendizaje predictivo, donde se combinan técnicas de clasificación, análisis por similitud, visualización de datos y Computación de Alto desempeño en su solución.

El problema a afrontar implica reconocer en un tiempo razonable ataques a una red, y de una manera lo más confiable posible. Para iniciar con esta tarea se define a $X = \{x_1, x_2, \ldots, x_n\}$ como una conexión de red, donde cada atributo representa los valores intervinientes en una comunicación. Por cada conexión, se evalúa intentando predecir si es normal, un ataque o una anomalía teniendo en cuenta los valores de cada uno de los atributos y sus relaciones.

Luego de definir el problema, se debe contar con los datos a analizar, ésto se logra mediante la recolección de los datos de la red. A continuación, se describen los pasos realizados para la normalización, utilizando la técnica z-score, y la selección de características, según la correlación de datos, ambas descritas anteriormente. Las tareas a desarrollar son:

- **Eliminación de datos Nulos y Anómalos:** Una vez que los datos son recolectados pueden haber valores fuera de lo normal o valores faltantes en algunos de los atributos recolectados. Por lo tanto, se deben eliminar aquellos datos donde no existe ni dirección IP origen ni destino, o son direcciones de multicast o broadcast limitado. Éstas últimas no aportan información útil para las reglas en la clasificación de los datos.

Aproximadamente el 20% de los datos son eliminados por ser del tipo multicast, broadcast ilimitado o poseer valores nulos. Las conexiones con valores nulos deben analizarse por separado para evaluar cuáles son los datos faltantes y si son anomalías.

- **Selección de Características:** La normalización de los datos es un proceso costoso desde el punto de vista computacional, más cuando se trabaja con mucha cantidad de datos como es este caso. Por ello, es necesario, previo a la normalización, determinar cuáles serán las características con las que se trabajará para determinar si existe un ataque, anomalía o no. Como se mencionó anteriormente, nosotros seleccionamos aquellas características independientes entre sí, en consecuencia, su determinación será según el coeficiente de correlación, el mismo se obtiene aplicando (2) en cada una de las características. Serán seleccionadas aquellas que estén más cercanas a cero o no superen un umbral de correlación definido.
- **Normalización de los Datos:** Al trabajar con la gran cantidad de datos circulantes en una red, existen muchos atributos con distintos rangos de valores. Por ejemplo, al convertir una dirección IP a un número decimal, el valor máximo es 4.294.967.295 si es un broadcast (255.255.255.255), sin embargo, otros atributos pueden tomar valores entre 0 a 1024 si se trata de evaluar puertos bien conocidos.

Para cada uno de los atributos seleccionados, se procede con su normalización aplicando la ecuación (1). Para la misma se debe calcular previamente la media y la varianza del conjunto de datos.

Una vez que los datos han sido recolectados, transformados, normalizados y seleccionados, se continúa con la siguiente etapa del proceso, por ejemplo, aplicar reglas de clasificación para determinar cuáles de esas conexiones son ataques conocidos generando *firmas*, y luego, establecer similitudes entre las conexiones y las *firmas* para determinar potenciales ataques.

4 Resultados Experimentales

En esta sección se presentan los experimentos realizados en el Laboratorio de Redes de la Universidad Nacional de San Luis y el análisis de los resultados obtenidos. Cada una de las etapas consideradas se realizaron de la siguiente manera:

- **Recolección de Datos.**
En nuestro caso usamos la herramienta *tshark* durante un día generando 24 archivos (1 por hora), donde los atributos recuperados son: direcciones IP origen y destino, puertos origen y destino, y protocolo utilizado en la comunicación. A partir de ellos se generan nuevos atributos tales como: clase, número de red y host de cada dirección IP. Esto es necesario para aplicar las reglas de clasificación a cada conexión.
El comando utilizado es:
tshark -f \tcp or udp or icmp{T elds |E separator=, -e frame.time relative -e ip.src -e ip.dst -e tcp.srcport -e tcp.dstport -e udp.srcport -e udp.dstport -nni eth0 > trafico_a_analizar.txt

- **La correlación entre los datos**
En la Tabla 1 se puede observar el nivel de correlación de los atributos donde las direcciones IP están fuertemente relacionadas con sus respectivas clases, red y host con valores cercanos a 1. Mientras que existe muy baja o nula dependencia con el resto de los atributos.

Tabla 1. Correlación entre los atributos de las conexiones.

	clase_o	red_o	host_o	pto_o	pto_d	prot	ip_d	red_d	host_d	clase_d
ip_o	0,75	0,75	-0,64	-0,15	-0,11	-0,46	-0,11	-0,08	-0,05	-0,05
ip_d	-0,05	-0,05	-0,08	-0,15	-0,11	-0,48	1,00	0,69	-0,66	0,69
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
prot	0.30	0.20	0.10	0.40	0.40	1.00	0.30	0.20	0.10	0.01

- **Selección de características:**

De las pruebas realizadas anteriormente se pudo determinar que el análisis de correlaciones entre los atributos nos permite decidir cuáles son los parámetros correlacionados o dependientes, estableciendo cuáles son los atributos significativos a ser considerados, por ejemplo, en el cálculo de la función euclidiana utilizados por el algoritmo *k-nn* para la evaluación de la similitud de las conexiones con las firmas de ataques conocidos. En este caso, de los 11 parámetros utilizados para las reglas de clasificación sólo son considerados 5 *(dirección IP origen y destino, puerto origen y destino y protocolo)* los que nos interesan. Esta reducción de atributos permite mitigar el alto costo computacional involucrado en el cálculo de la función euclidiana para cada una de las conexiones.

- **La normalización de los datos:**

Se realizó tomando los atributos seleccionados de cada conexión: (dirección IP origen y destino, puertos origen y destino) y aplicándoles la función z score a cada uno de ellos, dando como resultado los valores como se muestran en la Tabla 2.

Tabla 2. Normalización de los datos.

IP origen	IP destino	pto origen	pto destino	prot
-0,3152596	-0,33519186	-0,3274522	-0,48115293	-1,4115906
-0,3152595	-0,33519186	-0,4152103	-0,47791469	0,290621
……..	……..	……..	…….	……..
-0,31525969	-0,335191866	-0,42403127	-0,48569784	1,99283384

Estos valores se corresponden para el análisis de un día del tráfico de la red del Laboratorio, según los paquetes obtenidos y seleccionados después de la limpieza.

Una vez realizada la etapa de pre-procesamiento de los datos, se comprobaron los datos obtenidos aplicando los modelos que incluyen las reglas de clasificación utilizadas y el algoritmo de los *k*-nn más cercanos para la detección de valores similares a ataques conocidos. En la Tabla 3 se muestra el valor del cálculo de la función para distintas 5-uplas. Se incluye en dicha tabla una columna donde se especifica si es un ataque o no y otra columna con el valor de la función euclidiana.

Tabla 3. Valor de la función euclidiana

IP origen	IP destino	pto origen	pto destino	prot	es_ataque	f. euclidiana
-0,315	-0,335	-0,327	-0,481	-1,411	si	0
-0,315	0,652	2,937	-0,460	-1,411	no	0,987
-0,315	-0,335	-0,415	-0,477	0,290	no	1,702
-0,315	-0,335	-0,415	-0,477	0,290	no	1,702

En este caso el algoritmo *k-nn* se ejecuta con k=4, devolviendo las 4 filas mostradas en la Tabla 3. Donde si se realiza la inversa de la normalización, se puede observar que para los casos en que la función era 0, la conexión analizada es exactamente igual a la firma considerada como ataque, mientras que los otros valores muestran ser anomalías. En estos casos se observa que son ataques a otros puertos, a otro protocolo o son ataques sin puertos especificados. Esto se muestra en la Tabla 4 en un formato entendible para el experto del dominio de redes.

Tabla 4. Inversa de la Normalización

IP origen	IP destino	pto origen	pto destino	prot
10.230.34.73	10.255.255.255	1500	80	tcp
10.230.34.146	10.255.255.255	137	137	udp
10.230.34.146	10.255.255.255	137	137	udp
10.230.34.12	10.255.255.255			icmp

5 Conclusiones y Trabajos Futuros

Este trabajo presenta una metodología a aplicar en la etapa de pre-procesamiento de los datos en un sistema de Big Data, particularmente aplicado al dominio de seguridad en el tráfico de Redes de Computadoras. Para ello proponemos llevar a cabo la selección de características y la normalización de los datos mediante funciones estadísticas bien conocidas.

La aplicación de esta propuesta fue evaluada en una red, mostrando resultados satisfactorios, no sólo respecto a las respuestas obtenidas sino también al desempeño del sistema en general al realizarse la limpieza de datos.

Como líneas futuras, se propone aplicar técnicas de programación paralela en la normalización y análisis de correlación, como así también en la secuenciación de las etapas de manera de crear estructuras similares a arquitecturas de pipeline y *overlaping* de etapas, particularmente en la Etapa 1 y 2. Además, se prevé ampliar los conocimientos usando diversas técnicas de aprendizaje de máquina y/o redes neuronales a fin de comparar los modelos utilizados en este trabajo y, de ser necesario, mejorar los existentes.

Referencias

1. Tulasi.B, R. S. Wagh, and B. S., “High performance computing and big data analytics-paradigms and challenges,” International Journal of Computer Applications, vol. 116, Abril 2015.
2. Terzi,D. S., Terzi, R. and Sagiroglu, S. “Big data Analytics for Network Anomaly Detection from Netflow Data,” IEEE, 2017.
3. Ghimes, A. M., and Patriciu,V. V. “Neural network models in big data analytics and cyber security,”in 2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), pp. 1–6, June 2017.
4. Hernández Orallo, J.; Ramírez Quintana, M. J.; Ferri Ramírez, C. “Introducción a la Minería De Datos” ISBN eBook: 978-84-8322-558-5.
5. Barrionuevo M., Lopresti M., Miranda N., Piccoli F.. “Secure Computer Network: Strategies and Challenges in Big Data Era”. JCC&BD 2018. VI Jornadas de Cloud Computing & Big Data. La Plata (Buenos Aires), 25 al 29 de junio de 2018. ISBN 978-950-34-1659-4
6. Barrionuevo M., Lopresti M., Miranda N., Piccoli F.. “An Anomaly Detection Model in a LAN using K-NN and High Performance Computing Techniques”. Congreso Argentino de Ciencias de la Computación. CACIC 2017. http://sedici.unlp.edu.ar/handle/10915/63951
7. White, T.. “Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale”. O'Reilly Media, Inc. ISBN 1491901713, 9781491901717. 2015
8. Perrin, J. “Spark in Action, Second Edition: Covers Apache Spark 3 with Examples in Java, Python, and Scala”. 2do Edition. ISBN 1617295523, 9781617295522. Simon and Schuster, 2020.
9. Ikram,S., Kumar, C. “Intrusion detection model using fusion of chi-square feature selection and multi class SVM.” J. King Saud Univ. Comput. Inf. Sci. 29(4): 462-472. 2017.
10. Peiro Ucha, A.. “Coeficiente de correlación lineal”. Economipedia.com. 2015.
11. Azevedo, A., Santos, M. “KDD, SEMMA and CRISP-DM: a parallel overview”. IADIS European Conf. Data Mining. Pp 182-185. 2015.

ISBN 978-987-633-574-4

CACIC 2021

WORKSHOP INNOVACION EN SISTEMAS DE SOFTWARE

COORDINADORES

Pablo Fillottrani (UNS)
Marcelo Estayno (UNLZ)
Alicia Mon (ITBA)
Dante Zanarini (UNR)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978–987–633–574–4

El desafío de Implementar DevOps en una Organización del Estado en Tierra del Fuego

Ezequiel Moyano[1], Daniel Aguil Mallea[1], Cintia Aguado[1], Ana Karina Manzaraz[1]

[1] IDEI, UNTDF, H.Yrigoyen 879, 9410, Ushuaia, Argentina
{emoyano, daguil, caguado, amanzaraz}@untdf.edu.ar

Resumen. DevOps es una metodología de trabajo mediante el uso de nuevas herramientas y prácticas, que consiste en eliminar las barreras entre los equipos de desarrollo y operaciones en un área de IT. El objetivo de DevOps es optimizar y agilizar el ciclo de desarrollo (flujo de valor), potenciando la cultura de equipo (centrándose en la colaboración y comunicación de sus miembros), etc.; produciendo entregas continuas tendiendo a reducir el tiempo entre el momento en que se genera un cambio (productividad de desarrollo) y el momento en que se aplica en el entorno de producción (confiabilidad de las operaciones)

DevOps plantea la necesidad de un cambio cultural hacia la colaboración e integración, las empresas innovadoras y líderes disponen de equipos que visualizan todo el ciclo de vida del desarrollo y la infraestructura como parte de sus responsabilidades; acelerar del tiempo de comercialización, la mejora de la calidad, la satisfacción del cliente, el lanzamiento confiable, mejora de la productividad y la eficiencia son beneficios claves que motivan su implementación.

Sin embargo existen organizaciones con modelos de desarrollo tradicional que dificultan la adopción de estas prácticas sin antes realizar un fuerte cambio cultural y organizativo, adaptando o creando estructuras de equipo de trabajo acordes en sus áreas de IT.

El presente artículo refleja una experiencia de haber implementado DevOps en una organización del Estado en Tierra del Fuego.

Palabras clave: DevOps, Innovación de Software, Organismo Público, Full Stack Engineers.

1. Introducción

El término DevOps proviene de la unión de Desarrollo (dev) y Operaciones (ops); consiste en eliminar las barreras entre los equipos de desarrollo y operaciones, un cambio cultural hacia la coordinación y colaboración entre disciplinas aisladas[1].

DevOps tiene tres pilares, *cultura* que es la forma en la que las personas trabajan y colaboran en una organización, que facilite el desarrollo de equipos de alto rendimiento; *prácticas* que representan la implementación de la cultura en pos de la automatización y mejora continua; y *herramientas* (tecnologías específicas)[2].

El objetivo de DevOps es optimizar el value stream (flujo de valor) y el tiempo de entrega de un producto, eliminando fricciones humanas, procedimientos manuales, etc. permitiendo reducir el tiempo entre el momento en que se genera un cambio (productividad de desarrollo) y el momento en que se aplica en el entorno de producción (confiabilidad de las operaciones)[3][4]. Plantea la necesidad de un cambio cultural hacia la colaboración e integración, se trata de permitir que los diferentes equipos se comuniquen y trabajen mejor[5]. No es una estructura organizacional, más bien define una forma de organizar equipos independientes (multifuncionales o "full stack").

DevOps es un enfoque basado en principios ágiles y lean, donde los equipos de desarrollo y operaciones colaboran para entregar software de manera estable e ininterrumpida[6], a través de una implementación sincronizada en plataformas diferentes para reducir los costos de IT.

2. Beneficios de implementar DevOps

El uso de prácticas DevOps contribuyen a una mejor eficiencia organizativa[7]. Las empresas innovadoras y líderes disponen de equipos que visualizan todo el ciclo de vida del desarrollo y la infraestructura como parte de sus responsabilidades.

Esta nueva dinámica trata poner en alza algunos aspectos del trabajo diario (buena intercomunicación y colaboración entre equipos, reparto de responsabilidades, aprender de errores pasados, innovación, etc.); las empresas que tienen estructuras inadecuadas provocan equipos ineficientes. La falta de algunas prácticas hace difícil predecir cuánto tiempo llevará un desarrollo y, peor aún, es muy difícil saber qué tan avanzado está el proyecto; con DevOps en todo momento el equipo sabe dónde está, qué funciona y qué no.

Los principales motivos estratégicos por los que una organización debe considerar implementar un modelo de DevOps son la reducción de costos, mejora de la calidad, mejora de productividad, innovación, velocidad en el proceso de desarrollo de Software, escalabilidad de las aplicaciones y entornos colaborativos, entre otros.

Esta metodología ocupa un lugar distintivo en muchos proyectos a corto y medio plazo de varias áreas de IT a nivel mundial. La implementación de DevOps es cada vez más importante[8]. La figura 1 muestra, a nivel mundial, sobre un total de 100 empresas líderes, la implementación de DevOps.

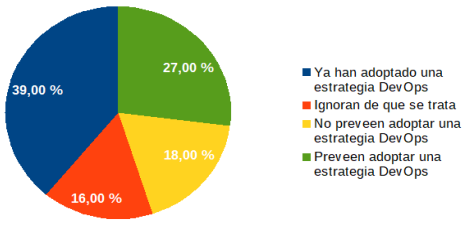


Figura 1. Adopción de DevOps.

3. Problemas presentes en empresas con equipos tradicionales

Los equipos de trabajo tradicionales están formados a partir de estructuras que surgen de la organización, siguiendo líneas jerárquicas establecidas y formales; la división del trabajo suele estar agrupada por las principales actividades, es decir, están constituidas en departamentos claramente identificados. Los líderes de los equipos son nombrados por la organización y tienen poder legítimo en el equipo, donde el mayor problema es la falta de comunicación entre los equipos funcionales.

En los equipos tradicionales existen barreras bien definidas, las aéreas suelen estar divididas en sectores independientes con jerarquías fijas, preocupados en resolver los problemas que le afectan directamente, falta de comunicación entre los equipos, etc.; lo que presenta muchos inconvenientes y problemáticas que afectan directamente a la productividad (la cual se ve comprometida) y, sobre todo, un gran interrogante ¿Por qué a las empresas tradicionales les cuesta innovar?

Si bien son varias son las causas de por qué a las empresas con una estructura organizativa tradicional les cuesta innovar, entre las más importantes es que su estructura organizativa no les permite adaptarse a los cambios. Adoptar una metodología de DevOps dentro de las organizaciones es fundamental, sin embargo, las organizaciones deben crear una estructura de equipo de trabajo adecuada para implementar sus prácticas.

Cuando nos referimos a organizaciones dependientes de la administración pública, estas problemáticas se acentúan muchísimo más. En el ámbito público, en la dinámica del desarrollo existen otros factores que afectan a los proyectos y a los procesos, que inciden en la calidad, en los plazos de entrega y en los costos asociados.

La falta de capacitación del personal es otro problema sustancial, puesto que se requiere que los equipos de desarrollo sean capaces de trabajar en forma sistemática, disciplinada y cuantificable; consecuentemente, los integrantes de estos equipos deben estar capacitados para hacerlo de esa manera. En el contexto regional del estudio del presente artículo, en el marco de la provincia de Tierra del Fuego AeIAS, los recursos humanos formados en la disciplina son particularmente escasos; motivo por el cual es usual encontrarse con dificultades al intentar que un equipo de desarrollo utilice enfoques metodológicos para obtener mejores resultados. Si sumamos los inconvenientes derivados de las planificaciones organizativas existentes en las áreas de sistemas, estos resultados son aún más pobres.

En esta realidad los proyectos de desarrollo tienen más probabilidad de no completarse con el tiempo; la falta de comunicación dentro de los equipos, el nivel de confusión y caos existentes, con frecuencia prevalecen en los grandes proyectos de software.

4. Incorporar DevOps en la organización

Para poder implementar una metodología DevOps en una organización, es importante conocer cuáles son sus principales prácticas. Existen unas cuantas prácticas fundamentales que ayudan a las organizaciones a innovar con mayor rapidez mediante la

automatización y la simplificación de los procesos; una práctica fundamental consiste en realizar actualizaciones muy frecuentes, pero pequeñas, de ese modo las organizaciones innovan con mayor rapidez para sus clientes[10]; a continuación se proporciona una vista general de las prácticas de DevOps más importantes[11]: Integración Continua, Entrega Continua (CD), Implementación automatizada, Supervisión continua, Control de versiones y comunicación y colaboración.

4.1 Consideraciones y desafíos de incorporar DevOps en el Estado

Implementar una metodología de DevOps con éxito en un área de desarrollo de una organización estatal, no es simplemente crear un equipo dedicado, es asegurarse de que los miembros del área de desarrollo de la organización promuevan las prácticas, la cultura y las metodologías de DevOps[12].

Uno de los principales problemas que nos encontramos es que en estas organizaciones no se puede cambiar el personal o incorporar a discreción, es decir, los cambios se deben producir con casi todo el personal que ya se encuentra trabajando. Por lo cual los miembros del equipo deben capacitarse con las habilidades necesarias para identificar y ejecutar todas las tareas inherentes a su responsabilidad.

Un desafío que deben afrontar estas áreas que desean incorporar estas prácticas es definir equipos con la capacidad no sólo de dirigir y organizarse para alcanzar sus objetivos, sino también de adaptarse para corregir y mejorar su propio desempeño.

El desafío está en organizar equipos de trabajo auto-organizados y multifuncionales (cross-functional), los equipos auto-organizados están formados por miembros que trabajan juntos y cooperan cada día para entregar una funcionalidad, producto o servicio. Una característica importante para estos equipos es que se les ha de delegar autoridad para la toma de decisiones[13].

Los equipos auto-organizados que funcionan responden a una cultura organizativa clara, que apuesta por el talento de sus profesionales en cada uno de sus niveles de responsabilidad. Para lo cual uno de los principales aspectos que una organización estatal debe considerar, al menos en sus áreas de IT, es la del cambio cultural y organizacional de la misma. Los equipos cross-functional o multifuncionales están formados por miembros de un mismo nivel jerárquico o expertos de diferentes áreas de trabajo (ej. development, operaciones, testing, infraestructura, etc.) que se reúnen para llevar a cabo una determinada tarea, permitiendo una alta cohesión, a efectos de optimizar esfuerzos para lograr los objetivos[14].

Otro aspecto a considerar es el tamaño de los equipos al implementar una metodología DevOps, siguiendo las concepciones de los equipos auto-organizados y multifuncionales; un equipo de trabajo debería contar con no más de 7 a 11 miembros; por debajo de 7 cualquier imprevisto sobre un miembro puede comprometer la previsión de objetivos y por encima de 11 la comunicación y colaboración real entre los miembros se hace más difícil y se forman subgrupos. Además el equipo debe ser estable y sus miembros deben cambiar lo mínimo posible.

Como se mencionó anteriormente, en las organizaciones estatales no se puede (generalmente) sacar personal o incorporar libremente. En este sentido, un aspecto a tener en cuenta es la formación del equipo de desarrollo, obtener un conocimiento acabado de las capacidades de los miembros del equipo, además de estudiar y analizar los programas de capacitación continua en esas áreas resulta esencial y es aspecto relevante a la hora de armar un equipo ajustado a métodos y permitirán obtener productos de calidad en tiempo adecuado, con los recursos disponibles y bajo presupuestos establecidos.

4.2 Aspectos a considerar a la hora del cambio

La adopción de DevOps en una organización, o en un área de IT, puede ser extremadamente compleja. Cambiar la cultura de una organización y normalizar los procesos y las herramientas requiere paciencia y persistencia.

La creación de nuevas metodologías de trabajo e interacción requieren cambios de actitud en los participantes. La falta de colaboración resulta otro aspecto a tener muy presente en las organizaciones a veces también cometen el error de seleccionar personal basándose únicamente en sus habilidades técnicas, en lugar de su capacidad para colaborar[5]. Y por supuesto, no implementar adecuadamente las prácticas de la metodología es otro factor importante de fracaso.

Por otro lado es importante tener bien en claro hacia dónde se quiere ir, es decir se puede adoptar por crear un nuevo departamento de DevOps dentro de la organización para todo el flujo de trabajo, o comenzar implementado la metodología en un área a los efectos de apoyar al departamento de desarrollo, también es un buen comienzo. A su vez se deberá considerar qué tipo de productos o servicios son lo que genera el área de IT, principalmente cambiar la cultura actual y pensar en equipo auto-organizados y multifuncionales que respondan a las metodologías vistas. Las empresas u organizaciones que tienen el propósito de introducir en sus mecanismos prácticas y herramientas DevOps deben considerar dos componentes clave: comunicación y colaboración.

De acuerdo a las necesidades que el área de IT o la organización tengan, se deben definir los primeros pasos a realizar: identificar el equipo con el cual se trabajará, identificar métricas y objetivos, implementación y capacitación, para finalmente realizar un análisis del resultado de la implementación[8].

Algunos interrogantes que debe hacerse la organización son: ¿Los equipos dentro del entorno se comunican eficazmente entre sí? ¿Pueden colaborar fácilmente? ¿Existen cuellos de botella en el proceso de desarrollo? ¿Existen brechas que están generando racionamiento de información? Estos son los tipos de preguntas que las organizaciones deben responder y abordar para implementar DevOps en sus áreas de IT.

El primer paso es que la organización determine cómo está implementada su área de IT, es decir, si el departamento de desarrollo está desvinculado del de operaciones o no, y cómo está definida la estructura del área de IT. En general existe un claro desajuste (independientemente que sean dos departamentos separados o no) entre desarrollo y operaciones, con objetivos diferenciados. Generalmente los equipos de desarrollo de software construyen código a un ritmo alto, pero no se sienten responsables

del proceso de despliegue que es realizado por el equipo de operaciones. El resultado es una acumulación de trabajo donde el dpto. de operaciones se ve desbordado por la cantidad de productos a desplegar y, por falta de tiempo, se desconocen las nuevas funcionalidades[11].

Un cambio en la organización cultural del área es uno de los objetivos para que mejore el nivel de rendimiento, utilizar prácticas DevOps de manera de: eliminar las barreras entre los equipos de desarrollo y operaciones, optimizar el value stream, eliminar fricciones humanas, procedimientos manuales, etc., lo que permitirá acelerar el desarrollo, la innovación, la entrega de productos y soluciones de alta calidad[5]. Una vez que el área logre incorporar estas habilidades se podrán incorporar otras más que mejoren sustancialmente la productividad y la eficiencia del área.

5. Cómo organizar el área de IT, la experiencia en la Dirección Provincial de Energía de Tierra del Fuego.

La Dirección Provincial de Energía es un Ente estatal autártico (autonomía presupuestaria) de la provincia de TDF AeIAS que brinda servicio de energía eléctrica a las ciudades de Ushuaia y Tolhuin. La organización dispone de un área de Sistemas encargada de todo lo concerniente a tecnología, desde el desarrollo de aplicaciones, base de datos, back-up, infraestructura, seguridad, redes, etc.

El área de IT estaba integrada por un jefe de departamento (responsable de toda el área) compuesta de dos divisiones (con un responsable y dos empleados cada una), una responsable de desarrollo (análisis, desarrollo, mantenimiento y otra de infraestructura, operaciones y seguridad; las cuales funcionaban de manera separada entre sí.

El proyecto abarca diferentes etapas de un plan trianual (2020-22), el presente trabajo explicita lo abordado en la primera mitad del mismo.

El objetivo principal consistió en poder comenzar a implementar algunas prácticas de DevOps para optimizar el flujo de valor (value stream), mejorar la productividad y los tiempos de respuestas (sobre todo en el desarrollo). Presentado el proyecto a la organización, y con el aval de sus directivos, se procedió a aplicar un nuevo concepto estructural (específico para el área de IT) el cual consistió en establecer que las responsabilidades sean asumidas por los tres jefes (todo se aprobaba por el triunvirato), si bien no se logró cambiar la estructura organizacional (cada uno mantenía su jerarquía), sí desde el punto de vista operativo; se logró de cada uno de los miembros se involucren en el proyecto y comenzaron a trabajar como un solo equipo.

Resuelto uno de los principales problemas (la consolidación del área como equipo), se procedió considerar otros aspectos para el éxito del proyecto; se procedió a identificar cada una de las habilidades de sus miembros, de manera de saber específicamente cuál sería su mejor aporte, que habilidades no se estaban cubiertas y, por otro lado, las necesidades de capacitación.

El resultado fue fundamental para definir las funciones de cada uno en torno a las nuevas prácticas y conocer las debilidades que se podrían enfrentar, a modo de ejemplo no existía nadie con los conocimientos necesarios para realizar testing. En base a eso se determinó el papel de cada empleado dentro del equipo, más la contratación de un profesional externo para colaborar con tareas específicas (en particular testing), con lo cual se estableció un equipo de trabajo de 8 miembros.

El segundo paso fue establecer un calendario (tentativo) de capacitación para todos los miembros en diversos aspectos según las necesidades (a cumplir en los tres años del proyecto), el objetivo final es alcanzar un full stack teams, el cual consiste en un equipo equilibrado que tiene varios miembros con habilidades combinadas, para diseñar, construir, implementar y operar a lo largo de todos los ciclos de desarrollo. Habilidades como:

- Codificar en varios lenguajes.
- Manejar diversos entornos (cloud, web, linux, u otros).
- Administrar una base de datos.
- Conocer de infraestructura.
- implementar varias herramientas de automatización de pruebas
- Conocimientos de testing.
- Front-end, back-end
- Gestión de proyectos,
- Administrador de versiones de desarrollo
- Habilidades de comunicación.

Otro aspecto fundamental para alcanzar las ventajas de implementar DevOps fue alinear los objetivos del área de IT con los objetivos de la organización; el cambio radical fue aplicar un nuevo paradigma de trabajo, en lugar de trabajar en proyectos y tareas asignadas en forma individual se pasó a una forma de trabajo orientada a objetivos, en que los miembros se centren en los objetivos del negocio; consolidando el trabajo colaborativo entre los miembros y obtener un propósito en su trabajo diario.

Concluida la primera etapa del proyecto, se determinó cuáles serían las herramientas adecuadas de DevOps a implementar en el área. En virtud de estar recién comenzado esta nueva estrategia se consideró aplicar algunas las prácticas de DevOps y, posteriormente incorporar otras con el tiempo. En función de las habilidades ya adquiridas en el equipo y la posibilidad de éxito, se decidió, dentro de las prácticas más importantes, comenzar a implementar las siguientes:

Integración Continua: Que el equipo desarrolle software en forma continua, es decir, con una cierta frecuencia o veces al día; cada integración se debe verificar mediante una compilación automatizada para detectar errores de integración lo más rápido posible.

Realización de pruebas automatizadas: Incorporar pruebas de código automatizadas y realizarlas a medida que el código se está creando o actualizando.

Control de versiones: El equipo deberá administrar todo el código por versiones, que permita realizar un seguimiento de las revisiones y del historial de cambios.

Monitoreo y registro: Llevar un tablero de control que permita medir el desempeño de las prácticas adoptadas.

Comunicación y colaboración: Esencial para el cambio propuesto de trabajo, los miembros deben adoptar mecanismos que permitan establecer la cooperación de manera de realizar con éxito las prácticas mencionadas y alineadas al trabajo por objetivos, estipulado como meta central.

Partiendo de la premisa que implementar DevOps no es un objetivo final, sino por el contrario un proceso continuo; una de las últimas actividades para conocer la efectividad del área, fue determinar que métricas serían las eficaces a utilizar para medir aspectos esenciales; en este caso de comenzó utilizando cuatro: frecuencia de despliegue, disponibilidad del servicio, tiempo de espera y tasa de errores. Los pasos descritos anteriormente no son de ninguna manera la única forma; las organizaciones deberán elegir los pasos que mejor les funcione y adapten a su entorno.

6. Resultados

Si bien el proyecto se encuentra en desarrollo, lo implementado a la fecha permite reflexionar sobre su alcance y obtener resultados parciales de los logros alcanzados.

En primer lugar se logró el compromiso de aceptar la innovación propuesta, que en el ámbito público a veces cuesta mucho más, lo que permitió por un lado romper la jerarquía tradicional e implementar un equipo de trabajo bajo la concepción de la metodología DevOps; bajo este aspecto se pudieron identificar las fortalezas y debilidades como equipos, es decir que habilidades existían y cuáles no, con el apoyo de la gerencia se estableció un plan de capacitación trianual para los miembro en aquellos tópicos más débiles.

Se logró un cambio profundo en la manera de trabajar, el trabajo orientado a objetivos, afianzo al equipo tanto en la eficiencia (la misma fue gradual y se evidenció más sobre el final) como así también sobre nuevas estrategias de comunicación adoptadas que facilitó la colaboración global del equipos; en este punto también el logro fue paulatino y gradual, los primeros meses llevo una alta carga de trabajo poder implementarlo (por los vicios de tanto tiempo trabajando de otra manera) y con el esfuerzo de todos se fue revirtiendo.

En lo referido a las prácticas y herramientas adoptadas de DevOps, se seleccionaron aquellas que pudieran implementarse para la primera etapa del proyecto, desde el momento de su puesta en práctica se siguen aplicando de acuerdo a lo planificado. Al no implementar muchas de ellas, se generan algunos cuellos de botellas que el equipo debe ir solucionando y evitando que se generen problemas; uno de los riesgo identificados en el proyecto era la convivencia de las nuevas prácticas con algunas viejas durante la transición, si bien hubo que realizar varios ajustes se logró mantener las prácticas adoptadas y se espera que se afiancen aún más para el final del proyecto.

Usando las métricas definidas podemos comparar resultados en forma cuantitativa de algunos parámetros, entre la forma tradicional de trabajo y al utilizar DevOps. La

figura 2 muestra la relación de los porcentajes de la tasa de errores y la disponibilidad de servicio, mientras que la tabla 1 refleja el resultado de otras métricas utilizadas.

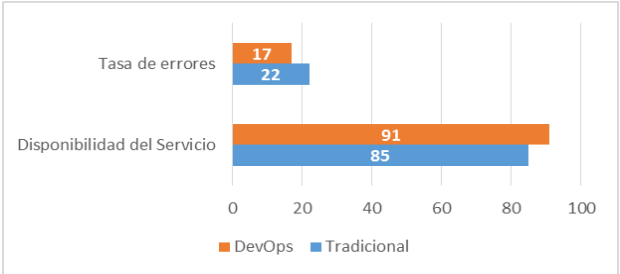


Figura 2. Comparación de métricas entre el equipo DevOps y tradicional

Tabla 1. Diferencias entre la organización tradicional y DevOps

D.P.E.	**Tradicional**	**DevOps**
Frecuencia de Despliegue	2/3 x semana	5 x semana
Tiempo de espera	3 / 4 días	1.5 / 2 días

7 Conclusiones

El éxito de innovar e implementar nuevas metodologías de desarrollo en el ámbito público está en el convencimiento de la gerencia y en el acompañamiento del equipo de trabajo actual que se desempeña en el área de IT. Es fundamental que los nuevos paradigmas en las áreas de IT deben diseñarse con equipos motivados y al servicio de los principales procesos productivos. Modificar las actuales estructuras jerárquicas rígidas por canales horizontales y transversales de comunicación, facilita que todos puedan ejercer las diferentes funciones y sentirse parte de los logros.

La experiencia en la Dirección Provincial de Energía nos demuestra que adaptar una estructura de trabajo tradicional hacia metodología DevOps es totalmente factible, lo que no quiere decir que sea fácil o que se logre de un día para el otro. Independientemente de todas las dificultades que tiene cualquier cambio organizacional en una empresa o área. Las ventajas que implica este paso son considerables pero al mismo tiempo todo un desafío para la organización, sus directivos y para cada miembro.

Identificar las habilidades del equipo de trabajo, junto a sus fortalezas y debilidades, resultará esencial a la hora de determinar que prácticas de la metodología DevOps podrán inicialmente poder en funcionamiento y cuando y de qué manera ir incluyendo las demás. Comenzar con algunas y familiarizarse con ellas es fundamental.

Por otro lado es muy importante capacitar a los miembros del área para que puedan incorporar las habilidades descritas, y si hace falta incorporar alguna persona más siempre tener presente que es fundamental buscar aquellas que tengan mucho interés

en aprender y colaborar, eso logrará que el área pueda implementar las prácticas DevOps con mayor facilidad.

DevOps no es un objetivo final de la organización, sino un proceso continuo que permita mejorar la eficiencia, la productividad y el rendimiento. Incentivar la colaboración y participación, así como la capacitación continua de los miembros del equipo será parte del éxito. Llevar un área tradicional a una metodología DevOps es posible, solo hay que proponérselo y trabajar en pos de ello.

Es importante evidenciar que lo expuesto es un caso concreto particular y que los pasos descritos no son de ninguna manera la única forma; las organizaciones deberán elegir los pasos que mejor les funcione y adapten a su entorno.

Si bien falta mucho por implementar en la DPE, se puede decir que el camino comenzado va dando los resultados esperados y es motivo de seguir avanzando.

Referencias

1. Jabbari R., Ali N., Petersen K. y Tanveer B.; *"What is DevOps? A systematic mapping study on definitions and practices"*; ACM International Conference Proceeding Series; (2016).
2. Paez N.; *Versioning Strategy for DevOps Implementations;* Congreso Argentino de Ciencias de la Informática y Desarrollos de Investigación (CACIDI); (2018)
3. Bass L., Weber I., Zhu L.; *DevOps A Software Architect's Perspective; Addison Wesley*; ISBN-13: 978-0-13-404984-7; (2015).
4. Villamarín E.; *Introducción a DevOps para la mejora de los procesos de desarrollo con herramientas Open Source*; (2019).
5. *Full Stack Teams,Not Engineers*; The DevOps Enterprise Forum; IT REVOLUTION; (2019)
6. The Journey to Positive Business Outcomes DevOps; IT REVOLUTION; (2016)
7. Jiménez Marco G.; *DevOps, la nueva tendencia en el desarrollo de sistemas TI, un caso práctico*; Universidad politécnica de Catalunya, (2016)
8. Forsgren N., Humble J., Kim G., *Accelerate The Science Of Lean Software And Devops Building and Scaling High Performing*; ISBN: 978-1942788331; (2018)
9. Minsal D. y Pérez Rodríguez, *Organización funcional, matricial… En busca de una estructura adecuada para la organización*; ACIMED; ISSN 524-9435; La Habana; (2007).
10. https://aws.amazon.com/es/devops/what-is-devops/
11. Remi J., Sangeetha M.; *From Dev to Ops – Introduction to Devops on understanding Continuous Integration and Continuous Delivery*; International Journal of Innovative Research in Computer and Communication Engineering ISSN(Online): 2320-9801; (2016)
12. https://www.mulesoft.com/resources/api/devops-team-structure
13. Baia W., Fengb Y.; *Organizational Structure, Cross-Functional Integration and Performance of New Product Development Team*; 13th Global Congress on Manufacturing and Management; ScienceDirect; (2016)
14. Floortje Blindenbach-Driessen; *The (In)Effectiveness of Cross-Functional Innovation Teams: The Moderating Role of Organizational Context*; IEEE Transactions On Engineering Management, VOL. 62; (2015)
15. T. Dingsøyr, T. E. Fægri, Tore Dybå; *Team Performance in Software Development*; IEEE SOFTWARE; (2016)
16. Wiedemann A., Neu-Ulm *Are You Ready For Devops? Required Skill Set For Devops Teams*; ECIS; (2018)

ISBN 978–987–633–574–4

Aprovechamiento de las características de las Aplicaciones Web Progresivas en las Redes Sociales

Rocío Rodríguez, Pablo Vera, Claudia Alderete, Mariano Dogliotti

Centro de Altos Estudios en Tecnología Informática (CAETI)
Facultad de Tecnología Informática
Universidad Abierta Interamericana (UAI)
Avenida Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina
{RocioAndrea.Rodriguez, PabloMartin.Vera, ClaudiaGabriela.Alderete,
MarianoGaston.Dogliotti }@uai.edu.ar

Resumen. Gran parte de la población mundial ha encontrado en las redes sociales un lugar para comunicarse y difundir sus pensamientos e ideas. El acceso a las redes sociales se produce mayormente desde dispositivos móviles, los cuales poseen distintas características, sistemas operativos y capacidades. El acceso debe ser asegurado desde distintos dispositivos móviles así como desde computadoras. El desarrollar una aplicación particular para cada sistema operativo (considerando distintos versionados) y contemplando las diversas características de los mismos, resulta cada vez más complejo. Las PWA (aplicaciones web progresivas) permiten simplificar y unificar el desarrollo, con la portabilidad propia de la web, agregregando características propias de las aplicaciones nativas, donde para el usuario final es indistinto si se trata de una aplicación PWA ó nativa. Este trabajo presenta un relevamiento de las princiaples redes sociales, analizando cuales de ellas están construidas mediante el principio de PWA y además analiza ciertas características para detectar si realmente están bien configuradas y cumplen con los lineamientos básicos y buenas prácticas de las PWA.

Palabras Clave: PWA, Redes Sociales, Service Worker, Manifiesto

1 Introducción

Las personas somos seres sociables y eso se ha reflejado a través del tiempo en el uso de distintos medios sociales en internet, “la extraordinaria capacidad de comunicación y de poner en contacto a las personas que tienen las redes ha provocado que un gran número de personas las esté utilizando con fines muy distintos” [1]. Pero la necesidad de estar conectados cobró mayor significancia en tiempos de pandemia. Y esto se vio reflejado en las redes sociales. “… Los medios sociales se han convertido en una parte indispensable de la vida cotidiana para todas las personas alrededor del mundo, el 2020 un año en el que el mundo cayó en un bloqueo, los usuarios de redes sociales crecieron a una tasa más grande que en 3 años…” [2]. "1,3 millones de nuevos usuarios se unieron a las redes sociales cada día durante 2020: 15 nuevos usuarios

cada segundo"[3]. “Esto significa que, por primera vez, más de la mitad de la población mundial ahora usa las redes sociales … [4].

“Al avanzar la tecnología móvil poco a poco se ha ido convirtiendo en el centro de la actividad online. Hoy en día, desde prácticamente cualquier dispositivo móvil podemos acceder a Internet, ya sea bien a través de navegadores o aplicaciones instaladas” [5]. Si se toma como ejemplo la red social Facebook en Argentina, el 97,5% de los accesos se realiza desde algún dispositivo móvil, mientras que un 2,5% lo hace desde notebook o computadora de escritorio, evaluando por cada usuario esos accesos se puede determinar que el 73,5% ingresa únicamente desde un dispositivo móvil (estadísticas publicadas en [6], en Enero del 2021). Esto pone en evidencia la necesidad de contar con soluciones que sean aptas para un sinfín de dispositivos móviles con características diferentes, así como en notebook o computadoras de escritorio. En este aspecto las aplicaciones web progresivas (PWA) pueden ofrecer una buena solución.

Este artículo se encuentra estructurado de la siguiente manera: en la sección 2 se definen las Aplicaciones Web Progresivas (PWA) y se presentan sus componentes; en la sección 3 se presentan las características a analizar en una PWA; en la sección 4 se listan las redes sociales que serán consideradas para someterlas a dicho análisis, en la sección 5 se presentan los resultados obtenidos y finalmente en la sección 6 las conclusiones.

2 PWA

Las aplicaciones web progresivas (PWA) son una evolución de las aplicaciones web, que consideran las bases de las aplicaciones web adaptativas [7], incorporando la apariencia junto con algunas funcionalidades que antes eran exclusivas de las aplicaciones nativas. “Las aplicaciones web progresivas son una evolución natural de las aplicaciones web que difuminan la barrera entre la web y las aplicaciones, pudiendo realizar tareas que generalmente solo las aplicaciones nativas podían llevar a cabo. Algunos ejemplos son las notificaciones, el funcionamiento sin conexión a Internet o la posibilidad de probar una versión más ligera antes de bajarte una aplicación nativa de verdad” [8]. “La PWA se basa en los conceptos de una sola aplicación para todas las plataformas al igual que el enfoque híbrido. Sin embargo, posee distintas capacidades, como carga instantánea, notificaciones push incluso en estado fuera de línea” [9].

Los componentes principales de una PWA son: (1) Archivo de Manifiesto, (2) Service Worker, (3) Almacenamiento Local, (4) Notificaciones

Toda PWA tendrá como punto de inicio un archivo de manifiesto en el que se definen las configuraciones básicas de la aplicación y será imprescindible contar con un service worker. “Un service worker es un proceso que el navegador web ejecuta en segundo plano y está asociado a un sitio web particular. Este proceso se programa en javascript y permite capturar las peticiones que el sitio web hace a la red e interceptarlas actuando como un proxy local. El capturar esas peticiones permite que el service worker responda en lugar de la red, haciendo posible que el navegador no salga a la red sino que se le devuelvan los datos localmente” [10]. Es decir, brinda la

posibilidad de recuperar los datos sin necesidad de la red y esto se logra gracias a métodos de almacenamiento local. Por último, las PWA incorporan la capacidad de mostrar notificaciones al usuario, aun cuando el navegador no está abierto, trabajando de forma similar a una aplicación nativa.

3 Análisis

Se toman en consideración diversas características sobre un conjunto de sitios de redes sociales (los cuales se listan en la sección siguiente), el análisis comenzará por la evaluación del archivo de manifiesto (la ausencia de este archivo implica que el sitio analizado no es PWA). Dentro del archivo de manifiesto se visualizan algunas propiedades básicas si están indicadas. Luego se examina el tamaño medido en Bytes que ocupa la solución, primeramente, en cuanto a consumo de memoria, del mismo modo se analiza el serviceworker (toda aplicación PWA posee un service worker) y además se analiza si de forma adicional emplean otros mecanismos de almacenamiento como bases de datos o almacenamiento interno en el dispositivo.

A continuación, se detallan las características consideradas.

3.1 Archivo de Manifiesto y Service Worker

Como fue especificado anteriormente para que una aplicación sea PWA debe tener estos dos elementos básicos (archivos de manifiesto y service worker).

El archivo de manifiesto es un archivo JSON que permite configurar ciertas características de la aplicación. En dicho archivo se analizará la definición de las siguientes propiedades:

- Presentación: Se analiza que estén definidos: (1) Los colores tanto de Tema como de Fondo, (2) El modo de visualización, definido mediante la propiedad Display.
- Iconos: De la aplicación en distintos tamaños, con estilo enmascarable. Debido a que la PWA podrá instalarse en distintos entornos cada uno de ellos tendrá una forma diferente de mostrar los íconos (cuadrados, circulares, cuadrado con borde redondeados…), es por eso que los íconos enmascarables permiten tomar un área desde el centro que será siempre visualizable en cualquier entorno y estos serán mostrados dentro del entorno de forma tal que sean visualizados igual que una aplicación nativa. En la figura 1, se muestra uno de los íconos creados para la PWA de Youtube y como sería visto al enmascararlos en un formato circular.

Fig. 1. A la izquierda Icono de Youtube y a la derecha enmascarado en una forma circular

- Shortcuts: Se pueden definir atajos para el acceso a determinadas funcionalidades (funcionarán cuando la aplicación es instalada)
- Screenshot: Capturas de pantallas que pueden ser incorporadas para tenerlas por defecto

Existen PWA que no tienen precisiones en los parámetros de su archivo manifiesto y es por ello por lo que resulta de interés relevar la completitud de estos.

En cuanto al service worker mediante la herramienta developer tools del navegador Google Chrome es posible visualizar si un sitio determinado tiene un service worker activo, su versión y estado. A modo de ejemplo la figura 2 muestra con esta herramienta la presencia de un service worker al acceder a Twitter.

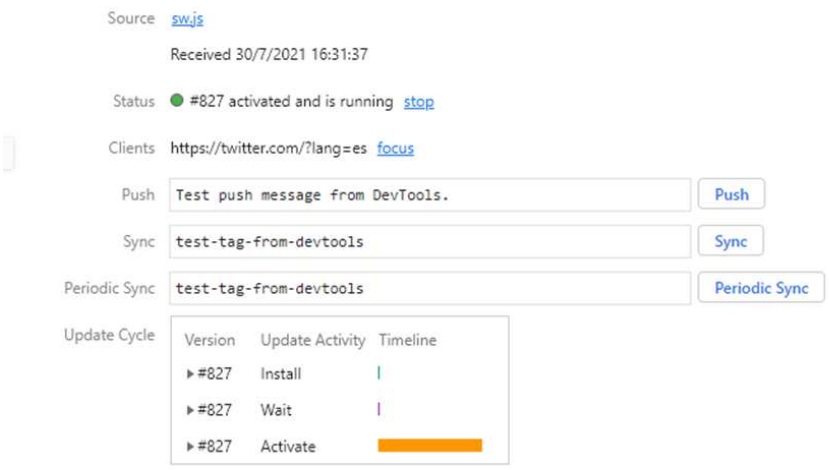


Fig. 2. Existencia de Service Worker – Ejemplo Twitter

3.2 Recursos

- Memoria: Se analiza el tamaño que ocupa en memoria la aplicación apenas el usuario se ha logueado a la misma (medido en MB).

- Service Worker: Se mide el espacio que consume el service worker (medido en KB)

A lo que se suman otros métodos de almacenamiento que pueden utilizar las aplicaciones PWA:

- IndexedDB: Base de datos local (medido en KB)
- Cache Storage: Almacenamiento local en el dispositivo (medidos en KB ó MB)

A modo de ejemplo se presenta el caso de Instagram en donde puede observarse que utiliza tanto IndexedDB como Cache Storage (en la descripción a la derecha en la figura 3) y además se observa el peso del service worker. Cabe aclarar que como todos los pesos están medidos en KB lo ocupado por Cache Storage en proporción es tan ínfimo en comparación de los restantes que no se ve en la gráfica de la izquierda.

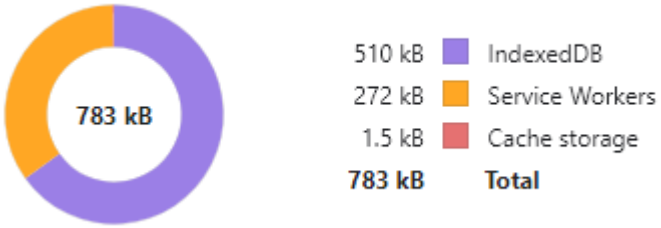


Fig. 3. Recursos – Ejemplo Instagram

3.3 PWA optimizada

Se complementa el análisis con una herramienta de código abierto “Lighthouse” [11], que puede ser ejecutada de diferentes maneras (en nuestro caso lo hemos utilizado como extensión del browser de escritorio) permitiendo a los desarrolladores analizar si la aplicación es una PWA Optimizada y el resultado que arroja es un valor del 1 al 8, según el cumplimiento de determinados parámetros:

1. Registra un service worker que controla las páginas y la url de inicio
2. Utiliza HTTPS
3. Contiene configuración para mostrar una página de inicio personalizada
4. Define un color de tema para la barra de direcciones
5. El contenido esta correctamente dimensionado al área de visualización (viewport)
6. Contiene el tag <meta name="viewport"> configurado con un ancho o escala inicial
7. Provee un apple-touch-icon válido
8. El manifiesto contiene un ícono enmascarable

3.4 Mediciones de Tiempos

"En la actualidad vivimos en tiempos donde todo el mundo parece tener prisa, queremos todo "ya" y no deseamos perder ni tan solo un minuto. Lo mismo sucede cuando navegamos por internet, buscamos información y somos tan exigentes que tiene que ser buena y presentada rápidamente" [12]. Actualmente Google a través de la iniciativa de Web Vitals [13] propone tres métricas que permiten analizar los tiempos de una solución web. Esta iniciativa se centra en tres aspectos de la experiencia del usuario: carga, interactividad y estabilidad visual, e incluye las métricas mostradas en figura 4 se presentan los umbrales de las tres primeras métricas.

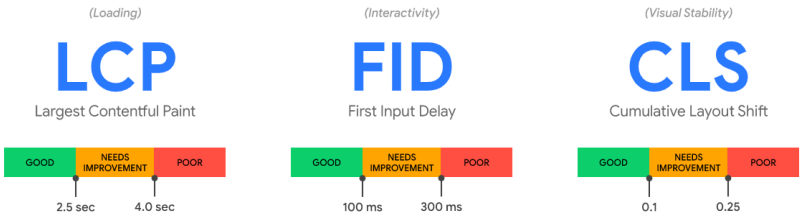


Fig. 4. Métricas consideradas

A continuación, se detallan las tres métricas propuestas:

- LCP (Largest Contentful Paint): Se refiere al tiempo para el despliegue del contenido más extenso, mide el rendimiento de la carga. Se establece que el tiempo debe producirse dentro de los 2,5 segundos desde el comienzo de carga de la página.
- FID (First Input Delay): Se refiere a la demora para la primera entrada, mide la interactividad. Para proporcionar una buena experiencia de usuario las páginas deben tener un FID menor a 100 milisegundos.
- CLS (Cumulative Layout Shift): Se refiere al cambio acumulativo en el diseño, mide la estabilidad visual. Se establece que el índicador, para proporcionar una buena experiencia de usuario, debe ser menor de 0,1.

De manera automática es posible medir el LCP y el CLS. "Las herramientas automáticas que cargan páginas en un entorno simulado sin un usuario no pueden medir FID" [13]. No obstante, muchas de ellas arrojan un valor de FID, pero el mismo no será considerado.

Por lo tanto, en esté análisis se utiliza la herramienta PageSpeed Insights [14], que permite obtener tanto el LCP cómo el CLS.

4. Relevamiento

Se consideró una muestra de diez redes sociales, indicadas en la tabla 1 (ordenadas alfabéticamente). Las tres redes que han sido destacadas (en negrita) no se consideran

PWA por no tener un archivo manifiesto, por lo cual se someten al análisis a las siete restantes.

Tabla 1. Listado de redes sociales consideradas para el relevamiento

Nombre	Objetivo	Año de Lanzamiento
Facebook	**Conectar con Personas**	**2004**
Instagram	Compartir fotografías y videos	2010
Linkedin	**Oportunidades laborales**	**2003**
Pinterest	Crear tableros personalizados con imágenes de interés	2010
Snapchat	**Mensajería con soporte multimedia**	**2011**
TikTok	Compartir Videos	2016
Tinder	Encuentros Online	2011
Twitch	Transmisiones en Vivo	2011
Twitter	Microblogueo	2006
Youtube	Compartir Videos	2005

5. Resultados Obtenidos

Como se mencionó previamente se descartan 3 redes sociales (Facebook, Linkedin y Snapchat) por no ser PWA, el resto de las redes sociales han sido analizadas con el procedimiento descripto previamente.

En base al archivo de manifiesto sobresale Twitter que es la única red social (de las relevadas) que posee la descripción de todas las características analizadas, siendo la única que incorpora shortcuts y screenhots.

En cuanto al uso de memoria todas las PWA ocupan un tamaño desde 15 a 40 MB, el uso de memoria cambia ligeramente en cada ejecución de la aplicación dado que su contenido es dinámico, no obstante, resulta interesante extender el análisis pudiendo observarse que el service worker más pesado lo tiene Instagram que es la única red social (de las analizadas) que utiliza tanto indexedDB como CacheStorage. Todas las redes sociales analizadas utilizan indexedDB, siendo tan sólo 3 las que usan CacheStorage: Instagram, TikTok y Twitch.

En cuanto al puntaje de PWA extraído desde el reporte generado con Lighthouse pude observarse en la figura 5 que las redes sociales Tinder, Twitch y Twitter son las que mejor están adaptadas a PWA cumpliendo con los 8 ítems analizados por la herramienta.

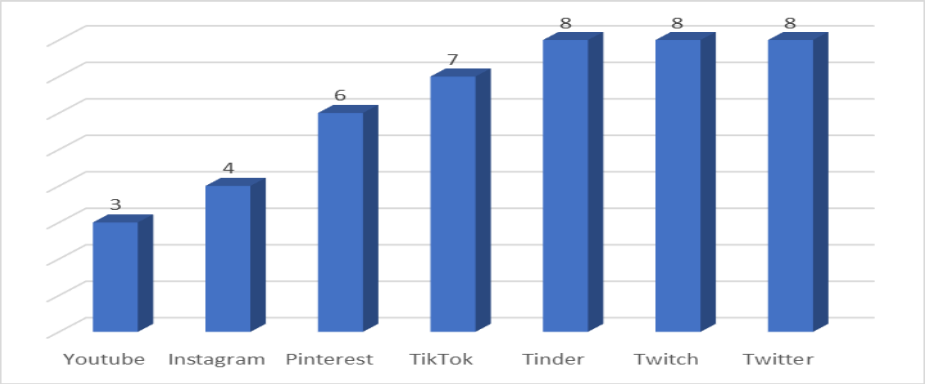


Fig. 5. Puntaje de PWA sobre un total de 8 puntos (obtenidos usando Lighhouse)

Tomando como ejemplo a las redes sociales mayormente asociadas con videos, puede advertirse las diferencias en el desarrollo posicionándose Youtube como la PWA menos optimizada (3 de 8), mientras que TikTok (7/8) está cerca de tener un desarrollo optimizado, sobresaliendo Twitch que se basa principalmente en streaming en vivo y es muy utilizada por radios y otros medios de comunicación, con una PWA optimizada (8 de 8).

En el análisis de tiempos mediante la herramienta PageSpeed Insights es posible observar que las tres redes sociales que estaban más optimizadas como PWA (Tinder, Twitch y Twitter), cumplen con los tiempos de carga tanto de FCP como de CLS. Como era de esperarse Twitch cumple con ambos parámetros mientras que Youtube sólo con 1 (CLS).

6. Conclusiones

Puede notarse como las redes sociales en general (70% de las relevadas) sacaron provecho de las posibilidades de las PWA, quedándose Facebook, Linkedin y Snapchat rezagadas en este aspecto. De las redes sociales desarrolladas como PWA un 43% de las analizadas están optimizadas para tal fin. Sobresale el caso de Twitter que al inspeccionar el archivo manifiesto se evidencia una muy buena definición de todas las características analizadas. Como se mencionó en la sección de resultados, más allá del enfoque de la red social (se compararon tres redes sociales basadas en videos) sus desarrollos pueden ser optimizados alcanzando la mejor calidad de PWA, lo que impactará sin lugar a duda en la reducción en la descarga de datos (utilizando almacenamiento interno en el dispositivo) y usabilidad (mediante los parámetros definidos en el manifiesto). En este sentido se destaca por los resultados obtenidos en el relevamiento a Twitter.

De las redes sociales analizadas la que resultó ser la PWA más efectiva fue Twitter (por su observación de archivo de manifiesto, cumpliendo con los 8 ítems de análisis de PWA y también con los parámetros asociados a los tiempos de carga), seguida por Tinder y Twitch (ambas con un manifiesto incompleto).

Finalmente es importante concluir que si bien la mayor parte de las redes sociales han dedicado sus esfuerzos a la construcción de una PWA muchas de ellas aún les falta optimizar el uso de almacenamiento interno, tener un archivo de manifiesto más detallado y tomar en consideración cuestiones simples de configuración que permitirán tener una PWA optimizada.

Como trabajo futuro se propone analizar el parámetro FID y por otra parte indagar que optimizaciones podrían realizarse en estas PWA para que impacten mejorando en su rendimiento actual.

Referencias

1. De Haro, J. J. Redes sociales en educación. Educar para la comunicación y la cooperación social, 27, 203-216. (2010).
2. Digital 2021. We are social. https://wearesocial.com/digital-2021
3. Digital Report 2021: el informe sobre las tendencias digitales, redes sociales y mobile https://wearesocial.com/es/blog/2021/01/digital-report-2021-el-informe-sobre-las-tendencias-digitales-redes-sociales-y-mobile
4. Digital 2020: July Global Statshot https://datareportal.com/reports/digital-2020-july-global-statshot?rq=wifi
5. Castell Ferreres, G. Desarrollo e implementación de una aplicación web progresiva (PWA) (Bachelor's thesis, Universitat Politècnica de Catalunya). (2020).
6. Branch. Estadísticas de la situación digital de Argentina en el 2020-2021 https://branch.com.co/marketing-digital/estadisticas-de-la-situacion-digital-de-argentina-en-el-2020-2021/
7. Rodríguez, R. A., Vera, P. M., Ramirez, M. A., Alderete, C. G., Conca, A. G., Dogliotti, M. G., & Zain, G. A. Análisis del Diseño Web Adaptativo Caso de estudio: Universidad Argentinas. Revista Abierta de Informática Aplicada (RAIA), 4(1), 51-62. (2020).
8. Ramírez Ivan. ¿Qué es una aplicación web progresiva o PWA? (2018) https://www.xataka.com/basics/que-es-una-aplicacion-web-progresiva-o-pwa
9. Adetunji, O., Ajaegbu, C., Otuneme, N., & Omotosho, O. J. Dawning of Progressive Web Applications (PWA): Edging Out the Pitfalls of Traditional Mobile Development. American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS), 68(1), 85-99. (2020).
10. Vera, P., Rodriguez R., Estrategias de Manejo de Cache para Aplicaciones Web Progresivas - Presentación de un Esquema optimizado. Congreso Nacional de Ingeniería Informática. (2020)
11. Google Developers, Lighthouse https://developers.google.com/web/tools/lighthouse/?utm_source=devtools
12. Echeverria, D. Tiempo de Respuestas y Experiencia de Usuario Estudio Experimental. Revista latinoamericana de ingeniería de software, 4(5), 231-234. (2016).
13. Philips Walton. Métricas esenciales para un sitio saludable. https://web.dev/vitals/
14. Google Developers, PageSpeed Insights https://developers.google.com/speed/pagespeed/insights/

ISBN 978–987–633–574–4

Ontology Metrics in the Context of the GF Framework for OBDA

Sergio Alejandro Gómez[1,2] and Pablo Rubén Fillottrani[1,2]

[1]Laboratorio de I+D en Ingeniería de Software y Sistemas de Información (LISSI)
Departmento de Ciencias e Ingenería en Computación
Universidad Nacional del Sur
San Andrés 4000, (8000) Bahía Blanca, Argentina
Email: {sag,prf}@cs.uns.edu.ar
[2]Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

Abstract. There is a need of generating high-quality ontologies for Semantic Web applications that need to access legacy, relational and non-relational data sources. We present GF, a tool for materialization of ontologies from relational and non-relational data sources such as H2 databases, CSV files and Excel spreadsheets. We evaluate a sample case generated by GF with a third-party ontology evaluation tool called OntoMetrics. The results obtained show that the ontologies generated with GF are reasonably good for being used in Semantic Web applications as they are validated correctly and pass all of the filters for the OWL2 main profiles, thus making them suitable for processing with lightweight reasoners. The metrics indicate that our application is lacking quality in the annotation area regarding the documentation of the classes and properties generated by the application.

Keywords. Ontologies, Ontology-Based Data Access, Metrics, Knowledge Representation

1 Introduction

The Semantic Web (SW) is a vision of the web where data resources have precise meaning given in terms of ontologies making them apt to be processed by computers. An ontology is a logical theory described in a language known as OWL and whose logical assertions are described in the RDF language. Thus OWL/RDF ontologies are considered as knowledgebases identified by an IRI and are composed of concepts/classes and relations/properties among classes and/or classes and values.

Ontology-based data access (OBDA) is concerned with enriching both relational and non-relational, usually legacy, data sources with an ontology describing the model of an application domain (also known as the business logic). One common approach to OBDA is materialization where data sources are translated into ontologies (both schema and instance information) and then enriched with application data. We have developed a framework, called GF, for performing OBDA with legacy datasources based on materialization, that is currently in

prototypical state but to which we are incrementally adding functionality following a problem-solving driven approach to information interoperability. The current implementation of GF allows to deal with H2 databases, CSV files and Excel spreadsheets.

Evaluation of the ontologies produced with our framework is important. The process of ontology evaluation is concerned with ascertain quality and correctness of ontologies. This ultimately allows quantifying the suitability of a given ontology to the purposes for which it has been built. Several metrics have been proposed for ontologies over the last years (see Sect. 3 and related work literature [1–7] for details). A notable third-party project that accrues the most part of the ontology metrics developed is Ontometrics that presents to the final user as web applcation and allows to obtain meaningful metrics from an ontology loaded as a text file.

In this paper, we revisit a case study on solving the construction of a university library ontology from a set of legacy datasources using GF [8]. We then evaluate the obtained ontology with a third-party ontology evaluation tool called OntoMetrics. We thus obtained several descriptors that we have used to debug our implementation and that indicate that in its current state is working correctly but it also needs some improvements. The results obtained show that the ontologies generated with GF are reasonably good for being used in SW applications as they are validated correctly and pass all of the filters for the OWL2 DL, OWL2 EL, OWL2 QL and OWL2 QL profiles, thus making them suitable for processing with lightweight reasoners. The metrics indicate that our application is lacking quality in the annotation area regarding the documentation of the classes and properties generated by the application.

The rest of the article is structured as follows. In Sect. 2, we review the basic functionality provided by the GF framework application. In Sect. 3, we review some of the most important approaches to metrics used for ontologies. In Sect. 4, we present the experiments that we performed to measure ontology metrics on ontologies produced by our application and the properties emerging from the cases that we observed. Finally, in Sect. 5, we present our conclusions and foresee future work.

2 Ontology-Based Data Access in GF

The GF framework allows to materialize OWL/RDF ontologies from heterogeneous, legacy, both relational and non-relational data sources. We explained its functionality in previous work (see [8] and references therein). Assume a relational database (RDB) as in Fig. 1, a typical workflow with GF to produce a working ontology as in Fig.(2.a) from such database along with Excel and CSV files would typically be: (1) Create a new ontology with IRI http://foo.org/ and file name OntoLibrary1.owl. (2) Establish a connection to the RDB Users-Theses-Loans.db. (3) Materialize an initial ontology with the DB contents. (4) Build intermediate classes Material, Printed, with their respective attributes. (5) Establish that Thesis and Printed are subclasses of Material. (6) Modify schema

expressing that Loan is now related to Material instead of Thesis by changing the range of property id and http://foo.org/Loan/ref-id from, select ref-id as object property name and *edit object property*), selecting http://foo.org/Material as new class range. (7) Create class PostgradThesis as a subclass of Thesis. (8) Create classes StudentUser, TeacherUser, GraduateThesis, MScThesis, PhDThesis with the SQL filters by introducing them directly or building them visually:

- StudentUser: SELECT "User"."userID" FROM "User" WHERE "User"."type"='S'
- TeacherUser: SELECT "User"."userID" FROM "User" WHERE "User"."type"='T'
- GraduateThesis: SELECT "Thesis"."id" FROM "Thesis" WHERE "Thesis"."type"='S'
- MScThesis: SELECT "Thesis"."id" FROM "Thesis" WHERE "Thesis"."type"='M'
- PhDThesis: SELECT "Thesis"."id" FROM "Thesis" WHERE "Thesis"."type"='D'

This also establishes that StudentUser and TeacherUser are subclasses of User, GraduateThesis is a subclass of Thesis, and both MScThesis and PhDThesis are subclasses of Thesis (which is made automatically by the system and we will later see that this option has to be actually optional). (9) Establish that PhDThesis and MScThesis are subclasses of PostgradThesis. (10) Establish disjoint classes MScThesis and PhDThesis. (11) Create class Magazine using Excel schema file magazines.xsc and Excel file Magazines.xlsx. (12) Establish Magazine as subclass of Printed. (13) Create class Book using CSV schema file books.sch and CSV file books.sch. (14) Establish Book as subclass of Printed. (15) Finally, save the ontology.

User(*userID*, *name*, *email*, *type*)
Thesis(*id*, *author*, *title*, *pubDate*, *type*, *institution*, *supervisor*)
Loan(*userID*, *id*, *date*, *timeDays*)

User

userID	*name*	*email*	*type*
1	John	john@nosite.com	S
2	Peter	peter@nosite.com	T

Loan

userNo	*id*	*date*	*timeDays*
1	1	2020-09-01	40

Thesis

id	*author*	*title*	*pubDate*	*type*	*institution*	*supervisor*
1	Marie Curie	Recherches sur les substances radioactives	1903-01-01	D	Faculte des Sciences de Paris	Gabriel Lippmann
2	Claude Shannon	A Symbolic Analysis of Relay and Switching Circuits	1937-01-01	M	Massachusetts Institute of Technology	Vannevar Bush

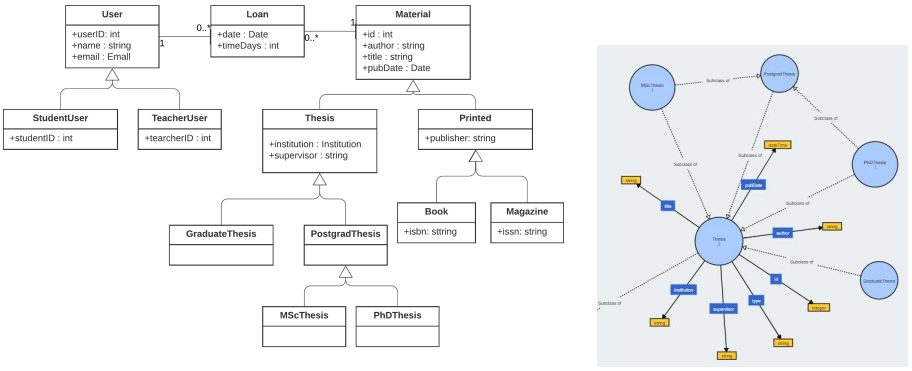
Fig. 1. Relational instance of the library's database concerning Users, Theses and Loans

The validation process of the ontology obtained[1] determined that it complies with OWL 2 DL, OWL 2 EL, OWL 2 QL, and OWL 2 RL profiles. But a visualization of the ontology[2] shows that there are redundant *is-a* relations (e.g. PhDThesis is subclass of both PostgradThesis and Thesis, where the latter is redundant as it does not need to be explicitly expressed and could be determined by an OWL reasoner). See Fig. (2.b).

3 Measuring Ontologies with *OntoMetrics*

Accountability is an integral part to the success of any endeavor. Accountability requires metrics that truly reflect the desired outcomes of a program. Met-

[1] An online validator can be found at http://visualdataweb.de/validator/
[2] OWL/RDF ontologies can be visualized with http://www.visualdataweb.de/webvowl/#



(a) (b)

Fig. 2. (a) Ontology for the university library and (b) Part of the ontology generated with GF

rics are measures of quantitative assessment commonly used for assessing, comparing, and tracking performance or production. *OntoMetrics*[3] is a third-party web-based tool that validates and displays statistics about a given ontology. Ontometrics implements several measurement frameworks such as [7, 9]. We now give an overview of the metrics computed by Ontometrics that we used on the experiments reported in Sect. 4. For further details, we refer the reader to the Ontometrics web page.

Base Metrics comprise of simple metrics, like the counting of classes, axioms, objects etc. These metrics show the quantity of ontology elements. *Class axioms metrics* count the number of subclass, equivalent classes and disjoint classes relations. *GCI* counts the number of the General Concept Inclusion (GCI). *HiddenGCI* counts the number of hidden GCIs in an ontology imports closure. A GCI is regarded to be a hidden GCI if it is essentially introduce via an equivalent class axiom and a subclass axioms where the LHS of the subclass axiom is named. For example, *A equivalentTo p some C*, *A subClassOf B* results in a hidden GCI. *Object property axiom metrics* and *Data property axioms metrics* quantify the presence of object properties and data properties, resp. *Individual Axioms metrics* measure are axioms concerning individuals in the extension of classes and properties.

Annotations can be used to associate information to ontologies, this information could be the version of the ontology or the creator. The annotation itself consists of an annotation property and an annotation value. *Annotation axiom metrics* count the number of annotation axioms in the given ontology.

Schema metrics address the design of the ontology. Although it is not possible to tell if the ontology design correctly models the domain knowledge, metrics in this category indicate the richness, width, depth, and inheritance of an ontology schema design. The most significant metrics in this category are described

[3] See https://ontometrics.informatik.uni-rostock.de/ontologymetrics/

next. The number of attributes (slots) that are defined for each class can indicate both the quality of ontology design and the amount of information pertaining to instance data. In general, it is assumed that the more slots that are defined, the more knowledge the ontology conveys. The *attribute richness* is defined as the average number of attributes (slots) per class, it is computed as the number attributes for all classes divided by the number of classes. The *inheritance richness* measure describes the distribution of information across different levels of the ontology's inheritance tree or the fan-out of parent classes. This is a good indication of how well knowledge is grouped into different categories and subcategories in the ontology. This measure can distinguish a horizontal ontology (where classes have a large number of direct subclasses) from a vertical ontology (where classes have a small number of direct subclasses). An ontology with a low inheritance richness would be of a deep (or vertical) ontology, which indicates that the ontology covers a specific domain in a detailed manner, while an ontology with a high IR would be a shallow (or horizontal) ontology, which indicates that the ontology represents a wide range of general knowledge with a low level of detail. The *relationship richness* metric reflects the diversity of the types of relations in the ontology. An ontology that contains only inheritance relationships usually conveys less information than an ontology that contains a diverse set of relationships. The relationship richness is represented as the percentage of the (non-inheritance) relationships between classes compared to all of the possible connections that can include inheritance and non-inheritance relationships. The relationship richness of a schema is defined as the ratio of the number of (non-inheritance) relationships (P), divided by the total number of relationships defined in the schema (the sum of the number of inheritance relationships (H) and non-inheritance relationships (P)). The following relationships are being counted as non-inherited relationships: Object Properties, Equivalent Classes, Disjoint Classes. The subclasses are being handled as inheritance relationships [9, 7]. The *Attribute-Class Ratio* metric represents the relation between the classes containing attributes and all classes. The difference to attribute richness is that not the amount of attributes is counted. It is only counted whether a class has attributes or not. The *Equivalence Ratio* calculates the ratio between similar classes and all classes in the ontology. The *Axiom Class Ratio* metric describes the ratio between axioms and classes. It is calculated as the average amount of axioms per class. The *Inverse Relations Ratio* metric describes the ratio between the inverse relations and all relations. The *Class Relation Ratio* describes the ratio between the classes and the relations in the ontology [9, 7].

The way data is placed within an ontology is also a very important measure of ontology quality because it can indicate the effectiveness of the ontology design and the amount of real-world knowledge represented by the ontology. *Instance metrics* include metrics that describe the knowledgebase as a whole, and metrics that describe the way each schema class is being utilized in the knowledgebase. These are collectivelly known as *Knowledgebase Metrics*. The *Average Population* (i.e. the average distribution of instances across all classes) measure is an indication of the number of instances compared to the number of classes. It can

be useful if the ontology developer is not sure if enough instances were extracted compared to the number of classes. Formally, the average population (AP) of classes in a knowledgebase is defined as the number of instances of the knowledgebase (I) divided by the number of classes defined in the ontology schema (C). The result will be a real number that shows how well is the data extraction process that was performed to populate the knowledgebase. For example, if the average number of instances per class is low, when read in conjunction with the previous metric, this number would indicate that the instances extracted into the knowledgebase might be insufficient to represent all of the knowledge in the schema. Keep in mind that some of the schema classes might have a very low number or a very high number by the nature of what it is representing. The *Class Richness* metric is related to how instances are distributed across classes. The number of classes that have instances in the knowledgebase is compared with the total number of classes, giving a general idea of how well the knowledgebase utilizes the knowledge modeled by the schema classes. Thus, if the knowledgebase has a very low Class Richness, then the knowledgebase does not have data that exemplifies all the class knowledge that exists in the schema. On the other hand, a knowledgebase that has a very high class richness would indicate that the data in the knowledgebase represents most of the knowledge in the schema. The class richness (CR) of a knowledgebase is defined as the percentage of the number of non-empty classes (classes with instances) (C') divided by the total number of classes (C) defined in the ontology schema.

Class Metrics examine the classes and relationships of ontologies. The *Class Connectivity* metric is intended to give an indication of what classes are central in the ontology based on the instance relationship graph (where nodes represent instances and edges represent the relationships between them). This measure works in tandem with the importance metric to create a better understanding of how focal some classes function. This measure can be used to understand the nature of the ontology by indicating which classes play a central role compared to other classes. The *connectivity of a class* is defined as the total number of relationships that instances of the class have with instances of other classes. The *Class Importance* metric calculates the percentage of instances that belong to classes at the inheritance subtree rooted at the current class with respect to the total number of instances. This metric is important in that it will help in identifying which areas of the schema are in focus when the instances are added to the knowledgebase. Although this measure does not consider the domain characteristics, it can still be used to give an idea on what parts of the ontology are considered focal and what parts are on the edges. The *importance of a class* is defined as the percentage of the number of instances that belong to the inheritance subtree rooted at in the knowledgebase compared to the total number of class instances in the knowledgebase.

The *Class Inheritance Richness* measure details the schema IR metric mentioned in schema metrics and describes the distribution of information in the current class subtree per class. This measure is a good indication of how well knowledge is grouped into different categories and subcategories under this class.

Formally, the inheritance richness (IRc) of class C_i is defined as the average number of subclasses per class in the subtree. The number of subclasses for a class C_i is defined as $|H^C(C_1, C_i)|$ and the number of nodes in the subtree is $|C'|$. The result of the formula will be a real number representing the average number of classes per schema level. The interpretation of the results of this metric depends highly on the nature of the ontology. Classes in an ontology that represents a very specific domain will have low IRC values, while classes in an ontology that represents a wide domain will usually have higher IRC values. The *Class Readability* metric indicates the existence of human readable descriptions in the ontology, such as comments, labels, or captions. This metric can be a good indication if the ontology is going to be queried and the results listed to users. Formally, the readability of a class is defined as the sum of the number of attributes that are comments and the number of attributes that are labels the class has. The result of the formula will be an integer representing the availability of human-readable information for the instances of the current class. The *Class Relationship Richness* is an important metric reflecting how much of the relationships defined for the class in the schema are actually being used at the instances level. This is another good indication of the utilization of the knowledge modeled in the schema. The *Class children* is a count-metric that measures the number of immediate descendants of a given class, also known as a number of children [9]. The *Class instances* metric displays the number of instances of a given class. *Class properties* metrics summarize the properties of an given class [7].

Graph or structural metrics calculate the structure of ontologies. *Cardinality* is a property of graphs which expresses a graph related number of specific elements. *Absolute root cardinality* is a property of a directed graph which represents the number of root nodes of the graph. *Absolute root cardinality* is a property of a directed graph which is related to leaf node sets and represents the number of leaf nodes of the graph. *Absolute sibling cardinality* is a property of a directed graph which is related to sibling node sets and represents the number of sibling nodes of the graph. *Depth* is a property of graphs which is related to cardinality of paths existing in the graph. The arcs which are considered are only *is-a* arcs but this only applies to directed graphs.

4 Experimental Results

As explained above, *OntoMetrics* is a web-based tool that validates and displays statistics about a given ontology, where a user can upload an OWL/RDF ontology source. Using this third-party application, we conducted several experiments for applying the metrics described in Sect. 3 to ontologies produced with the GF application. Here, we, in particular, present the results obtained in relation to the library example ontology described in Sect. 2.

The results of computing base and class-axioms metrics on the library ontology from Fig. 2 are shown in Table 1. The results of computing Object and data property axioms metrics on the library ontology from Fig. 2 are shown in Table 2. The individual and annotation axioms metrics on the library ontology

are presented in Table 3. The schema, knowledgebase and graph metrics are presented in Table 4. Finally, the class metrics computed for each of the classes of the library case study are shown in Table 5.

Table 1. Base and class axioms metrics of the library ontology in Fig. 2

Base metrics	
Axioms:	208
Logical axioms count:	150
Class count:	13
Total classes count:	13
Object property count:	2
Total object properties count:	2
Data property count:	35
Total data properties count:	35
Properties count:	37
Individual count:	8
Total individuals count:	8
DL expressivity:	$\mathcal{ALC}(D)$

Class axioms:	
SubClassOf axioms count:	12
Equivalent classes axioms count:	0
Disjoint classes axioms count:	1
GCICount:	0
HiddenGCICount:	0

Table 2. Object and data property axioms metrics of the library ontology in Fig. 2

Object property axioms	
SubObjectPropertyOf axioms count:	0
Equivalent object properties axioms count:	0
Inverse object properties axioms count:	0
Disjoint object properties axioms count:	0
Functional object properties axioms count:	0
Inverse functional object properties axioms count:	0
Transitive object property axioms count:	0
Symmetric object property axioms count:	0
Asymmetric object property axioms count:	0
Reflexive object property axioms count:	0
Irreflexive object property axioms count:	0
Object property domain axioms count:	2
Object property range axioms count:	2
SubPropertyChainOf axioms count:	0

Data property axioms	
SubDataPropertyOf axioms count:	0
Equivalent data properties axioms count:	0
Disjoint data properties axioms count:	0
Functional data property axioms count:	0
Data property domain axioms count:	35
Data Property range axioms count:	35

Table 3. Individual and annotation axioms metrics of the library ontology in Fig. 2

Individual axioms	
Class assertion axioms count:	12
Object property assertion axioms count:	2
Data property assertion axioms count:	49
Negative object property assertion axioms count:	0
Negative data property assertion axioms count:	0
Same individuals axioms count:	0
Different individuals axioms count:	0

Annotation axioms	
Annotation axioms count:	0
Annotation assertion axioms count:	0
Annotation property domain axioms count:	0
Annotation property range axioms count:	0

Now we provide an interpretation of the results that we obtained. The class count coincides with the number of classes that we can manually count in the UML diagram of Fig. 2. The data property count exceeds the number of attributes seen in the UML diagram because every attribute of the database triggers the creation of one data property but in order to model the class hierarchy we have to manually create the attributes of each of the corresponding superclasses thus producing some overlap and consequently some redundancy too. The new functionality of modifying relations (see step (6) in the workflow described in Sect. 2) allows to continue having only 2 object properties (where one remained from Loan to User but another one that moved from Loan to Thesis towards Loan

Table 4. Schema, knowledgebase and graph metrics of the library ontology in Fig. 2

Schema metrics	
Attribute richness:	2.692308
Inheritance richness:	0.923077
Relationship richness:	0.2
Attribute class ratio:	0.0
Equivalence ratio:	0.0
Axiom/class ratio:	16.0
Inverse relations ratio:	0.0
Class/relation ratio:	0.866667

Knowledgebase metrics	
Average population:	0.615385
Class richness:	0.692308

Graph metrics	
Absolute root cardinality:	3
Absolute leaf cardinality:	8
Absolute sibling cardinality:	13
Absolute depth:	37
Average depth:	2.466667
Maximal depth:	4
Absolute breadth:	15
Average breadth:	2.5
Maximal breadth:	4
Ratio of leaf fan-outness:	0.615385
Ratio of sibling fan-outness:	1.0
Tangledness:	0.153846
Total number of paths:	15
Average number of paths:	3.75

Table 5. Class metrics of the library ontology in Fig. 2

Class metrics	Book	GraduateThesis	Loan	MScThesis	Magazine	Material	PhDThesis	PostgradThesis	Printed	StudentUser	TeacherUser	Thesis	User
Class connectivity:	0	0	2	0	0	0	0	0	0	0	0	0	0
Class fulness:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Class importance:	0.125	0.0	0.125	0.125	0.25	0.125	0.125	0.25	0.375	0.125	0.125	0.75	0.5
Class inheritance richness:	0.0	0.0	0.0	0.0	0.0	1.3	0.0	6.5	6.5	0.0	0.0	2.166	6.5
Class readability:	0	0	0	0	0	0	0	0	0	0	0	0	0
Class relationship richness:	0	0	0	0	0	0	0	0	0	0	0	0	0
Class children count:	0	0	0	0	0	10	0	2	2	0	0	6	2
Class instances count:	1	0	1	1	2	9	1	2	3	1	1	6	4
Class properties count:	0	0	0	0	0	0	0	0	0	0	0	0	0

to Material, making a loan more generic than before). The number of subclassof axioms count (12) exceeds the expected number (10) from the UML diagram in Fig. (2.a) because GF produces redundant is-a relationships as it can be seen in Fig. (2.b)—there are two paths from MScThesis to Thesis and another two from PhDThesis to Thesis instead of one in each case. The results of Table 2 show that indeed the ontologies produced by GF have only property axioms defining attributes and dispense with other type of richer semantic relations that would impose an excessive time complexity worst-case on the reasoner, thus adhering to the tenets of the OBDA paradigm that propose only using light-weight ontologies to guarantee polynomial time complexity reasoning. Table 3 shows that only 12 of 13 classes have individuals, that is consistent with our case study where there are no StudentUser of the library. One important aspect revealed by the annotation axioms metrics is that GF is not impacting them at all, meaning that the usability of the ontologies produced would be at stake and pointing out a place of the framework where to introduce further improvement. In Table 4, we see that 3 roots are detected (User, Loan and Material) and 8 leaves, that is all consistent with the UML diagram in Fig. (2.b). According to Table 5, Thesis appears to be the most important class followed by User, results that coincide with the richness of the relational database diagram.

5 Conclusions and Future Work

We used a third-party application called OntoMetrics and use it for conducting several experiments for applying ontology metrics to ontologies produced with the GF application. We presented a particular case study on measuring an on-

tology produced in previous work, what led to adding new functionality to GF (viz., modifying domain and range of materialized properties from database relations). The results obtained show that the ontologies generated with GF are reasonably good for being used in SW applications as they are validated correctly and pass all of the filters for the main OWL2 profiles, thus making them suitable for processing with lightweight reasoners. One important aspect revealed that GF does not impact the annotation axioms metrics, meaning that the usability of the ontologies produced would be at stake and therefore pointing out a place of the framework where to introduce further improvement.

Acknowledgments. This research is funded by Secretaría General de Ciencia y Técnica, Universidad Nacional del Sur, Argentina and by Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC-PBA).

References

1. Raad, J., Cruz, C.: A survey on ontology evaluation methods. In: Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. (November 2015)
2. García, J., F.J., F.G.P., Therón, R.: A survey on ontology metrics. In Lytras, M., Pablos, P.O.D., Ziderman, A., Roulstone, A., Maurer, H., Imber, J., eds.: Knowledge Management, Information Systems, E-Learning, and Sustainability Research. WSKS 2010. Communications in Computer and Information Science. Volume 111. Springer, Berlin, Heidelberg (2010)
3. Franco, M., Vivo, J.M., Quesada-Martínez, M., Duque-Ramos, A., Fernández-Breis, J.T.: Evaluation of ontology structural metrics based on public repository data. Briefings in Bioinformatics **21**(2) (March 2020) 473—485
4. Bansala, R., Chawlab, S.: Evaluation metrics for computer science domain specific ontology in semantic web based irscsd system. International Journal of Computer (IJC) **19**(1) (2015) 129–139
5. Plyusnin, I., Holm, L., Törönen, P.: Novel comparison of evaluation metrics for gene ontology classifiers reveals drastic performance differences. PLOS Computational Biology (november 2019) 1–27
6. Tovar, M., Pinto, D., Montes, A., González-Serna, G.: A metric for the evaluation of restricted domain ontologies. Comp. y Sist. **22**(1) (jan 2018)
7. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: A theoretical framework for ontology evaluation and validation. In: SWAP 2005 - Semantic Web Applications and Perspectives, Proceedings of the 2nd Italian Semantic Web Workshop, University of Trento, Trento, Italy (December 2005) 14–16
8. Gómez, S.A., Fillottrani, P.R.: Specification of the schema of spreadsheets for the materialization of ontologies from integrated data sources. In Pesado, P., J., E., eds.: Computer Science – CACIC 2020. Volume 1409., Cham, Springer International Publishing (2021) 247–262
9. Tartir, S., Arpinar, I.B., Sheth, A.P.: Ontological evaluation and validation. Theory and Applications of Ontology: Computer Applications (2010) 115–130

ISBN 978–987–633–574–4

DeepSeed: aplicación multiplataforma para estimar la calidad de granos de maíz

Máximo Librandi[1], Joshua Corino[1], Paula Tristan[1], Laura Felice[1],

[1]Instituto de Investigación en Tecnología Avanzada – INTIA - Facultad de Ciencias Exactas.
Universidad Nacional del Centro de la Pcia. de Buenos Aires.
{maximolibrandi, joshuc98}@gmail.com, {ptristan,
lfelice}@exa.unicen.edu.ar

Abstract. En los últimos años, la implementación de nuevas tecnologías en la producción agropecuaria ha permitido un importante salto en las magnitudes producidas. Sin embargo, la determinación de calidad de los granos como requisito para establecer su precio aún no ha generado ningún cambio o innovación durante su proceso. Luego de décadas, la tarea de clasificación y estimación del grado de calidad comercial continúa realizándose de forma manual por los peritos clasificadores de granos. En este artículo se presenta DeepSeed, una aplicación multiplataforma que, utilizando el modelo de Deep Learning FasterRCNN Resnet152 COCO provisto por TensorFlow, determina el grado de comercialización del maíz, a través del procesamiento de la imagen de una pequeña muestra.

Keywords: Clasificación, Calidad de granos, Inteligencia Artificial, Deep Learning, Faster RCNN, Progressive Web App.

1 Introducción

En las últimas dos décadas el sector agrícola argentino registró grandes transformaciones; cambios en las formas organizacionales, en las técnicas productivas y en la tecnología aplicada que dieron lugar a un salto de gran magnitud en las cantidades producidas.

El contexto de trabajo de DeepSeed está centrado en una etapa de la cadena agroexportadora donde la evolución tecnológica aún no ha introducido cambios: la determinación de la calidad de los granos. Se entiende por calidad de un cereal al conjunto de defectos que desmejoran una partida. Estos defectos son los factores que se tienen en cuenta para determinar la calidad de una partida en función de la cantidad o la intensidad que los mismos se encuentran presentes en una muestra.

Las metodologías de clasificación y medición de la calidad de los granos no han cambiado desde sus orígenes, ya que aún continúan realizándose de forma manual por los Peritos Clasificadores de Granos. En este contexto, la efectividad depende en gran medida de la capacidad y experiencia que tenga el perito, así como de otros factores externos como estrés o cansancio. De este modo, la automatización de este proceso

impondría numerosos beneficios a lo largo de la cadena de comercialización, permitiendo ofrecer una herramienta de soporte y asistencia a la tarea de los peritos. Además, brindaría asistencia a los productores agropecuarios durante el proceso de recolección, permitiendo aplicar ajustes en la maquinaria en caso de ser necesario, o realizar una adecuada segmentación de su cosecha previa a su almacenamiento.

La Inteligencia Artificial (IA) [15], se ha afianzado y constituye hoy la gran tendencia en desarrollo de software y tecnología. El Deep Learning (DL) [3], [5], [8] es sin duda el área de investigación más popular dentro del campo de la IA. El área de Computer Vision [1] se ha vuelto algo mucho más fácil e intuitivo gracias a los recientes avances en DL, el cual ha revolucionado el reconocimiento de patrones. Este reconocimiento se hace a través de las redes neuronales convolucionales (CNNs) [11], una técnica de DL que utiliza filtros con el fin de extraer características de las imágenes y aprenderlas, para luego poder realizar tareas como detección de objetos o clasificación de imágenes. En el uso de nuevas tecnologías en el desarrollo de aplicaciones multiplataforma, cabe mencionar las Progressive Web Apps (PWA) [23], aplicaciones web que proporcionan un producto instalable y una experiencia de aplicación en computadoras de escritorio y dispositivos móviles, que se crean y entregan directamente a través de la web.

Se presenta aquí un resumen del desarrollo de una aplicación multiplataforma (DeepSeed) que, utilizando técnicas de DL, permite detectar y clasificar los objetos presentes en una foto de una muestra de granos de maíz y así determinar su calidad de acuerdo a las normas vigentes.

Este documento está organizado de la siguiente forma: en la sección 2 se introducen los trabajos más importantes relacionados con la determinación de calidad de diversos cultivos. En la sección 3 se detalla cada etapa de la aplicación propuesta. La sección 4 presenta las métricas de performance analizadas sobre el modelo de DL y analiza los resultados obtenidos por DeepSeed en cuatro casos de estudio. En la sección 5 se describen las conclusiones arribadas y los trabajos futuros que surgen del análisis del trabajo finalizado. Finalmente, en la sección 6, se encuentran las referencias bibliográficas.

2 Estado del arte

En esta sección se resume el trabajo de algunos proyectos de investigación académica y a nivel empresarial de la clasificación de maíz y de otros cultivos, permitiendo entender la evolución de las diferentes investigaciones y desarrollos.

El trabajo de Saleres (2018) [2], considerado antecesor principal de DeepSeed, se basa en una aplicación web que permite determinar el grado de calidad de una muestra de granos de maíz a partir de una imagen de la misma, utilizando técnicas de procesamiento de imágenes.

En cuanto a los trabajos relacionados con el maíz, se destacan el trabajo de Bhurtel et al. (2019) [10] donde se implementa un sistema de clasificación de calidad de un lote de semillas de maíz utilizando técnicas de DL. A partir de imágenes que incluyen semillas en buen estado, semillas dañadas y materia extraña, el sistema categoriza la

calidad del lote como excelente, buena, promedio, mala y pésima. En la propuesta de Huang et al. (2019) [13] se comparan las técnicas de Convolutional Neural Networks (CNNs) y Transfer Learning con los algoritmos tradicionales de Machine Learning en la clasificación de semillas de maíz.

En el tratamiento de otras semillas, se puede mencionar a Riat et al. (2018) [19] creadores de CamWheat: una tecnología que permite conocer la calidad del trigo en 8 segundos, también basada en IA. Suseendran et al. (2020) [7] aplican las técnicas Multi-Layer Perceptron (MLP) Neural Network y Neuro-Fuzzy Classifier para la clasificación entre trigo, arroz y maíz.

Por otro lado, ZoomAgri [24] es una empresa nacional que desarrolla tecnología de determinación de calidad de commodities agrícolas, por medio de Procesamiento de Imágenes, Inteligencia Artificial e Internet de las Cosas (IoT). Su primera gran disrupción es ZoomBarley, un escáner que permite determinar la variedad de cebada en menos de 5 minutos y a una fracción del costo de los métodos actuales.

3 Metodología aplicada

La Figura 1 resume la metodología empleada para obtener la aplicación multiplataforma objetivo. Básicamente, a partir de una imagen de una muestra de maíz, se determina el grado de calidad comercial de la misma, utilizando un modelo de DL entrenado y técnicas de procesamiento de imágenes. Seguidamente se describen las partes correspondientes al modelo diseñado.

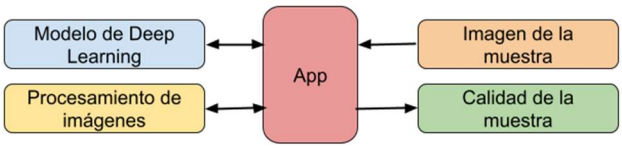


Fig. 1. Metodología de trabajo propuesta.

3.1 Modelo de Deep Learning

La tarea de construcción del modelo de DL, sin dudas la más importante de esta propuesta, consta de la ejecución de varias tareas que se detallan a continuación.

3.1.1 Generación del dataset

En primer lugar, y con el fin de reentrenar la red neuronal convolucional, es necesario crear el dataset. Para esto, basados en el caso de aplicación, los peritos de la Cámara Arbitral de la Bolsa de Cereales de Buenos Aires separaron y etiquetaron objetos dentro de las distintas clases a tener en cuenta. Entre estas clases se encuentran: granos de maíz (semillas en buen estado), granos dañados, granos quebrados, materia extraña, semillas de chamico y granos picados.

Posteriormente se capturaron numerosas imágenes variando las clases y cantidades de objetos, concluyendo en un total de 909 imágenes de 4096 x 3024 píxeles. Debido a que se trabaja con un algoritmo de detección de objetos supervisado, resulta necesario etiquetar los objetos presentes en cada imagen, entregando al modelo no solo la imagen, sino también los cuadros delimitadores y la clase de cada objeto que figura en la misma. Para cumplir con esta tarea se utilizó LabelImg [22], que contabilizó un total de 4948 etiquetas en el dataset.

Finalmente, el dataset fue dividido aleatoriamente, tomando el 80% de las imágenes para entrenamiento, y el 20% restante para evaluación. Como resultado de esta división, 3898 objetos fueron utilizados para entrenamiento, mientras que los otros 1050 para evaluación.

3.1.2 Entrenamiento del modelo

Una vez generado el dataset adecuado, el siguiente paso es el entrenamiento del modelo. Esta compleja tarea requiere de la ejecución iterativa de varias acciones. En particular, la evaluación y validación resultan necesarias para asegurar la correctitud del mismo. En la Figura 2 se puede visualizar el pipeline propuesto para obtener un modelo de DL capaz de detectar y clasificar los objetos presentes en una imagen de una muestra de maíz. A continuación, se describe detalladamente cada paso del entrenamiento y las decisiones tomadas en cada uno.

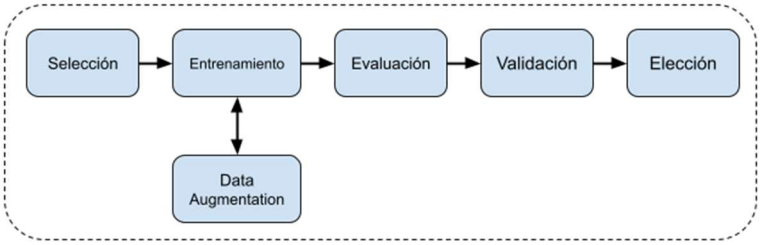


Fig. 2. Pasos para el entrenamiento del modelo de Deep Learning.

Selección del modelo

En principio, es necesario seleccionar el modelo de DL al que se le aplicará Transfer Learning entre los modelos que provee la API de detección de objetos de TensorFlow [18].

En primer lugar, se estudió la posibilidad de emplear un modelo basado en el algoritmo Single Shot Detector (SSD) [17]. Sin embargo, luego de tiempo de investigación y debido a las restricciones de tamaño de imágenes del SSD, lo cual lo desfavorecen en la detección de pequeños objetos, la siguiente alternativa fue el algoritmo Faster RCNN [14], el cual ha probado ser uno de los más eficientes en la familia de los RCNN para la detección de objetos en la actualidad [12].

Se optó por el modelo Faster RCNN ResNet 152 800x1333 COCO. Dicho modelo posee como extractor de características una ResNet 152 [9], una red residual de 152 capas. Otra característica principal, tal como su nombre lo específica, es que el

modelo define 800 píxeles como valor mínimo y 1333 como máximo para la redimensión de las imágenes durante su aprendizaje, manteniendo la relación de aspecto. TensorFlow ofrece este modelo pre-entrenado con el dataset de COCO [16], sobre el cual se aplicará la técnica de Transfer Learning para ser utilizado en un set de datos más específico: el de granos de maíz.

Si bien el modelo elegido no puede ser deployado en una aplicación móvil, requiere una conexión a Internet mínima para su funcionamiento a través de una API y demanda mayor tiempo y recursos para su entrenamiento, brinda mayor precisión en la detección de objetos pequeños al aceptar imágenes de mayor tamaño.

Entrenamiento del modelo

Con el modelo adecuado elegido, la siguiente tarea es el reentrenamiento del mismo para adaptarlo al caso de aplicación. El reentrenamiento se realizó utilizando Google Colab [21], para lo cual fue necesario la configuración de parámetros, como por ejemplo, el número de clases a detectar, el tamaño del dataset, entre otros.

Durante este proceso fue necesario aplicar técnicas de Data Augmentation, para ampliar el tamaño del dataset realizando una serie de cambios aleatorios en las imágenes. En este trabajo se aplicaron operaciones de flip y rotación, sin aplicar operaciones de transformación del color, brillo u otros factores que reducen la sensibilidad del modelo al color, ya que fueron probadas causando que el modelo confunda las principales clases asociadas al maíz.

Para poder evaluar el entrenamiento del modelo, se generan archivos de control o checkpoint cada 5 mil steps, a partir de los cuales se construye el grafo de inferencia utilizado luego para la validación del modelo.

Evaluación del modelo

La evaluación aporta una noción de cómo el modelo está aumentando su capacidad de generalización. Esta tarea fue automatizada modificando la versión original provista por TensorFlow para que se ejecute cada vez que el entrenamiento genere un nuevo checkpoint. Como resultado de este proceso se obtuvieron las métricas que se listan a continuación:

- RPN Localization Loss: significa la pérdida de localización de los cuadros delimitadores generados por la red de propuesta de región (RPN o Region Proposal Network) que incluye la arquitectura Faster RCNN.
- RPN Objectness Loss: es la pérdida del clasificador de la RPN que determina si un cuadro propuesto es un objeto de interés o forma parte del fondo de la imagen.
- Box Localization Loss: representa el error de los recuadros de los objetos identificados, es decir, de las coordenadas sugeridas para cada objeto.
- Box Classification Loss: determina el error existente en la clasificación de los objetos detectados, donde se define a qué clase pertenece cada objeto.
- Total Loss: es la suma total de las cuatro pérdidas definidas anteriormente.

En la sección 4 -Resultados- se muestran los gráficos de la evolución de dichas métricas a lo largo del entrenamiento.

Validación del modelo

Durante el proceso de validación del modelo se procesan imágenes no utilizadas para el entrenamiento y para la evaluación, para así poder observar su comportamiento a través de métricas de performance y en base a ello tomar decisiones de implementación. Si se utilizan métricas más significativas sobre un conjunto de datos acotado, es posible seleccionar el modelo que brinda los mejores resultados en la generalización. Este proceso también se realizó automáticamente cada vez que se generaba un nuevo checkpoint durante el entrenamiento, pero esta vez sobre imágenes similares a las que la aplicación procesará una vez deployada para su uso (es decir, imágenes con más de 100 granos). De este modo, fueron tomadas 27 nuevas imágenes que contenían un total de 3119 objetos.

Cada nuevo checkpoint significa un nuevo posible modelo que debía ser validado, lo que requiere que se realice la detección de objetos en cada una de las 27 imágenes destinadas a la validación. Se utilizaron tres alternativas para llevar a cabo la detección de objetos: una con la imagen en tamaño original; la segunda, basada en [6], en donde la imagen es dividida en cuatro subimagenes a las cuales se le aplica la detección, y los resultados son unidos eliminando los que poseen un alto grado de superposición (IoU); y la última, una combinación de las dos versiones anteriores que obtiene los resultados con mayor confidencialidad. Por último, se compararon los resultados obtenidos utilizando una matriz de confusión para cada alternativa de detección, en donde se tienen filas y columnas por cada clase que detecta el modelo.

Elección del modelo

Se seleccionó el modelo a incorporar en la versión final de la aplicación utilizando la métrica conocida como Macro F1 Score [4], la cual permite evaluar la performance de un algoritmo de clasificación multiclase utilizando la aproximación “one vs all”. En esta aproximación, los valores de una clase particular son seleccionados y convertidos en ejemplos positivos, mientras que el resto de las clases, en negativos. Este proceso se repite para cada clase detectada por el modelo y se obtienen las métricas correspondientes a cada una de ellas.

De este modo, utilizando los conceptos de verdaderos positivos, falsos positivos y falsos negativos, se calcularon las métricas Precision, Recall y F1 Score para cada clase. Posteriormente, con el objetivo de comparar la performance de los modelos entre sí, se combinaron los F1 Scores en un solo número: el F1 Score general del modelo. La denominada Macro F1 Score se calcula como la media aritmética simple de los F1 Scores por clase. En la sección 4 -Resultados- se ilustra la evolución de dicha métrica durante el entrenamiento.

El modelo correspondiente al checkpoint 315.000 fue el seleccionado para incluir en la aplicación resultante. A pesar de que por una mínima diferencia no fue el que logró el mayor valor para la métrica Macro F1 Score, se eligió ya que consiguió el valor más alto de F1 Score para la clase granos dañados.

3.2 La Aplicación

Se desarrolló una Progressive Web App (PWA) que, a partir de la imagen de la muestra de maíz que se desea analizar, determina su grado de calidad. Esta app se implementó como una aplicación web, utilizando HTML, CSS y JavaScript, y se le añadieron los componentes necesarios para convertirla en una PWA.

La API Rest desarrollada es una aplicación Python en donde se emplea Flask [20] para definir los endpoints necesarios. La API, la cual ya tiene el modelo de DL precargado con el objetivo de optimizar el tiempo de respuesta, una vez que recibe la imagen, realiza la detección de los objetos, obteniendo las coordenadas de los cuadros delimitadores, las clases a las que pertenecen y las confidencias con las que fueron detectados. Luego, utilizando procesamiento de imágenes, más precisamente la técnica de binarización, se aproxima el valor del peso hectolítrico (PH) y se calculan los porcentajes de superficie que ocupa cada rubro de calidad, utilizando la hoja A4 como objeto de referencia. Con estos datos, la API determina el grado de calidad de la muestra y retorna los resultados a la PWA en formato JSON.

La PWA está hosteada en los servidores de Firebase de Google, mientras que la API Rest desarrollada corre en sistemas locales. Si se deseara escalar la aplicación, solo sería necesario subir y correr la API en un servidor o servicio en la nube.

4 Resultados

En primer lugar, en esta sección se muestran los análisis de evaluación y validación realizados sobre el modelo de DL que permitieron concretar la selección de la mejor opción del modelo. Posteriormente, se presentan los resultados obtenidos con la aplicación en el análisis de cuatro muestras de maíz.

4.1 Precisión del Modelo

Como resultado del proceso de evaluación del modelo de DL se obtuvieron las métricas RPN Localization Loss, RPN Objectness Loss, Box Localization Loss, Box Classification Loss y Total Loss. La evolución de la pérdida total en los 500.000 pasos del entrenamiento puede visualizarse en la Figura 3. Se observa cómo el valor de las pérdidas tiende a converger en un valor mínimo.

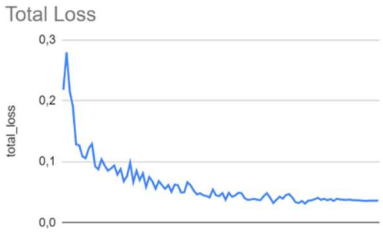


Fig. 3. Evolución de la métrica de evaluación Total Loss a lo largo del entrenamiento.

Para la validación del modelo se procesaron otras 27 imágenes que contenían 3119 objetos. Luego, utilizando la matriz de confusión, se calcularon las métricas definidas a continuación para cada clase detectada, en las distintas alternativas de detección.

- Precisión (P): la precisión es intuitivamente la capacidad del modelo de no etiquetar como positiva una muestra que es negativa.
- Recall (R): responde a la pregunta ¿qué proporción de los positivos reales se clasifica correctamente?
- F1 Score: es un número real que combina la precisión y el recall, otorgando un mayor peso a los números más bajos. Se calcula utilizando una media armónica.

El siguiente paso consiste en combinar los F1 Scores por clase en el F1 Score general del modelo, para lo cual se utilizó la métrica Macro-F1 Score, calculada como la media aritmética simple de los F1 Scores por clase.

$$Macro\ F1\ Score = \frac{\sum_{i=1}^{K} F1\ Score(K)}{K}$$

En la Figura 4 se puede observar el comportamiento de la métrica Macro F1-Score para las tres alternativas de detección durante el proceso de validación. Se deduce que la opción que combina las dos alternativas de detección obtiene mejores resultados en la mayoría de los casos. Se utilizó el modelo correspondiente al checkpoint 315.000 para la versión final de la aplicación, el cual obtuvo un valor de Macro F1 Score de 0.89928, logrando el segundo mayor valor. El modelo 305.000 lo superó con un valor de 0.90735. Sin embargo, el modelo 315.000 consiguió un F1 Score para la clase grano dañado de 0.7535, siendo el mejor para este rubro.

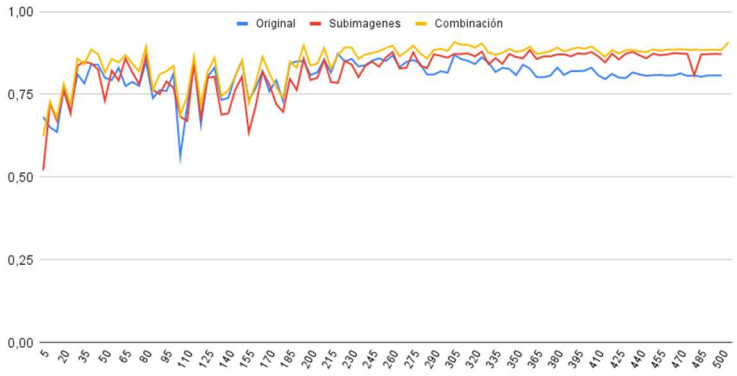


Fig. 4. Macro F1 Score para las tres alternativas de detección propuestas.

4.2 Calidad del maíz

En esta sección se detallan los resultados obtenidos por DeepSeed con el objetivo de evaluar su desempeño. Para esto, se adquirieron cuatro submuestras de maíz de 50 gramos a partir de cuatro muestras clasificadas por peritos clasificadores de granos de la Cooperativa Agropecuaria Gral. Necochea. Las submuestras fueron dispersadas individualmente sobre una hoja A4 blanca y fotografiadas. Estas cuatro pruebas,

llamadas casos de estudio, fueron evaluadas por DeepSeed y sus resultados se comparan con el análisis realizado por los peritos clasificadores en la Tabla 1.

Tabla 1. Comparación de resultados obtenidos por DeepSeed y peritos.

	Caso estudio 1		Caso estudio 2		Caso estudio 3		Caso estudio 4		
Rubro	Deep Seed	Peritos	Deep Seed	Peritos	Deep Seed	Peritos	Deep Seed	Peritos	Error
G. dañados	1.40	0.80	0.00	0.00	0.00	0.00	0.71	0.40	0.2275
G. quebrados	0.00	0.20	0.83	0.10	0.16	0.30	0.00	0.20	0.3175
Mat. extraña	1.81	1.00	0.46	0.05	0.00	0.10	0.65	1.00	0.4175
P.Hectolítrico	71.0	71.0	73.0	73.0	77.72	77.8	75.1	75.0	0.045
Grado	3	3	2	2	1	1	1	1	0

Se concluye que la aplicación desarrollada logra un buen comportamiento al momento de determinar el grado de calidad de cada submuestra analizada, ya que logra enmarcar los cuatro casos de estudio en el mismo grado de calidad resultantes de la tarea manual. Se puede deducir que DeepSeed obtuvo un error absoluto promedio de 0.2015. Cabe destacar que las submuestras analizadas por DeepSeed no son las mismas que las analizadas manualmente por los expertos, por lo que los errores obtenidos pueden deberse a esta diferencia. Por su parte, los altos valores alcanzados por la aplicación para ciertos rubros de calidad son causados por la no homogeneidad de las submuestras seleccionadas.

5 Conclusiones y trabajos futuros

El presente trabajo intenta cubrir la vacancia en la automatización de la clasificación de granos, introduciendo al sector agropecuario en el uso de la IA. Estas nuevas tecnologías aplicadas a la agricultura son sumamente necesarias para el incremento de la productividad.

La aplicación desarrollada, además de incorporar el uso de DL para clasificar los granos de maíz, posee el agregado de ser una PWA, lo que posibilita el acceso desde cualquier dispositivo. Para su uso solo se requiere depositar la muestra sobre una hoja blanca tamaño A4 y tomar una fotografía.

Esta propuesta permite conocer la calidad de los granos durante la cosecha y en sus posteriores etapas de almacenamiento y comercialización, elevando así los estándares de exportación y suministrando trazabilidad a la cadena agroexportadora. Además, constituye una importante herramienta de soporte para los peritos clasificadores.

Con el objetivo de perfeccionar la herramienta, será necesario aumentar el volumen de los datos para entrenar el modelo de DL, ya sea incorporando nuevas muestras al dataset o tomando nuevas capturas bajo diferentes condiciones y entornos. Otra variante seria utilizar un modelo de DL que esté pre entrenado con un set de datos más relacionado a los objetos que posteriormente reconocerá.

Finalmente, una vez optimizado el modelo para la clasificación de granos de maíz, resulta interesante extender la herramienta a otros cultivos como trigo, soja, girasol y

cebada. Esta extensión implica nuevos entrenamientos del modelo para detectar los rubros determinantes de calidad en cada cultivo.

6 Referencias

1. A. Jain, Deep Learning for Computer Vision – Introduction to Convolution Neural Networks. 2016. Disponible en: https://www.analyticsvidhya.com/blog/2016/04/deep-learning-computer-vision-introduction-convolution-neural-networks/ (accedido Ago. 2021)
2. A. S. Saleres, Aplicación web para la clasificación de granos de maíz. Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires. 2018.
3. A. Zhang, Z. C. Lipton, M. Li y A. J. Smola, Dive into Deep Learning. 2020.
4. B. Shmueli, Multi-Class Metrics Made Simple, Part II: the F1-score. 2019. Disponible en: https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1 (accedido Ago. 2021)
5. C. Aggarwal, Neural Networks and Deep Learning: A Textbook. Springer. 2018.
6. F. Ozge Unel, B. O. Ozkalayci y C. Cigla, The Power of Tiling for Small Object Detection. 2019.
7. G. Suseendran, E. Chandrasekaran, D. Akila y D. Balaganesh, Automatic Seed Classification by Multi-Layer Neural Network with Spatial-Feature Extraction. 2020.
8. I. Goodfellow, Y. Bengio y A. Courville, Deep Learning. 2016.
9. K. He, X. Zhang, S. Ren y J. Sun, Deep Residual Learning for Image Recognition. 2015.
10. M. Bhurtel, J. Shrestha, N. Lama, S. Bhattarai, A. Uprety y M. Kumar Guragain. Deep Learning based Seed Quality Tester. 2019.
11. M. Manav, Introduction to Convolutional Neural Networks (CNN). 2021. Disponible en: https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/ (accedido Ago. 2021)
12. P. Sharma, A Step-by-Step Introduction to the Basic Object Detection Algorithms (Part 1). 2018. Disponible en: https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/ (accedido Ago. 2021)
13. S. Huang, X. Fan, L. Sun , Y. Shen y X. Suo, Research on Classification Method of Maize Seed Defect Based on Machine Vision. 2019.
14. S. Ren, K. He, R. Girshick y J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. 2015.
15. S. Rusell y P. Norvig, Artificial Intelligence: A modern approach (Fourth Edition). 2020.
16. T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick y P. Dollár, Microsoft COCO: Common Objects in Context. 2014.
17. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu y A. C. Berg, SSD: Single Shot MultiBox Detector. 2015.
18. API de detección de objetos de TensorFlow. GitHub. Disponible en: https://github.com/tensorflow/models/tree/master/research/object_detection
19. Crearon una tecnología que permite conocer la calidad del trigo en 8 segundos. Puntobiz.com. Disponible en: https://puntobiz.com.ar/lo-nuevo/2018-8-9-6-1-0-crearon-una-tecnologia-que-permite-conocer-la-calidad-del-trigo-en-8-segundos (accedido Ago. 2021)
20. Flask. Disponible en: https://flask.palletsprojects.com/en/2.0.x/
21. Google Colaboratory. Disponible en: https://colab.research.google.com/
22. LabelImg. Disponible en: https://github.com/tzutalin/labelImg
23. Progressive Web Apps. Disponible en: https://web.dev/progressive-web-apps
24. ZoomAgri. Disponible en: https://zoomagri.com/

ISBN 978–987–633–574–4

Análisis de Comunicaciones en Aplicaciones Móviles 3D para Domótica

Diego Encinas, Sebastián Dapoto, Federico Cristina, Cristian Iglesias, Federico Arias, Pablo Thomas, Patricia Pesado

Instituto de Investigación en Informática (III-LIDI). Facultad de Informática, Universidad Nacional de La Plata - Centro Asociado CIC. Buenos Aires, Argentina
{dencinas, sdapoto, fcristina}@lidi.info.unlp.edu.ar
{cristianniglesias, fede98.arias}@gmail.com
{pthomas, ppesado}@lidi.info.unlp.edu.ar

Resumen El presente trabajo expone el estudio y análisis de la performance en las comunicaciones de una aplicación móvil orientada a redes de sensores con tecnología en la Nube o *Cloud Computing*. La aplicación móvil reproduce un entorno visual 3D que está vinculado a diferentes dispositivos y sensores de una vivienda u oficina. Se desarrolló una comunicación bidireccional entre los dispositivos y sensores con la aplicación 3D desarrollada con la herramienta Unity. Se utilizaron servidores físicos como también virtuales (*Cloud Computing*). Además, se obtuvieron métricas de rendimiento de comunicaciones como latencia y throughput del sistema.

1 Introducción

Una vivienda domótica integra un conjunto de automatismos en cuanto a electricidad, electrónica, robótica, informática y telecomunicaciones, con el objetivo de asegurar al usuario una mejora en el confort, la seguridad, el ahorro energético, las facilidades de comunicación y las posibilidades de entretenimiento. La domótica se centra en dos puntos de vista: el del usuario y el tecnológico. Desde el punto de vista del usuario, una vivienda automatizada permite una mayor calidad de vida a través de las nuevas tecnologías, reduciendo el trabajo doméstico, mejorando el bienestar y, por ende, mejorando el control del consumo. Desde el punto de vista tecnológico, se encuentra la capacidad de los distintos objetos pertenecientes a una vivienda, con la posibilidad de intercomunicarse entre sí, a través de un soporte de comunicaciones [1]. El protocolo de comunicaciones más utilizado para la interconexión de los distintos dispositivos es el protocolo MQTT. Sus siglas en inglés corresponden a transporte de telemetría de cola de mensajes y es un protocolo abierto de máquina a máquina (M2M). Se caracteriza por ser orientado a mensajes permitiendo la comunicación entre los dispositivos de forma asincrónica y eficiente. Es uno de los estándares más utilizado para internet de las cosas (IoT, Internet of Things) [2]. En este trabajo, se explica el

ISBN 978–987–633–574–4

desarrollo de una comunicación bidireccional entre dispositivos móviles y sensores con una aplicación móvil 3D desarrollada mediante la herramienta Unity [3]. Además, se realiza un análisis por medio de métricas de rendimiento de comunicaciones del sistema.
Este trabajo se organiza del siguiente modo: a continuación se menciona la aplicación 3D generada; luego se plantean los componentes relacionados a comunicaciones del sistema. En el capítulo 4 se detalla el desarrollo de la comunicación bidireccional; seguido a esto se muestra la experimentación realizada. En el capítulo 6 se explican las métricas obtenidas. Finalmente se presentan los resultados y las conclusiones.

2 Aplicación móvil 3D

El cliente es una aplicación móvil 3D desarrollada en Unity. Éste es un framework de desarrollo de aplicaciones 3D que se destaca por la cantidad de documentación disponible, una numerosa y muy activa comunidad de usuarios, gran variedad de componentes pre-desarrollados (assets) y plugins que facilitan la integración con otras herramientas. Además, Unity ofrece variadas opciones de plataformas de publicación.
La aplicación móvil 3D permite recrear el conjunto de ambientes de una vivienda u oficina, incluyendo los objetos presentes en dichos espacios. Se cuenta con la posibilidad de manejar diferentes tipos de dispositivos, representados mediante hardware que simulan objetos del mundo real tales como ventiladores, luces, lámparas, televisores, entre otros. Entre las funciones con las que cuenta la aplicación se destacan:

- Creación de una casa. Al utilizar la aplicación por primera vez, es necesario crear la estructura de la casa que se desea controlar. La aplicación permite crear los modelos de todos los ambientes de la casa, con el nombre y tamaño correspondientes.
- Edición de una casa. En cualquier momento es posible modificar la estructura de una casa.
- Colocación de dispositivos. Mediante este módulo es posible seleccionar una posición en la pared, piso o techo de un ambiente determinado y colocar un dispositivo.
- Configuración. Permite entre otras cosas configurar la dirección del broker y asociar cada dispositivo con el sensor que lo maneja.
- Control de una casa. Este módulo permite visualizar los ambientes de una casa y controlar sus dispositivos presionando directamente sobre la pantalla. Como se puede observar en la ***Figura 1***, cuando se presiona sobre un electrodoméstico se accede a un menú con las posibles funciones del dispositivo.
- Conexión al Broker. Permite realizar la comunicación con el broker.

ISBN 978–987–633–574–4

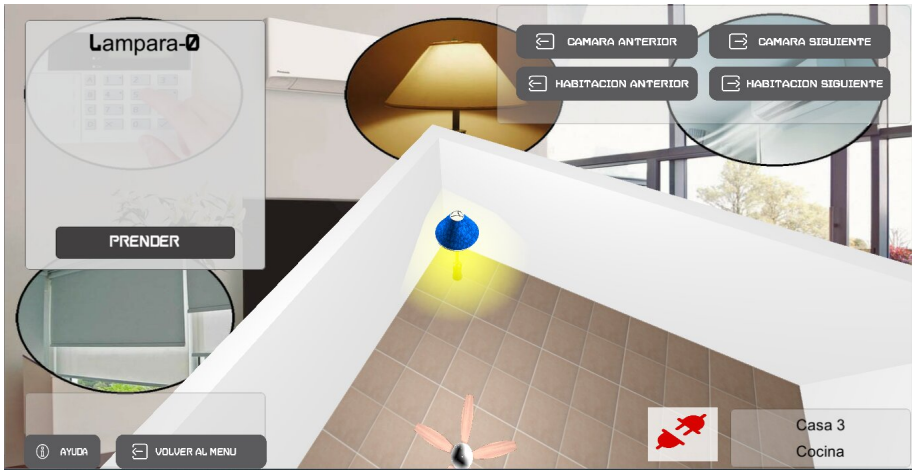


Figura 1. Control de dispositivos

3 Componentes del sistema

Para realizar la comunicación entre dispositivos y sensores, se cuenta con tres componentes principales:

- Componente de interfaz: es el encargado de la interacción con el usuario por medio de una aplicación móvil 3D mencionada en el capítulo 2.
- Componente de comunicación: es el responsable del procesamiento de las solicitudes y mensajes. Dicho procesamiento es realizado por un Broker de MQTT.
- Componente de control: permite llevar a cabo las solicitudes en el mundo físico que provengan del componente de interfaz. Los nodos (microcontroladores NodeMCU [4]) son adaptados a los diferentes dispositivos para proveer las señales necesarias y el funcionamiento deseado de cada uno.

3.1 Protocolo de comunicación

Para realizar la comunicación entre el cliente (desde la aplicación) y los componentes de control se utilizó un procesamiento de solicitudes y mensajes. Dicho procesamiento es realizado por un Broker de MQTT. MQTT es un protocolo de red liviano y simple del tipo publicación-suscripción, permite ser utilizado por dispositivos de recursos limitados y que no dispongan de gran ancho de banda. Este protocolo se monta típicamente sobre TCP/IP [5]. MQTT define dos entidades de red, un servidor, llamado Broker, y un número de clientes conectados a dicho Broker. Los mensajes se organizan por tópicos y funcionan de la siguiente manera: cuando un cliente quiere realizar una publicación, debe definir el tópico al cual será publicada, el Broker se encargará de distribuir el mensaje enviado entre todos los clientes que se hayan suscripto a dicho tópico. Un tópico puede

ISBN 978–987–633–574–4

contener varios subtópicos utilizando "/", por ejemplo "Habitación0/Lámpara-1". Los datos enviados por la aplicación contienen la información necesaria para que el nodo receptor analice y realice la acción correspondiente para modificar los dispositivos físicos. El tópico estará compuesto por RESIDENCIA/HABITACIÓN/DISPOSITIVO, por ejemplo, si se tiene una residencia llamada Juan con una habitación y se quiere realizar el apagado/encendido de una luz, el tópico generado será Juan/Habitación-0/Luz-1. A su vez se tendrá otro tópico el cual se usará para que el microcontrolador (nodo) envíe el estado del dispositivo, permitiendo saber si se ha manipulado de manera externa a la aplicación. Se tendrán, entonces, dos tópicos, uno que envía información de la aplicación al nodo de control, y otro que envía información del estado de un determinado dispositivo, del nodo de control a la aplicación.

4 Comunicación Bidireccional

Bajo la comunicacion bidireccional definida, tanto el cliente como los nodos de control envıan y reciben mensajes. Estos mensajes son almacenados en un servidor el cual los distribuirá a través de la red. La posibilidad de enviar y recibir mensajes permite un mejor control del estado de los dispositivos. Es decir, no solo se tiene el control interno desde la aplicación, sino que también se cuenta con la información de las alteraciones realizadas a los objetos físicos de forma externa, y por lo tanto, es posible realizar las acciones necesarias para que el estado del objeto se visualice de forma correcta en tiempo real.
Los firmwares de los NodeMCU y su depuración fueron realizados por medio de la herramienta Arduino IDE. Este entorno multiplataforma fue desarrollado en java, y se lo publica bajo la Licencia Pública General de GNU, admite lenguajes C y C++, y adicionalmente cuenta con una biblioteca de software, que proporciona un conjunto de procedimientos de E/S [6]. Además, se utilizó un servidor en la nube, Amazon Cloud [7], en el cual se instaló Mosquitto, un agente de mensajes de código abierto que implementa el protocolo MQTT [8].
En la ***Figura 2*** se muestra la bidireccionalidad adquirida, la aplicación publicará un mensaje según el accionar del usuario (Unity), mientras que desde el nodo de control (desarrollado en Arduino) se publicará el estado en el que se encuentra el dispositivo físico. Estos mensajes son distribuidos desde un servidor en la nube (AWS) o en un servidor local.

A cada dispositivo físico le corresponde un nodo de control, es decir, cada nodo controla un único objeto, por lo tanto, cada uno tiene un tópico que lo identifica. Al iniciar la aplicación, se envían al Broker los tópicos pertenecientes a cada objeto. Cada nodo identifica y se suscribe al tópico que le corresponde. Además, para poder recibir el estado de los dispositivos físicos en tiempo real, la aplicación también se suscribe a los tópicos en los cuales los nodos publicarán el estado de cada dispositivo. Esta última funcionalidad permite que la aplicación esté notificada en todo momento sobre lo que sucede en el mundo real, y en base a ello pueda realizar las modificaciones necesarias dentro de la aplicación, brindando información actualizada al usuario.

ISBN 978–987–633–574–4

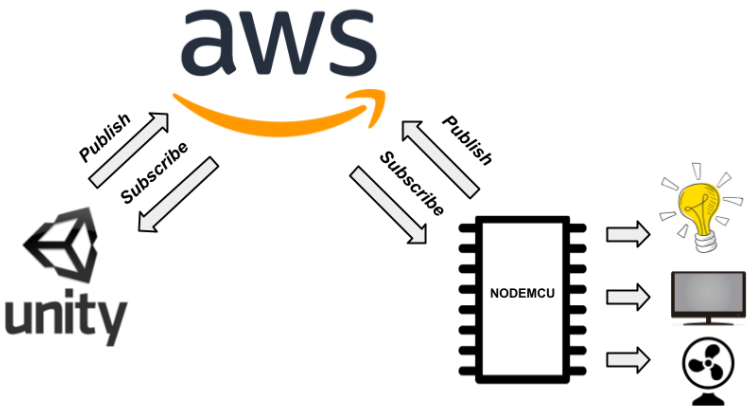


Figura 2. Bidireccionalidad

5 Experimentación

Una aplicación con estas características permite ser utilizada en cualquier ambiente, ya sea un hogar, un edificio, una oficina. Además, la funcionalidad de comunicación de la aplicación puede estar ligada a dos tipos de servidores, local o en la nube, por lo que es posible analizar el comportamiento de cada tipo de servidor en un mismo escenario de prueba. Esto permite tener una visión de cómo se comporta el protocolo de comunicación utilizado, como también la logística de programación a medida que se lo fuerza con solicitudes de mensajes. Como servidor local se utilizó una notebook con sistema operativo Ubuntu, y para el servidor en la nube se utilizó la plataforma Amazon Web Service, que proporciona Amazon EC2, servicio web que brinda capacidad de procesamiento en la nube [9].

Como funcionalidad básica de prueba se optó por realizar el encendido/apagado de una luz. Como se puede observar en la ***Figura 3***, el mensaje inicia en el Cliente (Tablet), se envía al servidor o broker (AWS), y desde alli al Cliente (NodeMCU). Este último responde de dos formas: por un lado envía un mensaje de vuelta indicando el estado de la luz y por otro lado enciende/apaga dicha luz. Aprovechando la comunicación bidireccional, se espera que la tasa de recepción de mensajes por parte del nodo de control y la aplicación, sea la misma que la tasa a la cual se envían los mensajes.

Los mensajes estaban compuestos por dos palabras alternadas, “ON” correspondientes a 2 bytes y “OFF” correspondientes a 3 bytes. Las pruebas tienen un total de envío de 100 mensajes. Se enviaron diferentes cantidades de mensajes en cada segundo. El escenario de prueba se basó en el envío de 1, 2, 4, 5 y 10 mensajes independientemente, cada 1 segundo. Estos mensajes se enviaban al Broker

ISBN 978–987–633–574–4

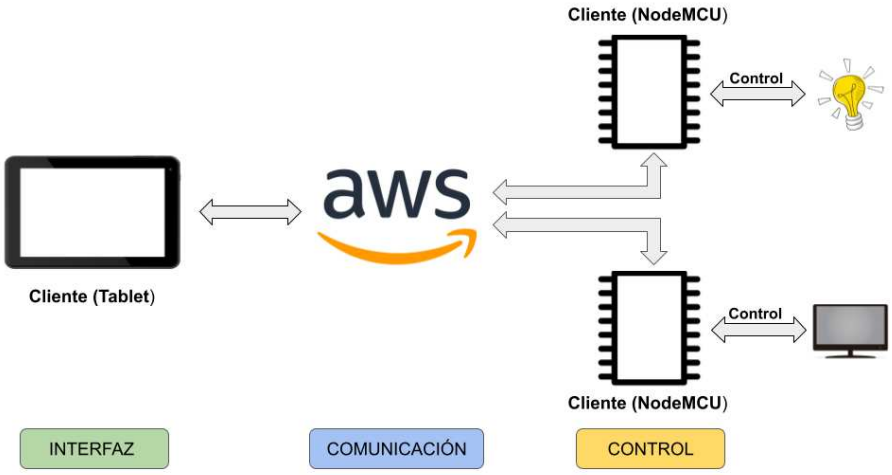


Figura 3. Relación entre los componentes

mediante el tópico correspondiente desde el cliente, es decir, la aplicación móvil. Los mensajes se recibian por el NodeMCU en el cual se registraba el tiempo de llegada de cada mensaje a partir del último recibido. El microcontrolador, analiza el mensaje arribado y responde con otra publicación en un nuevo tópico. La aplicación está suscripta a este tópico, por lo que se procede a tomar el tiempo en el que llega el mensaje del lado del cliente. Asi, se tiene una aproximación del tiempo de envío, como también, del tiempo de respuesta al mensaje arribado al microcontrolador. A partir de la suma de estos dos valores, en milisegundos, es posible obtener el tiempo que tarda el mensaje en llegar y la respuesta del mismo dando como resultado la latencia de la comunicación bidireccional. Con los datos medidos se generaron métricas de rendimiento de comunicación como latencia, throughput, entre otros.

- La latencia se entiende como la suma de tiempos entre el momento en el que un mensaje es publicado y el momento en el que dicho mensaje es recibido por el suscriptor y vuelto a publicar llegando al suscriptor quien es el que había iniciado los envíos de mensajes.
- El throughput se entiende como la tasa de paquetes que se envían a través de un canal de comunicación.

6 Métricas obtenidas

Para una mejor comprensión del tipo de experimentación realizada, a continuación se detallan especificamente los mecanismos de resolución y análisis utilizados en las diferentes etapas de las dos primeras pruebas. Las métricas resultantes se reflejan en graficas que muestran el comportamiento del canal de comunicación

ISBN 978–987–633–574–4

del sistema, observándose los tiempos de arribo de los mensajes a la aplicación móvil 3D y al nodo de control, como también, el número del mensaje enviado.

6.1 Servidor en la nube

Para llevar a cabo las pruebas en un servidor en la nube, se utilizó la plataforma Amazon Web Services con la región ubicada en América del Sur (São Paulo) sa-este-1. Para confirmar la correcta comunicación con el servidor, se realizó un ping desde la dirección IP con la que se realizaron todas las pruebas. El tiempo de respuesta fue de 33 milisegundos y ningún paquete perdido.

1ra Prueba:

Se comenzó con el envío de 1 mensaje cada 1 segundo, con un total de 100 mensajes. Por lo tanto 50 mensajes fueron de 2 bytes y los 50 restantes de 3 bytes, teniendo un envío de 2000 bits en total. El tiempo promedio entre mensajes recibidos en el nodeMCU fue de 1019 milisegundos, en relación con el tiempo promedio en la aplicación que fue de 962 milisegundos. Por lo tanto se tiene una latencia en la bidireccionalidad de 1,981 segundos. Esto indica que los envíos se hicieron de forma normal, ya que como se envían mensajes cada 1 segundo al nodo de control, este recibe cada mensaje cada 1 segundo, y por lo tanto responderá también cada 1 segundo, logrando un tiempo total de aproximadamente de 2 segundos en el envío y respuesta.

En la ***Figura 4*** se puede observar la relación de los intervalos de tiempo entre mensajes en el microcontrolador (Color azul) y la aplicación (Color naranja).

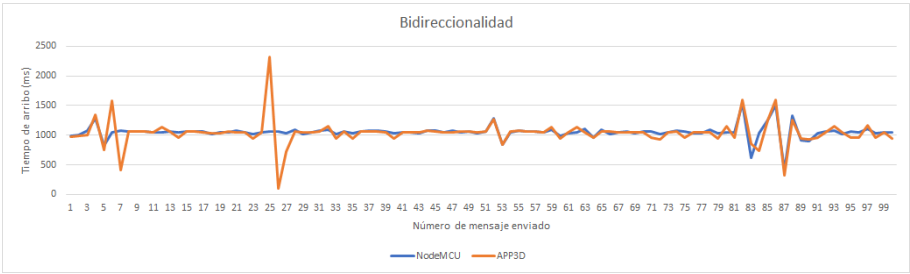


Figura 4. Intervalo de tiempos entre mensaje y mensaje (1ra prueba)

En esta prueba no se observó que se hayan perdido mensajes en ningún momento. Se puede apreciar que hay momentos que del lado de la aplicación el tiempo de arribo de mensajes fue superior a lo normal. Sin embargo, una vez recibido el mensaje, el siguiente mensaje llega al instante. Puede existir una demora en el tiempo de procesamiento o análisis de los mensajes cuando éstos arriban al microcontrolador, provocando que los próximos mensajes a arribar se acumulen en la cola del buffer del microcontrolador. Una vez finalizado el análisis y vuelto a publicar, es posible que los mensajes que se encuentran en el buffer hayan sido analizados más rapidamente, haciendo que su envío sea

ISBN 978–987–633–574–4

casi instantáneo. Lo explicado anteriormente se puede visualizar en la ***Figura 4*** en los intervalos entre los mensajes 24,25 y 25,26. Se destaca que gráficamente, tanto el envío desde la aplicación y la respuesta desde el microcontrolador siguen aproximadamente la misma tasa de respuesta.

2da Prueba:

Se enviaron 2 mensajes cada 1 segundo con un total de 100 mensajes. Del lado del nodo (nodeMCU) se recibieron 100 mensajes, 50 de 2 bytes y el resto de 3 bytes. Mientras que del lado de la aplicación se recibieron 99 mensajes dando por sentado que se perdió 1 mensaje. Cómo se enviaron 2 mensajes por segundo, el tiempo entre mensaje y mensaje fue de 0.5 segundos.

Si se reciben 2 mensajes cada 1 segundo, y se tiene un total de 100 mensajes, el tiempo que debería transcurrir en recibir dichos mensajes es de 50 segundos. En las pruebas se tiene un total de 52 segundos aproximadamente de ambos lados en recibir los mensajes, teniendo un retraso de casi 2 segundos. El tiempo promedio de recepción de mensaje en el nodo fue de 516 ms, mientras que en la aplicación fue de 510 ms.

En la ***Figura 5*** se pueden observar picos que corresponden a lo explicado en la primera prueba.

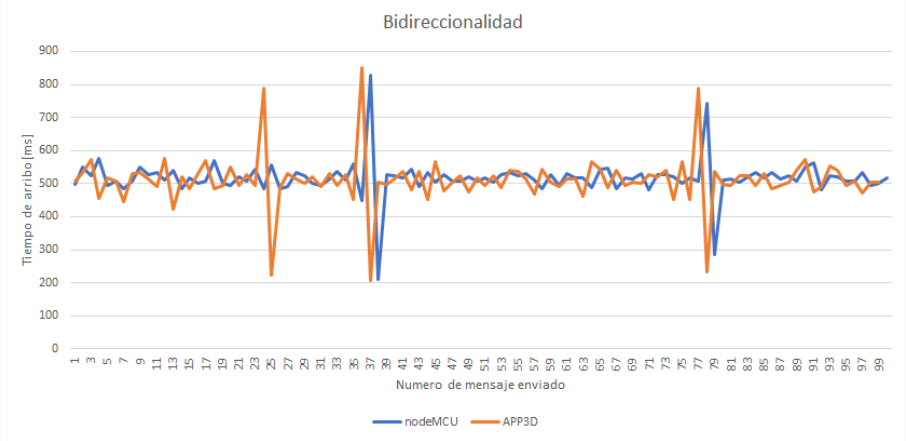


Figura 5. Intervalo de tiempo entre mensajes (2da prueba)

Finalmente, teniendo en cuenta los datos obtenidos se destaca que no se tuvo exigencia en la funcionalidad de la comunicación, pero se perdió un mensaje en el último envió por parte de la aplicación.

6.2 Servidor local

Como se mencionó en el apartado Experimentación, como servidor local se utilizó una notebook con sistema operativo Ubuntu. Dicha notebook se encontraba bajo la misma red local que el microcontrolador nodeMCU o nodo de control.

ISBN 978–987–633–574–4

1er Prueba

La prueba es similar a la realizada en la sección anterior en donde se envía 1 mensaje cada 1 segundo. El tiempo de recepción de mensajes del lado del nodeMCU fue de 1012 ms y del lado de la aplicación de 1005 ms. El tiempo total que llevó el envío de mensajes de la aplicación al nodo fue de 101 segundos. En la ***Figura 6*** se muestran las respuestas de arribo de mensajes.

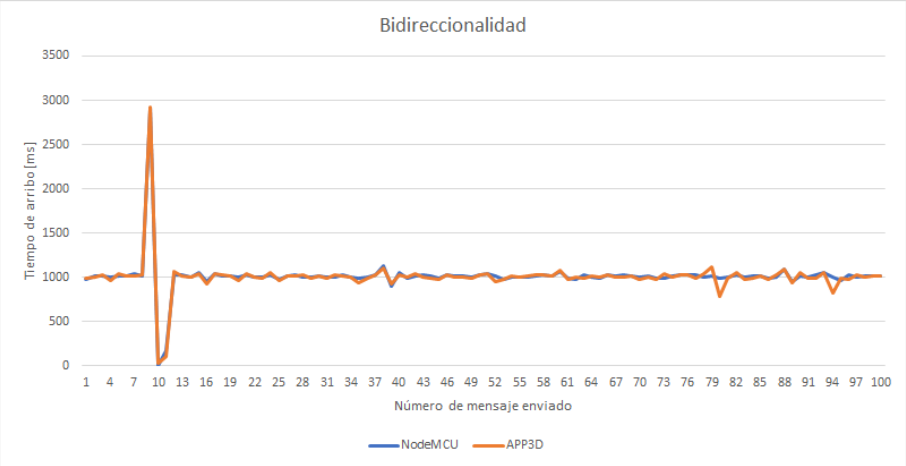


Figura 6. Intervalo de tiempo entre mensaje y mensaje (1ra prueba)

Se puede apreciar claramente que el comportamiento bidireccional de la comunicación es casi exactamente igual, con aproximadamente el mismo tiempo de respuesta, arribaron los 100 mensajes y no se perdió ninguno. El tiempo de arribo en ambos lados es casi una constante.

2da Prueba

Se enviaron 2 mensajes por segundo, es decir, 1 mensaje cada 500 milisegundos. El tiempo promedio de arribo del lado del nodeMCU fue de 525 milisegundos y del lado de la aplicación de 522 milisegundos. En la ***Figura 7*** se puede observar que el tiempo de arribo en ambos lados fue similar en gran parte de la comunicación. Solo se generaron picos de retardo en ciertos instantes, pero no afectaron a la comunicación bidireccional.

7 Resultados

En el ***Cuadro 1*** se resumen los resultados obtenidos en las 5 pruebas realizadas utilizando el servidor en la nube. En cuanto a las pruebas utilizando un servidor local, en el ***Cuadro 2*** se resumen los 6 resultados obtenidos.

ISBN 978–987–633–574–4

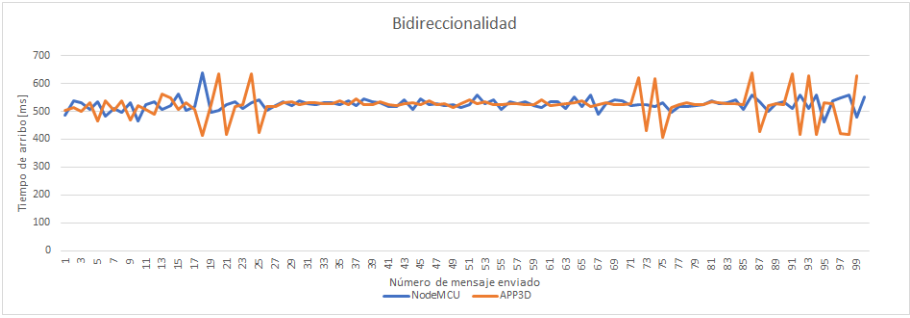


Figura 7. Intervalo de tiempo entre mensaje y mensaje (2da prueba)

Cuadro 1. Resultados utilizando un servidor en la nube.

Número de Prueba	Throughput	Tiempo Promedio Bidireccional [ms]	Número de mensajes perdidos enviados a la aplicación
1	1 msj/seg	2093	0
2	2 msj/seg	1028	1
3	4 msj/seg	595	2
4	5 msj/seg	451	10
5	10 msj/seg	284	25

Cuadro 2. Resultados utilizando un servidor local.

Número de Prueba	Throughput	Tiempo Promedio Bidireccional [ms]	Número de mensajes perdidos enviados a la aplicación
1	1 msj/seg	2017	0
2	2 msj/seg	1047	0
3	4 msj/seg	526	1
4	5 msj/seg	420	1
5	10 msj/seg	232	1
6	1 msj/0,05 seg	159	23

Estas pruebas surgieron de la necesidad de analizar el rendimiento de la comunicación bidireccional desarrollada. Por lo tanto, se procuró obtener métricas de comunicaciones entre la aplicación móvil 3D y un nodo de control. Como conclusión, se destaca que, al ser una comunicación bidireccional los tiempos de envío y recepción de mensajes tuvieron la misma tasa de respuesta, que era lo esperado. La utilización de dos servidores posibilitó ver que tan confiable es establecer una comunicación con una determinada cantidad de mensajes por segundo. Se destaca que un servidor local tiene una muy buena respuesta en todo

ISBN 978–987–633–574–4

tipo de escenarios, comparado con un servidor en la nube, en donde en ciertas pruebas hubo perdida de mensajes. Adicionalmente, considerando que el servidor utilizado se encuentra en Brasil y las pruebas fueron realizadas desde la ciudad de Berisso (Argentina), es posible que los tiempos de respuesta disminuyan en caso de utilizar servidores con mayor cercanía.
Finalmente, la aplicación desarrollada es apta para ser implementada en cualquier entorno. Sin embargo, teniendo en cuenta el análisis realizado, se podría decir que un servidor local sería la elección correcta en el caso de una vivienda. Aunque, el usuario debería contar con conocimientos específicos ante la aparición de algún inconveniente en el servidor. Otra observación es que debido a la situación sanitaria actual (COVID-19) los precios de los servicios en la nube han disminuido considerablemente. Entonces, con la elección de un servidor en la nube se podría desligar al usuario del mantenimiento requerido [10].
Por otro lado, existe la posibilidad de implementar el sistema en un edificio lo que implica la utilización de una gran cantidad de nodos (NodeMCU). Considerando la capacidad de procesamiento que se obtuvo en las pruebas, la utilización de un servidor en la nube podría llevar a la saturación del canal de comunicaciones y la consecuente perdida de paquetes. Para este tipo de instalaciones, un servidor local permitiría una menor exigencia al mismo, una comunicación más fluida y un menor tiempo de respuesta.

8 Conclusiones

En el presente trabajo se ha realizado un análisis e investigación relacionada especificamente a las comunicaciones de un sistema de control domótico a través de aplicaciones móviles.
Se han desarrollado e implementado diversos módulos que corresponden a la aplicación cliente, servidor en la nube y sensores (NodeMCU). Este aporte ha permitido realizar el control del estado de los diferentes dispositivos electrónicos (nodos de control) en tiempo real y visualizarlo en un modelo tridimensional en la aplicación móvil, utilizando una arquitectura en la nube como servidor.
A partir de las pruebas experimentales, con servidores físicos y virtuales, se han obtenido métricas para cuantificar el rendimiento del flujo de mensajes, permitiendo validar y seleccionar posibles configuraciones del sistema.
Como trabajo futuro se propone realizar un análisis del rendimiento de las comunicaciones utilizando otras infraestructuras en la Nube existentes, especificamente orientadas a Edge Computing y 3D.

Referencias

1. Mateus Cruz, Diego Alejandro. Molina Castañeda, Jonathan Esteban. Jaramillo Benavides, Maria Camila, “Domótica, el hogar digital,” 2018. [Online]. Available: https://answersingenesis.org/answers/magazine.Accedido:4-2021
2. MQTT, “Scalagent. benchmark of mqtt servers.” 2015. [Online]. Available: https://mqtt.org/.Accedido:4-2021

3. Unity, “Unity homepage.” [Online]. Available: https://unity.com/.Accedido:4-2021
4. NodeMCU, “Datasheet nodemcu.” [Online]. Available: https://www.esploradores.com/datasheet-nodemcu/.Accedido:4-2021
5. R. A. Light, “Mosquitto: server and client implementation of the mqtt protocol,” *Journal of Open Source Software, 2(13), 265*, 2017.
6. N. Zlatanov, “Arduino and open source computer hardware and software,” *IEEE Computer Society*, 2005.
7. AWS, “Página oficial amazon web services.” [Online]. Available: https://aws.amazon.com/es/.Accedido:4-2021
8. MQTT, “Mqtt homepage.” [Online]. Available: https://mosquitto.org/.Accedido:4-2021
9. AWS, “Amazon ec2. capacidad informática segura y de tamaño ajustable que admite prácticamente cualquier carga de trabajo.” [Online]. Available: https://aws.amazon.com/es/ec2/?ec2-whats-new.sort-by=item.additionalFields.postDateTime&ec2-whats-new.sort-order=desc.Accedido:4-2021
10. Telesemana, “Telesemana.com. la pandemia impulsa el consumo de la nube en el tercer trimestre de este año,” Diciembre 2020. [Online]. Available: https://www.telesemana.com/blog/2020/10/30/la-pandemia-impulsa-el-consumo-de-la-nube-en-el-tercer-trimestre-de-este-ano/.Accedido:4-2021

ISBN 978–987–633–574–4

Implementación Técnica de una Arquitectura Orientada a Integrar Conocimiento Externo Heterogéneo en Motor de Reglas.

Marcos Maciel[1], Claudia Pons[2]
[1] CAETI UAI, Buenos Aires, Argentina
Mmaciel03@hotmail.com
[2] Universidad Nacional de La Plata, UAI, Buenos Aires, Argentina
Claudia.pons@uai.edu.ar

Abstract. En un contexto de negocios globalizado donde la completitud de la información es la suma de varias partes, resolver problemas se convierte en una tarea que involucra tiempo, análisis y experiencia. Una organización ve limitado su ámbito de acción porque necesita información de terceros para evaluar en forma íntegra y completa una colección de datos. Para superar estos problemas se propone implementar un motor de reglas capaz de interactuar mediante reglas con servicios usando Json como mensajería de intercambio de datos. El modelo propuesto mejora la capacidad de conocimiento al compartir información entre sistemas heterogéneos usando los estándares de la comunidad para resolver problemas complejos.

Keywords: Motor de reglas, DSS, Lógica Simbólica, Arquitectura Orientada a Servicio, SOA.

1 Introducción

Un motor de reglas es una herramienta de software con reglas que evalúan de forma encadenada bloques de conocimiento y experiencia configurados por un experto en la materia. Cada regla es una porción de información expresada: **Si** condición **Entonces** conclusión.

Estos sistemas ayudan a las personas con la tarea de recolectar información para unificar conocimiento (Peña, 2006), analizar relaciones entre los datos y emitir resultados con el objetivo de apoyar la toma de decisión. Son menos propensos a cometer errores porque razonan como un humano lo haría (Agarwal, 2014), resuelven cálculos simples y complejos en un tiempo inferior comparado con una persona (Palma & Marín, 2008), después de ser configurado también puede ser usado por personas no expertas.

Algunos sistemas expertos desarrollados recientemente implementan la inteligencia con diversos métodos, por ejemplo, para diagnosticar miocardiopatía dilatada (Bahrami et al, 2014) usa un árbol de decisión de 46 reglas con respuestas del tipo si/no integrando lógica y conocimiento con el lenguaje de programación Clips (Clips, s.f). Con 16 reglas programadas y un árbol de decisión este sistema experto desarrollado en Php y Mysql genera un diagnostico sugerido sobre 16 tipos de enfermedades oftalmológica (Munaiseche et al, 2018).

En el campo de la telemedicina Massaro et al, desarrollaron un sistema para soportar decisiones de multiniveles apoyado en inteligencia artificial con una arquitectura no orientada a servicios.

En Prado et al, se propone un sistema experto para apoyar la enseñanza en combinación lineal de vectores usando Prolog como lenguaje de programación.

En relación con el aprendizaje Mohd et al, propone un sistema de apoyo de decisión basado en la predicción del estilo de aprendizaje de estudiantes, usando como input la interacción usuario-sistema a través del comportamiento corporal. El diseño técnico de la solución responde a un motor de inferencia con una base de datos para centralizar el conocimiento. En la industria de fabricación de línea blanca se propone la recolección de datos desde sensores IoT (Internet of Things) y la ejecución de un DSS (Decision Support System) desarrollado en Minitab® con el objetivo de incrementar la productividad y eliminar el stock (Shady & Eltawil, 2018). Una arquitectura orientada a servicios permite a una herramienta CIG (Computer-Interpretable Guidelines) utilizar múltiples fuentes y formato de datos, pero delega por completo el razonamiento a una aplicación monolítica desarrollado en Prolog (Chapman & Curcin, 2019).

Berona et al, propone un DSS para predecir la calidad ecológica necesaria para el cultivo de microalgas en exterior, para ello usan una arquitectura monolítica de machine learning y algoritmos en Python con resultados expuestos en la web. Relacionado a medio ambiente (Yu et al,.2020) presento una arquitectura con GIS (sistema de información geográfico) para monitorear y predecir la calidad del agua en pozos privados, aunque hace uso de servicios para extender el sistema implementado este se encuentra acoplado a un software propietario. Para el ámbito de la agricultura se desarrolló un DSS para asistir a los granjeros con información basada en el clima y envió de SMS (Short Message Service) como mecanismos de comunicación (Soyemi & Adesola, 2018), el prototipo usa una arquitectura tradicional de sistema experto con una base de datos y un módulo de predicción. Thaker & Nagori analizo las funciones usadas en sistemas expertos que utilizan lógica difusa en sistemas de recomendaciones tradicionales con base de conocimiento centralizada. En la industria 4.0 Cheng et al, propone una arquitectura basada en la nube donde combina IoT con árboles de decisión para mejorar la producción de prótesis dentales. El uso de DSS para el desarrollo de smart cities elaborado en Bartolozzi et al, usa una arquitectura de consulta a múltiples repositorios con información divida por área de interés.

Para un ser humano tomar decisiones en base al conocimiento experto de un dominio es una tarea que involucra tiempo, análisis y experiencia. Decidir conlleva asociado un proceso cognitivo divido en etapas: invertir tiempo para recolectar grandes volúmenes de información, depurar y/o clasificar mediante sesgos y heurísticas, determinar la probabilidad de ocurrencia y un porcentaje de exactitud sobre cada posible respuesta encontrada (Rampello, 2019), por último, asegurar el respaldo de la información procesada en soporte digital etc. Todas estas tareas pueden resultar difíciles de gestionar debido a que, con la globalización las personas cada día generan más contenido y el volumen de información se torna incontrolable. Los motores de reglas son una solución ideal para gestionar y almacenar el conocimiento.

En este trabajo se propone desarrollar el diseño de la arquitectura propuesta en (Maciel, 2020), integrando conocimiento mediante la combinación de regla ejecutadas en código tradicional y regla ejecutadas en servicios Api Rest (propietarios o externos) para dar soporte a una variedad de funcionalidades de negocios desplegadas en la nube e independiente unas con otras. El objetivo es extender las fronteras del conocimiento incorporando una variada gama de sistemas heterogéneos en la ejecución de reglas, establecer relaciones de negocios globales, compartir

conocimiento, reusar desarrollos tecnológicos propios y de terceros y optimizar los procesos de validaciones, entre otras.

Este artículo está organizado de la siguiente forma: integración de conocimiento mediante Json son detallados en la sección II, implementación técnica es presentada en la sección III, testing y evolución es descripta en la sección IV, y por último se hallan conclusiones y trabajos futuros en la sección V.

2 Integración de conocimiento mediante Json.

Json (JavaScript Object Notation) (JSON, s.f). es una representación lógica, organizada y de fácil lectura que tienen como finalidad intercambiar datos entre sistemas sin importar el software subyacente que lo creo. Los valores o datos están contenidos por atributos que representan un modelo completo o parcial de un negocio, por ejemplo, {"first_name": "Jhon"} donde {atributo: valor}. Es posible leer jerárquicamente su estructura para recuperar el par atributo-valor como padres e hijos usando las llaves ({}). Los servicios SOA (por sus siglas en ingles Service-Oriented Architectures) permite aprovechar las ventajas de este formato porque comparten contrato formal, son débilmente acoplados, abstractos, sin estado y reusables (Rosado Gomez & Jaimes Fernández, 2018). Las características antes mencionadas hacen del formato de intercambio Json y de los servicios un medio indicado para unir sistemas heterogéneos sin dependencia del lenguaje de programación, de una representación exacta de objetos de negocios, del medio de almacenamiento que soporta los objetos de negocios (base de datos, xml, archivos, etc.), del tipo de base de datos (relacional o no relacional), etc. Para integrar conocimiento de distintas fuentes de datos sin las preocupaciones de la tecnología que los soportan se propone un sistema experto capaz de llamar a servicios usando Json como medio para el intercambio de información con los siguientes pasos:

2.1 Configuración de entidades y tipo de datos a partir de un Json no propietario.

A continuación, se presenta un mensaje Json recuperado de una plataforma de pagos, primero se configura al sistema experto mapeando las entidades y atributos como level-entity donde cada entity tiene un tipo de datos asociado. El sistema tiene preconfigurado una lista de funciones o hechos que se asignan dependiendo del tipo de datos (Maciel, 2020), por ejemplo "area_code" se asocia al tipo de datos numérico que tiene preconfigurado algunas funciones matemáticas como por ejemplo sumas, restas o comparaciones del tipo $x > 0 < x$. Configurados los pasos previos se pueden construir las reglas desde un mensaje u objeto Json de terceros.

```
curl -X POST \ 'https://api.client.com/v1/customers' \
  -H 'Authorization: Bearer ACCESS_TOKEN_ENV' \ -d '
"customer": {"email": "jhon@doe.com", "first_name": "Jhon", "last_name": "Doe",
 "phone": {"area_code": "55", "number": "991234567"},
 "identification": { "type": "CUIT", "number": "12345678900" }, "default_address": "Home",
 "address": {"id": "123123", "zip_code": "01234567", "street_name": "Av Corrientes",
"street_number": "123"},
```

"date_registered": "2000-01-18", "description": "Description user", "default_card": "None",
"file": "JVBERi0xLjQKJeLjz9MKMiAwIG…"}

Table 1. Lista de atributos recuperado de un mensaje json.

Atributo	Date	Number	Identity	Invoice	String
email					x
first_name					x
last_name					x
phone.area_code		x			
phone.number		x			
identification.type			x		
identification.number		x	x		
default_address					x
address.id		x			
address.zip_code					x
address.street_name					x
address.street_number		x			
date_registered	x				
description					x
default_card					x
file				x	

La tabla 1 contiene la categorización de cada valor del atributo(entity) con un tipo de datos donde date, number y string son los tipos básicos, por otro lado, Identity e Invoice son tipos de datos agrupadores de atributos por ej. {"identification": {"type": "CUIT", "number": "12345678900"}} == Identity. Este tipo de datos propietario del motor de reglas permite asignar una función que valida un conjunto de datos como una unidad o si es necesario llamar a un servicio externo componer el input. Las funciones son porciones de código capaces de ejecutar lógica simbólica mediante programación tradicional, expresiones regulares, matemáticas o algebraicas o llamar a servicios externos al motor de reglas.

Por ejemplo, el tipo de datos Invoice del atributo file permite ejecutar una regla intermedia, invocando a un endpoint cuyo input es un string en base64(factura afip pdf) y usando un procedimiento OCR (Optical Character Recognition) retorna un json que el motor de reglas inyecta al mensaje principal según:

"customer": {
"invoice": {"category": "CAE", "identity_type": "80", "identity_number": "20000000001",
"sales_point": "1", "type": "A", "number": "00141787", "date": "20101014",
"amount": "300.8", "authorization_code": "60428000005029", "receptor_identity_type": "80",
"receptor_identity_number": "300000000007"}}

Después de la configuración de un mensaje Json a un modelo de negocios el motor de reglas esta disponible para crear paquetes de reglas.

2.2 Desarrollo de un paquete de reglas a partir de un objeto de negocios.

A partir de cada par entidad-atributo(level-entity) y con el tipo de dato asociado se pueden crear las reglas de negocios. Una función tiene como objetivo validar una condición de verdad mediante rutinas matemáticas, expresiones de comparación booleanas, validaciones del tipo expresiones regulares o llamadas a servicios.

Lista de reglas configuradas en base al json de trabajo – Paquete de regla:

Regla 1: first name is not null
Regla 2: last name is not null
Regla 3: phone.area_code is not null Λ phone.area_code is number Λ phone.area_code == 2 dígitos
Regla 4: phone.number is not null Λ phone.number is number Λ phone.number > 7 dígitos Λ phone.number < 12
Regla 5: email is not null Λ email is valid (api reg exp)
Regla 6: trusth email domain (api externa)
Regla 7: email not in black list (api interna)
Regla 8: address.zip_code is not null Λ address.zip_code is number Λ address.street_name is not null Λ address.street_number is not null Λ address.street_number is number
Regla 9: phone.area_code belongs to address(api interna)
Regla 10: Afip invoice (api interna)
Regla 11: category is not null Λ category between CAE-CAI-CAEA
Λ identity_number is not null Λ identity_number is Identity Λ sales_point is not null Λ sales_point is number Λ type is not null Λ type between A-B-C Λ number is not null number is number Λ date is not null Λ date is date Λ amount is not null Λ amount > 0 Λ authorization_code is not null Λ authorization_code is number Λ receptor_document_type is not null Λ receptor_document_type is number Λ receptor_identity_number is not null Λ receptor_identity_number is Cuit
Regla 12: AFIP invoice valid (SOA Service externo)
Regla 13: IP not in Blacklist
Regla n: …

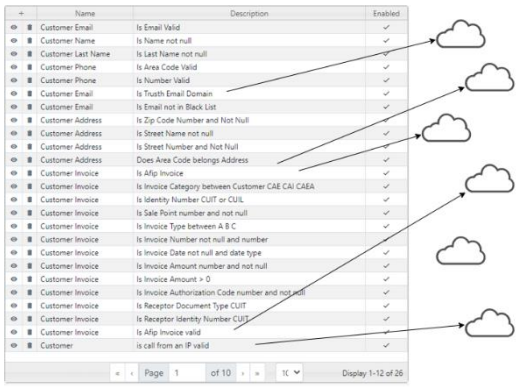


Fig. 2. Paquete de reglas: Prevencion de Fraude.

Cada paquete de reglas se puede configurar en forma independiente a otros paquetes, reutilizando los tipos de datos y las funciones parametrizadas la Fig. 2 resume las reglas construidas para el paquete llamado Prevención. La lista contiene reglas que se ejecutan en código del motor de reglas para validaciones base y otras reglas que consumen servicios señaladas con flechas a la nube. Este paquete de reglas demuestra la flexibilidad y creatividad para construir validaciones robustas que involucren no solo código y datos propietarios sino código y datos externos.

3 Implementación Técnica.

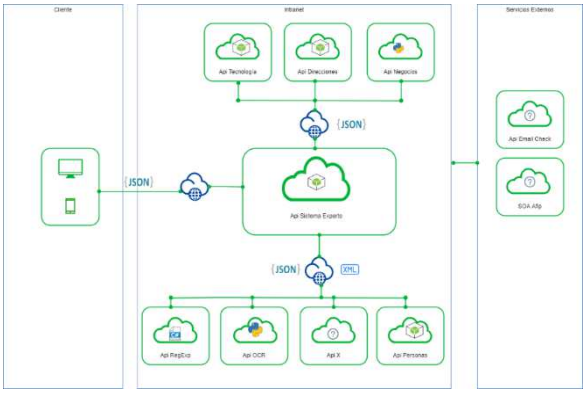


Fig. 3. Arquitectura del motor de reglas.

La Fig. 3 expone la comunicación entre el motor de reglas (Maciel, 2020) con servicios propietarios hosteados tanto en intranet y otros no propietarios hosteados en internet. Un cliente envía un mensaje json al sistema, el motor recupera el paquete de reglas correspondiente y ejecuta regla a regla según su definición Fig. 4. En los casos de invocaciones hacia endpoint no propietario se dispara una llamada a un servicio de intranet cuya responsabilidad es interactuar con el servicio externo.

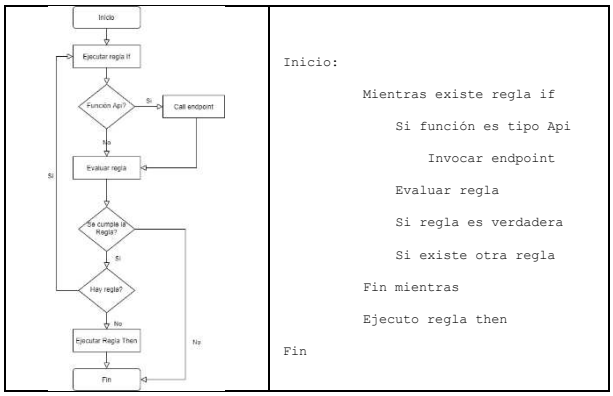


Fig. 3. Diagrama de flujo de datos del procesamiento de paquetes de reglas.

Disgregación de reglas – endpoint:

Regla: Formato de Email
Entidad: email {customer.email}
parámetros: {customer.email: mmaciel03@hotmail.com }
Función: is email format: curl --location
--request GET 'https://{domain}:105/emailformat/{parámetros}'
Tipo servicio: Api Regular Expression

Respuesta: {"email": "mmaciel03@hotmail.com","isok": "true","type": "hotmail.com","user": "mmaciel"}

Cuando se ejecuta esta regla la función extrae el valor del customer.email y pasa el parámetro al servicio validando la respuesta con status 200 y el valor isok==true.

Regla: Dominio de email confiable.
Entidad: email {customer.email}
parámetros: {customer.email : mmaciel03@hotmail.com }
Función: is trusth email domain: curl --location
--request GET 'https://{domain}:106/esdominioemailconfiable/{parámetros}'
Invocación a Api externo:
curl --location --request GET 'https://mailcheck.p.rapidapi.com?domain=mmaciel03@hotmail.com' \
--header 'x-rapidapi-key: xxx' \ --header 'x-rapidapi-host: mailcheck.p.rapidapi.com'

Respuesta: {"email": "mmaciel03@hotmail.com","isok": "true","type": "hotmail.com","user": "mmaciel"}
Tipo servicio: Api Business

Cuando se ejecuta esta regla la función extrae el valor del {customer.email} y pasa el parámetro al servicio validando la respuesta con status 200 y el valor isok==true. En este ejemplo además existe una llamada a un servicio externo y sobre la respuesta se crea un nuevo objeto.

Regla: Factura Afip
Entidad: Factura {customer.file}
parámetros: "JVBERi0xLjQKJeLjz9MKMiAwIG…"
Función: is AFIP invoice: curl --location --request POST 'https://{domain}:108/getafipinvoice /' \
--header 'Content-Type: application/json' \ --data-raw '{"data": {parámetros}'
Tipo servicio: Api OCR

Respuesta: "invoice": {"category": "CAE", "identity_number": "20000000001","sales_point": "1",
"type": "A", "number": "00141787", "date": "20101014", "amount": "300.8",
"authorization_code": "60428000005029", "receptor_document_type": "80",
"receptor_identity_number": "300000000007","isok": "true"}

Esta regla la función extrae el valor del customer. file y pasa el parámetro al servicio validando la respuesta con status 200 y el valor isok==true, adicionalmente inyecta un objeto nuevo para usar en la siguiente regla.

Regla 12: factura aprobada x afip

Entidad: Factura {customer.invoice}
parámetros: {customer.invoice}
Función: is AFIP invoice valid: curl --location
--request POST 'https://{domain}:106/checkafipinvoice /' \
--header 'Content-Type: application/json' \ --data-raw '{customer.invoice}'
Invocación a servicio externo
https://wswhomo.afip.gov.ar/WSCDC/service.asmx?op=ComprobanteConstatar
Tipo servicio: Api Business
Respuesta: {"response": "A", "isok": "true"}

En este último, caso la regla usa información generada por una regla anterior y llama a un servicio que a su vez se comunica con un ente gubernamental para validar información que originalmente llego en base64, logrando un procesamiento más completo.

4 Testing y evaluación.

Se realizaron pruebas de performance, tiempo de respuesta y kb de datos enviados y recibidos, el motor de reglas propuesto actúa como evaluador de datos entre un sistema cliente que envía json y un sistema backend responsable del negocio por este motivo el tiempo incurrido para ejecutar un paquete de reglas no debe ser considerable.

Table 3. Testing de performance.

Rules	Time avg	Min	Max	Std. Dev.	Throughput	Received KB/sec	Sent KB/sec	Avg. Bytes
Rule 1- Is Email Valid	528	146	960	232.44	5.73	3.13	0.83	560
Rule 2 - Is Name not null	538	95	961	275	5.59284	2.85	0.75	521
Rule 3 - Is Last Name not null	583	138	949	278.97	5.11038	2.6	0.68	521
Rule 4 - Is Area Code Valid	623	149	982	288.51	4.75105	2.38	0.62	512
Rule 5 - Is Number Valid	660	158	990	282.93	4.40141	2.3	0.61	536
Rule 6 - Is Trusth Email Domain	705	185	1074	288.98	4.09601	2.25	0.6	563
Rule 7 - Is Zip Code Number and Not Null	743	166	1132	289.32	3.80228	1.89	0.49	509
Rule 8 - Is Street Name not null	787	223	1171	286.01	3.58526	1.93	0.51	551
Rule 9 - Is Street Number and Not Null	798	266	1169	264.44	3.43265	1.7	0.44	506
Rule 10 - Does Area Code belongs Address	806	298	1162	241.78	3.32094	1.71	0.77	527
Rule 11 - Is Afip Invoice	801	317	1169	222.36	3.23792	1.57	61.73	497
Rule 12 - Is Invoice Category between Customer CAE CAI CAEA	799	398	1180	205.01	3.16016	1.67	0.44	542
Rule 13 - Is Identity Number CUIT or CUIL	780	386	1164	194.59	3.11798	1.68	0.45	551
Rule 14 - Is Sale Point number and not null	777	440	1166	183.56	3.0792	1.51	0.39	503
Rule 15 - Is Invoice Type between A B C	762	535	1163	166.63	3.04841	1.49	0.39	500
Rule 16 - Is Invoice Number not null and number	756	565	1148	160.04	3.03656	1.58	0.42	533
Rule 17 - Is Invoice Date not null and date type	739	528	1125	161.49	3.02572	1.52	0.4	515
Rule 18 - Is Invoice Amount number and not null	734	541	1181	151.26	3.03785	1.53	0.4	515
Rule 19 - Is Invoice Amount > 0	718	451	1162	173.31	3.08318	1.6	0.42	530
Rule 20 - Is Invoice Authorization Code number and not null	690	385	1113	186.9	3.16236	1.63	0.43	527
Rule 21 - Is Receptor Document Type CUIT	650	280	1127	214.69	3.28299	1.64	0.43	512
Rule 22 - Is Receptor Identity Number CUIT	611	189	1164	239.21	3.45829	1.85	0.49	548

Rule 23 - Is Afip Invoice valid		1954	1098	2962	468.94	3.21481	1.35	1.66	429
Rule 24 - Is call from an IP valid		396	18	1183	370.88	3.70178	1.94	0.51	536
Rule 25 - Is Email not in Black List		351	18	1117	346.89	4.01671	2.23	0.59	569
	Total	732	18	2962	378.49	48.76907	24.98	44.51	524.5

La tabla 3 resume las métricas de la prueba de performance sobre 50 ejecuciones concurrentes evaluado con la herramienta Apache JMeter (JMeter, sf). Average es el tiempo promedio tomado por las 50 ejecuciones, el Min es el tiempo más corto ocupado y el Max el tiempo más largo, Std Dev es la desviación estándar también sobre el tiempo, Throughput numero de request procesados por unidad de tiempo a mayor valor mejor resultado. Por último, Received KB/sec, Sent KB/sec, Avg. Bytes informa el tamaño del mensaje intercambiado ida y vuelta. A priori el tiempo de respuesta se mantiene por abajo del segundo excepto el caso de dependencia con un servicio externo.

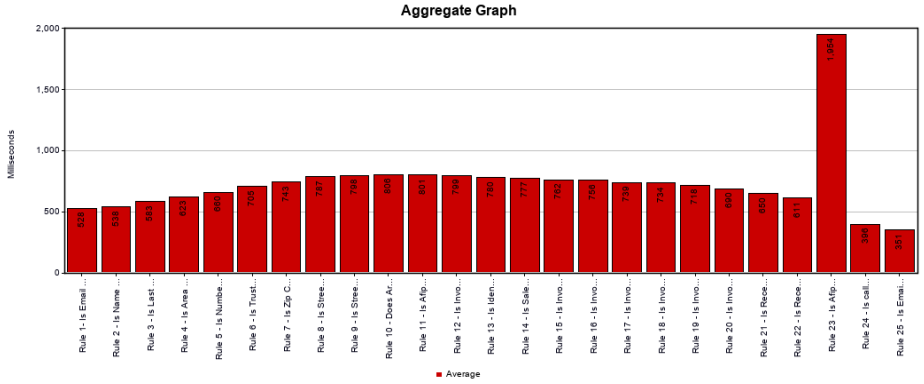


Fig. 4. Gráfico de tiempo medio de resolución de paquete de regla.

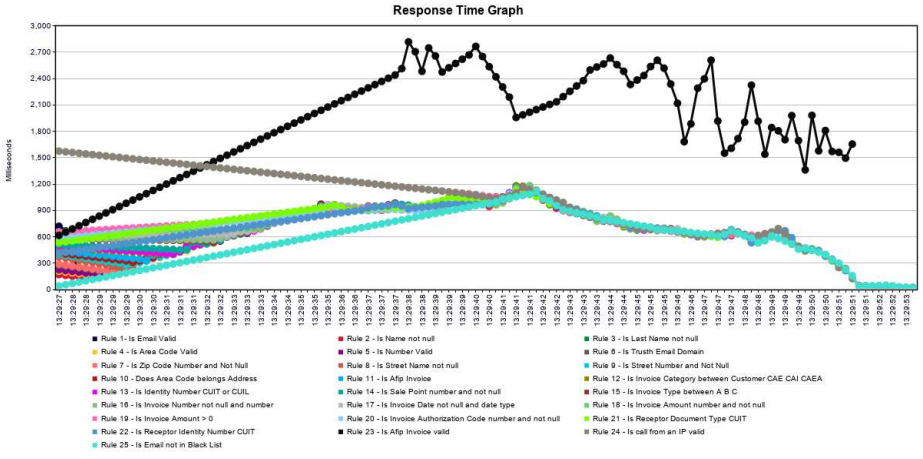


Fig. 5. Gráfico del tiempo de respuesta por cada ejecución de regla.

En la Fig. 4 resume el tiempo medio de ejecución de cada regla para una prueba de 50 corridas y la Fig. 5 gráfica el tiempo individual de la regla todo medido en ms.

Ambiente de pruebas: OS Name Microsoft Windows Server 2012 R2 Standard
Version 6.3.9600 Build 9600 System Type x64-based PC
Processor Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz, 2195 Mhz, 6 Core(s), 6 Logical Processor(s)
Installed Physical Memory (RAM) 38.0 GB Drive C: File System NTFS
Size 79.48 GB (85,343,596,544 bytes) Free Space 22.40 GB (24,051,150,848 bytes)

5 Conclusión y trabajos futuros

En un contexto globalizado donde analizar información ya no depende exclusivamente de datos propios, esta propuesta está orientada a acortar la distancia entre sistemas extendiendo las fronteras del ámbito de la información y la tecnología, hacienda uso de servicios y mensajería para el intercambio de información. El motor de reglas permite configurar un mix de reglas tanto para validar datos de forma tradicional como para recuperar nueva información proveniente de fuentes externas y reutilizarla en una instancia posterior mediante otras reglas. Esta flexibilidad para construir validaciones de negocios potencia el acceso a datos externos, mejora el acceso a soluciones externas independientemente de la tecnología subyacente, permite reutilizar conocimiento, desarrollar reglas de negocios más robustas e incrementar la certidumbre de los resultados optimizando la toma de decisiones entre otras mejoras. Como trabajo futuro se puede mencionar la incorporación de lógica no simbólica para soportar algoritmos de machine learning, y la implementación de microservicios con un service bus para mejorar la experiencia de acceso a servicios.

References

Agarwal, M. & Goel, S. (2014) Expert System and it's Requirement Engineering Process. International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), Jaipur, 2014, pp. 1-4. 2015 de http://ieeexplore.ieee.org [Acceso Junio 2021]

Bahrami, A., Roozitalab, N., Jafari, S. and Bahrami, A. 2014. An expert system for diagnosing dilated cardiomyopathy. International Journal of Engineering Science Invention. 3,3 (March, 2014), 38--42.

Bartolozzi, Marco & Bellini, Pierfrancesco & Nesi, Paolo & Pantaleo, Gianni & Santi, Luca. (2015). A Smart Decision Support System for Smart City. 10.1109/SmartCity.2015.57.

Berona, Elyzer & Buntag, Daibey & Tan, Mary Jane & Coronado, Armin. (2016). Web-Based Decision Support System for Water Quality Monitoring and Prediction for Outdoor Microalgae Cultivation. IOSR Journal of Computer Engineering. 18. 2278-661. 10.9790/0661-1803061620.

Chapman, M. D., & Curcin, V. (2019). A Microservice Architecture for the Design of Computer-Interpretable Guideline Processing Tools. In 18th IEEE International Conference on Smart Technologies IEEE Computer Society Press. https://doi.org/10.1109/EUROCON.2019.8861830

Cheng, Yu-Jie & Chen, Ming-Huang & Cheng, Fu-Chi & Cheng, Yu-Chi & Lin, Yu-Sheng & Yang, Cheng-Jung. (2018). Developing a Decision Support System (DSS) for a Dental Manufacturing Production Line Based on Data Mining. Applied System Innovation. 1. 17. 10.3390/asi1020017.

Clips C Language Integrated Production System http://www.clipsrules.net/ [Acceso Junio 2021].

JMeter. Apache JMeter™ https://jmeter.apache.org/ [Acceso Agosto 2021]

JSON (JavaScript Object Notation - Notación de Objetos de JavaScript) https://www.json.org/json-es.html [Acceso Agosto 2021]

Maciel, Marcos (2020). Motor de Reglas desacoplado orientado a formato JavaScript Object Notation. XXVI Congreso Argentino de Ciencias de la Computación, CACIC Buenos Aires, Argentina. 489-498 ISBN: 0201633612

Massaro, Alessandro & Galiano, Angelo & Scarafile, Domenico & Vacca, Angelo & Frassanito, Antonella & Melaccio, Assunta & Solimando, Antonio & Ria, Roberto & Calamita, Giuseppe & Bonomo, Michela & Vacca, Francesca & Gallone, Anna & Attivissimo, F.. (2020). Telemedicine DSS-AI Multi Level Platform for Monoclonal Gammopathy Assistance. 1-5. 10.1109/MeMeA49120.2020.9137224.

Mohd, Fatihah & Yahya, Wan & Ismail, Suryani & Jalil, Masita & Maizura, Noor. (2019). An Architecture of Decision Support System for Visual-Auditory-Kinesthetic (VAK) Learning Styles Detection Through Behavioral Modelling. International Journal of Innovation in Enterprise System. 3. 24-30. 10.25124/ijies.v3i02.37.

Munaiseche, Cindy & Kaparang, Daniel & Rompas, Parabelem. (2018). An Expert System for Diagnosing Eye Diseases using Forward Chaining Method. IOP Conference Series: Materials Science and Engineering. 306. 012023. 10.1088/1757-899X/306/1/012023.

Palma J., Marín R (2008). Inteligencia Artificial: Métodos, técnicas y aplicaciones, pp. 83-97 Madrid: McGraw-Hill.

Peña, A.A.: Sistemas basados en conocimiento: Una Base para su Concepción y Desarrollo. Instituto Politécnico Nacional México (2006)

Power, Daniel J., "Decision Support Systems: Concepts and Resources for Managers" (2002). Faculty Book Gallery. 67.

Prado, C. S., León, A. D. L. C. C., & Martín, T. R. T. (2020). AUTOAPRENDIZAJE SOBRE COMBINACIÓN LINEAL DE VECTORES UTILIZANDO UN SISTEMA EXPERTO. Revista Tecnología Educativa, 5(2).

Rampello, S. (2019). “Los sesgos en la toma de decisiones”. Revista Perspectivas de las Ciencias Económicas y Jurídicas, Vol. 9, N° 1 (enero-junio). Santa Rosa: FCEyJ (UNLPam); EdUNLPam; ISSN 2250-4087, e-ISSN 2445-8566. DOI http://dx.doi.org/10.19137/perspectivas-2019-v9n1a06

Rosado Gomez, Alveiro & Fernández, Juan. (2018). REVISIÓN DE LA INCORPORACIÓN DE LA ARQUITECTURA ORIENTADA A SERVICIOS EN LAS ORGANIZACIONES.. REVISTA COLOMBIANA DE TECNOLOGIAS DE AVANZADA (RCTA). 1. 10.24054/16927257.v31.n31.2018.2769.

Shady Salama, Amr B. Eltawil, A Decision Support System Architecture Based on Simulation Optimization for Cyber-Physical Systems, Procedia Manufacturing, Volume 26, 2018, Pages 1147-1158, ISSN 2351-9789, https://doi.org/10.1016/j.promfg.2018.07.151.

Soyemi, Jumoke & Adesola, Adesi. (2018). A Web-based Decision Support System with SMS-based Technology for Agricultural Information and Weather Forecasting. International Journal of Computer Applications. 180. 1-6. 10.5120/ijca2018916338.

Thaker, Shaily & Nagori, Viral. (2018). Analysis of Fuzzification Process in Fuzzy Expert System. Procedia Computer Science. 132. 1308-1316. 10.1016/j.procs.2018.05.047.

Yu Lan , Wenwu Tang , Samantha Dye & Eric Delmelle (2020) A web-based spatial decision support system for monitoring the risk of water contamination in private wells, Annals of GIS, 26:3, 293-309, DOI: 10.1080/19475683.2020.1798508

ISBN 978–987–633–574–4

Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia.

Pablo Soligo, Germán Merkel, and Jorge Ierache

Universidad Nacional de La Matanza,
Florencio Varela 1903 (B1754JEC) San Justo, Buenos Aires, Argentina
{psoligo,jierache}@unlam.edu.ar
{gmerkel}@alumno.unlam.edu.ar
http://unlam.edu.ar

Abstract. En este trabajo exploramos la posibilidad de encontrar anomalías automáticamente en telemetría satelital real. Comparamos dos técnicas de aprendizaje automático diferentes como alternativa al control de límites clásico. Intentamos evitar, en la medida de lo posible, la intervención de un experto, detectando anomalías que no se pueden encontrar con los métodos clásicos o que se desconocen de antemano. La mezcla gaussiana y Rolling Means se aplican en la telemetría del subsistema de potencia de un satélite órbita baja. Algunos valores de telemetría se modificaron artificialmente para generar un apagado en un panel solar para intentar lograr una detección temprana por contexto o por comparación. Finalmente, se presentan los resultados y la conclusión.

Keywords: Satellites, Ground Segment, Platform, Telemetry, Machine Learning, Data Mining, Anomaly Detection

1 Introducción

El Grupo de Investigación y Desarrollo de Software Aeroespacial (GIDSA) [1] tiene como objetivo proponer y probar prototipos de soluciones de software para el área aeroespacial de nueva generación. El trabajo desarrollado incluye prototipos que utilizan interpretes de propósito general para decodificar telemetría y scripts de comandos, adopción de estándares bien probados en la industria del software, almacenamiento masivo de telemetría y detección de fallas [2], [3] y [4]. El prototipo funcional del segmento terreno se encuentra público en internet y funciona con datos de satélite reales, principalmente obtenidos de la red SatNOGS [5] y [6].
La detección temprana de anomalías en sistemas complejos como satélites artificiales son de vital importancia teniendo en cuenta el costo de las misiones y la dificultad de reparar daños. El control de límites superior e inferior para muchas variables de telemetría suelen ser la técnica más común para detectar comportamientos anómalos[7]. Como se indicó en un artículo anterior [8], la

ISBN 978–987–633–574–4

salud del satélite se controla con la ayuda constante de un experto, utilizando poca potencia computacional. Mientras tanto, en la industria del software, el aprendizaje automático se utiliza actualmente para diferentes tipos de detección de anomalías, como fraudes con tarjetas de crédito y detección de intrusiones entre otros [9] y [10] . El objetivo de utilizar el aprendizaje automático es lograr una detección temprana de fallas evitando, en la medida de lo posible, la evaluación constante por parte de expertos así como detectar tipos de anomalías desconocidas previamente. El aprendizaje automático ofrece una interesante variedad de posibilidades de predicción y detección de anomalías. Hay dos tipos de algoritmos de aprendizaje automático: aprendizaje automático supervisado y no supervisado. El primero depende de los datos de entrada etiquetados, es decir, el conjunto de datos de entrada debe haber definido si un dato se considera una anomalía o no. El aprendizaje no supervisado no depende de los datos de entrada etiquetados, sino que aprende la representación interna del conjunto de datos y genera patrones [11]. Una anomalía es cualquier dato que se desvía de lo esperado o normal. En la literatura estadística, también se les conoce como valores atípicos o outliers. Cada dato que es procesado por el prototipo será clasificado usando etiquetas binarias: un dato es una anomalía o no [9]. Para detectar anomalías, los algoritmos de aprendizaje automático de detección de fallas crean un modelo del patrón nominal en el conjunto de datos, luego calculan una puntuación para cada valor como medida de cuan atípico es. Dependiendo del algoritmo, esta puntuación atípica toma en en cuenta la correlación con diferentes características o no [9]. En este trabajo y en el caso de series de datos de tiempo, buscamos una secuencia de valores atípicos que determina una anomalía en lugar de un dato particular, buscamos un comportamiento anormal del sistema en lugar de un valor incorrecto.
El UNLaM Ground Segment (UGS) posee el control de límites clásico desde su primera versión. En versiones posteriores se implementaron módulos prototipo que modifican los limites dinámicamente[8]. En este trabajo, en lugar de trabajar con límites, buscamos obtener una medida de éxito en la detección de un comportamiento anómalo del subsistema de potencia, comportamiento que no puede ser detectado por el control de limites clásico. El trabajo actual se exploran dos métodos de aprendizaje automático diferentes, mezcla gaussiana y rolling means. Estos dos algoritmos son investigados y comparados entre sí para estudiar la viabilidad de aplicarlos en la detección de patrones y comportamientos en un prototipo de control de salud en tiempo real.

2 Materiales y métodos

Para estos experimentos usamos nuestro propio conjunto de datos de telemetría real [12]. La fuente de la telemetría es un satélite científico de órbita baja. Puntualmente se utilizará telemetría del subsistema de potencia, incluyendo voltaje medio de batería, sensores de corrientes redundados, una bandera que indica si la batería esta en proceso de carga o descarga, y corrientes medidas de forma independiente en 24 paneles solares. La tabla 1 muestra el significado de cada

ISBN 978–987–633–574–4

campo según descripción disponible en la documentación del fabricante. Todos los valores, excepto *vBatAverage* y *BatteryDischarging* están en bruto(raw), sin embargo, los datos siempre se normalizan antes de ser procesados. Desafortunadamente, el conjunto de datos no cuenta con fallas documentadas por un experto. La telemetría comienza en 2015-05-27 08:51:06 +00:00 y termina en 2015-06-05 23:34:06 +00:00. Para estos experimentos usamos solo los dos primeros días, desde 2015-05-27 08:51:22 +00:00 hasta 2015-05-29 08:50:59 +00:00.

Para crear una anomalía artificial, se corta modificando la telemetría, parcialmente la generación energía del panel solar 24 poniendo en 0 la corriente (128 en bruto) en el conjunto de datos de prueba. El corte es progresivo y cubre 1079 tuplas. Esto es similar a dejar el panel eclipsado (según posición orbital), independientemente del contexto real. Se debe tener en cuenta que el límite clásico el control no puede manejar este comportamiento, debido al hecho de que las corrientes cercanas a 0 son perfectamente válidas en períodos de eclipse reales pero no son esperadas cuando existe exposición al sol.

Feature	Meaning
vBatAverage	Average of Battery voltage used by supervisions
BatteryDischarging	Flag True/False if battery is discharging
ISenseRS1	IsenseRS1 current (battery current)
ISenseRS2	IsenseRS2 current (battery current)
V_MODULE_N_SA	Current in solar panel #N con 0<N<25

Table 1: DataSet Features

Se utilizan dos algoritmos de aprendizaje automático diferentes para detectar anomalías en Telemetría satelital: Mezcla gaussiana y Rolling Means. El primero se aplica la telemetría del subsistema de potencia en su conjunto, utilizando la correlación entre variables, mientras que la última se aplica a cada variable de telemetría de manera aislada. Ambos modelos siguen enfoques estadísticos clásicos: ambos utilizan medidas estadísticas como media, desviación estándar y probabilidad.

2.1 Mezcla Gaussiana

Todas las telemetrías del subsistema de potencia están altamente correlacionadas como se muestra en la matriz de correlación 1. Los coeficientes de correlación cercanos a 1 o -1 muestran una alta interdependencia entre las variables. Por razones de tamaño, se muestra la correlación de solo 4 de las 24 características de telemetría pertenecientes a los 24 paneles solares.

ISBN 978–987–633–574–4

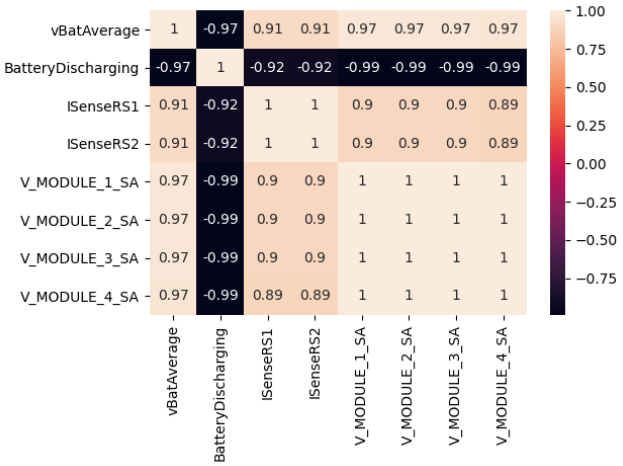


Fig. 1: Correlaciones entre valores de telemetría del subsistema de potencia

Usando la biblioteca sklearn [13], creamos una Modelo de Mezcla Gaussiana, del Inglés Gaussian Mixture Model (GMM). GMM puede utilizarse para agrupar datos sin etiquetar, GMM puede ayudar a detectar un comportamiento lejano o poco probable que el comportamiento nominal. Cualquier punto muy alejado de las funciones gaussianas podrían considerarse una anomalía. El conjunto de datos se dividió en dos subconjuntos de datos: conjunto de datos de entrenamiento y conjunto de datos de prueba . La prueba se realiza durante dos días de telemetría. El 20% final del conjunto de datos se utiliza para la prueba mientras que el otro 80% forma el conjunto de datos de entrenamiento. Un modelo es obtenido al ejecutar el algoritmo sobre el conjunto de entrenamiento, donde la cantidad de componentes y el tipo de covarianza se seleccionan en un proceso iterativo que analiza information-theoretic criteria (BIC), cubriendo los 4 tipos de covarianza y la cantidad de componentes (grupos o clusters) entre 1 y 20. La puntuación mínima de valores atípicos se establece como límite para la prueba.

2.2 Rolling Means

Rolling Means utiliza un enfoque estadístico simple para la detección de anomalías sobre un conjunto de datos de serie de tiempos. Dada una serie de referencias y una ventana de tamaño fijo N, el algoritmo obtiene primero la media de los N registros iniciales de la serie. Entonces la ventana se "mueve hacia adelante" en uno, recalculando la media de la ventana. Este proceso se repite hasta que la ventana final incluye el dato final. Una vez que todos las medias se ha obtenido, el algoritmo etiqueta como valores atípicos todos los puntos cuyo desvío de la media es S veces mayor que la desviación estándar que corresponde al punto.

Se aplica Rolling Means mediante un algoritmo [14] a cada subconjunto de datos, uno para cada variable de telemetría, y para cada uno genera un modelo

ISBN 978–987–633–574–4

de normalidad. Para utilizar este algoritmo, se debe establecer el tamaño de la ventana y un número fijo de desviaciones estándar. Para decidir el valor de estos parámetros, se ejecutan varias iteraciones con diferentes valores en el mismo conjunto de datos, y finalmente, conociendo la naturaleza de los datos y tomando el rol de experto, los mejores valores se utilizan. Se eligió Rolling Means ya que es un algoritmo sensible a valores anómalos, siendo simple de implementar. Se basa en la desviación estándar, teniendo en cuenta el cambio en la serie de tiempo usando la ventana de tamaño fijo.

2.3 Otros métodos

También se probó el método de distribución normal multivariable, pero se descartó a favor de la Mezcla Gaussiana, dado que el primero necesita que sus datos sigan una distribución normal y no puede manejar varias campanas. Isolation Tree también fue analizado, pero se descartó dado que se etiquetaron incorrectamente la mayor parte de el conjunto de datos "normal" como anomalías, sin tener la posibilidad de utilizar un parámetro para cambiar su comportamiento.

3 Resultados

3.1 Modelo de Mezcla Gaussiana

Para obtener un grafico que nos brinde una aproximación visual al modelo generado utilizamos inicialmente solo dos características $V\_MODULE\_24\_SA$ y $vBatAverage$ sobre los datos de entrenamiento.La figura 2 muestra graficamente las funciones gaussianas en verde y las diferentes agrupaciones generadas. Se testean los datos de prueba con el modelo previamente generado. La figura 3 muestra como los datos de prueba sin modificación artificial ajustan al modelo.

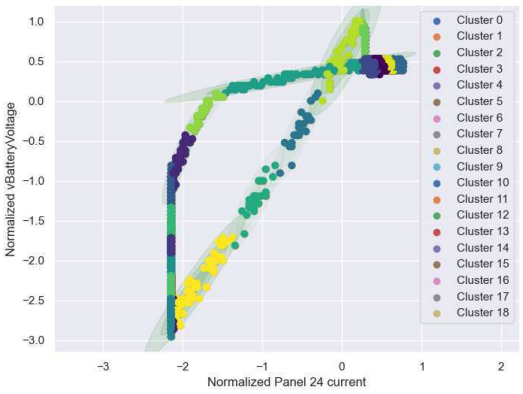


Fig. 2: Grupos o Clusters creados para 2 variables usando mezcla gaussiana

ISBN 978–987–633–574–4

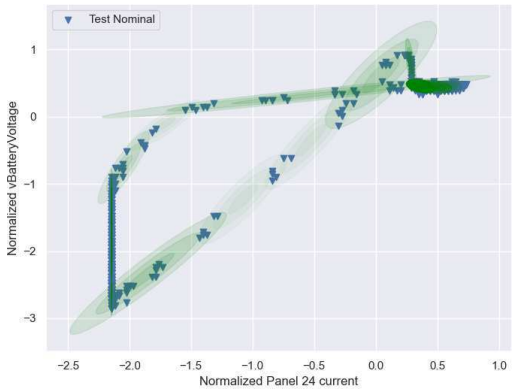


Fig. 3: Mezcla Gaussiana aplicada al dataset de prueba sin anomalías generadas artificialmente. En verde las funciones gaussianas, todos los datos se ajustan al modelo. No hay falsos positivos

Los resultados con el conjunto de datos modificado artificialmente, con solo 2 variables ($V\_MODULE\_24\_SA$ y $vBatAverage$), simulando corriente 0 en el panel solar 24, se muestran en la figura 4. Se detectan 880 anomalías. Si bien la caída progresiva de corriente no permite separar de forma claro cual dato es anómalo y cual no, la cantidad obtenida sobre el total de datos es una buena medida del estado general del sistema.

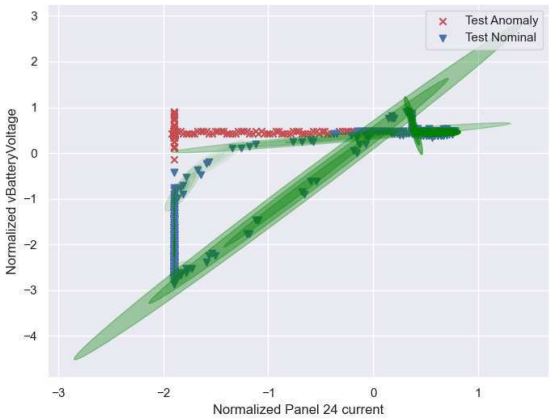


Fig. 4: Mezcla Gaussiana aplicada al dataset de prueba con anomalías

ISBN 978–987–633–574–4

Si utilizamos las 28 características disponibles en el dataset 1 obtenemos 11 falsos positivos con el conjunto de datos sin modificar, es el 0,25% del conjunto de datos de prueba, si, también usando todas las características 1 pero con el dataset modificado artificialmente obtenemos 925 anomalías. Los resultados del experimento con 2 y 28 características, para datos originales o modificados artificialmente son mostrados en la tabla 2.

# Variables	Dataset Normal	Dataset c/anomalias
2	0	880
28	11	925

Table 2: Anomalías detectadas por GMM

3.2 Rolling Means

Se utiliza el subconjunto de datos compuesto por los datos de la corriente del Panel 24. Para este subconjunto de datos, el algoritmo Rolling Means etiqueta los datos de cada variable como anomalías o no, según el "modelo de normalidad".

Ejecutando el algoritmo, con un tamaño de ventana de 1000 (una ventana que es la mitad del número de anomalías insertadas), utilizando 1 y 2 desviaciones estándar obtiene los próximos resultados. Cada subconjunto de datos se traza con líneas azules que representan datos considerados nominales y lineas rojas que representan los puntos de anomalías que detectó el algoritmo. La tabla 3 muestra, para una y dos desviaciones estándar la cantidad de anomalías detectadas en el dataset original y el modificado artificialmente.

#desviaciones estándar	Dataset Normal	Dataset c/anomalías
1	260	449
2	71	6

Table 3: Anomalías detectadas por Rolling Means para una y dos desviaciones estándar sobre dataset original y modificado

ISBN 978–987–633–574–4

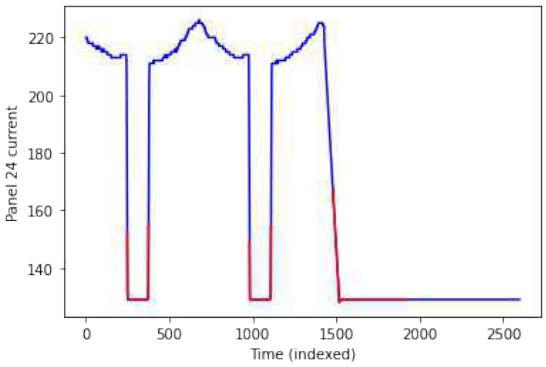


Fig. 5: Rolling Means aplicado al Panel 24 usando una desviación estándar

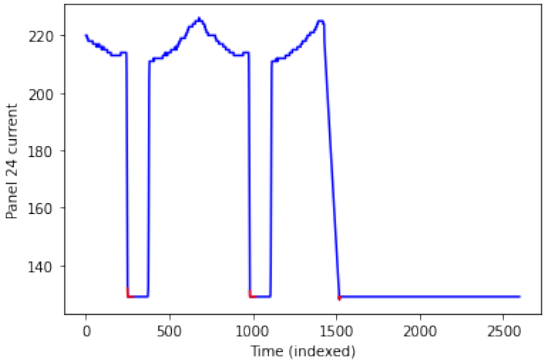


Fig. 6: Rolling Means aplicado al Panel 24 usando dos desviaciones estándar

4 Conclusiones

La tabla 4 muestra una comparación entre los resultados obtenidos en los dataset con y sin anomalías, utilizando una o dos desviaciones estándar y solo dos características o todo el conjunto para Mezcla Gaussiana.

ISBN 978–987–633–574–4

Algoritmo	Var/Desv	Dataset Normal	Dataset c/anomalías
Rolling Means	1	260	449
Mezcla Gaussiana	2	0	860
Rolling Means	2	71	6
Mezcla Gaussiana	28	11	925

Table 4: Rolling means vs Mezcla Gaussiana

En el caso de Rolling Means, el modelo no es sensible a la correlación y define si un dato es anómalo basándose en la tendencia del conjunto de datos de series de tiempo. Un pico aislado en el gráfico se etiquetará como una anomalía, pero puede ser el resultado de una acción contextual esperada. Aunque Rolling Means es un algoritmo simple, con pocas necesidades computacionales, al no tener en cuenta el contexto y las correlaciones no pueden manejar anomalías específicas y dependientes del contexto. Usando una desviación estándar parece detectar las anomalías introducidas, pero también etiqueta erróneamente los datos válidos. Usando dos desviaciones estándar, contrariamente a lo esperado, se comporta de la misma manera, etiquetando incorrectamente aún más datos. Rolling Means es un método válido para detectar valores atípicos producidos por ruido, pero no puede considerarse un algoritmo válido para la detección de anomalías. También necesita la intervención de un experto para establecer los parámetros iniciales.
Por otro lado, el método de Mezcla Gaussiana muestra resultados prometedores. Se detectaron anomalías, sin etiquetar incorrectamente una gran cantidad de registros (signo de que el modelo no se ha sobreentrenado). Estas anomalías introducidas no pueden ser detectadas por los sistemas de control de límites dado que los valores probados son normales en un contexto determinado. La covarianza y la cantidad de clusters se obtuvieron automáticamente, sin un intervención experta.

5 Trabajo Futuro

Los resultados dan una vista informativa de los diferentes algoritmos, pero no pueden ser evaluado objetivamente ya que no hay datos etiquetados disponibles para compararlos. Si se pudieran obtener datos preetiquetados, se utilizarían métricas estadísticas para evaluar los resultados y ajustar los parámetros de los modelos para minimizar el número de falsos positivos producidos por el prototipo. Entre los requerimientos típicos de estos sistemas se encuentra la detección de anomalías. Otros algoritmos como DBScan, y técnicas de aprendizaje profundo deben ser exploradas como alternativas a los métodos analizados en el presente trabajo.

ISBN 978–987–633–574–4

References

1. Gidsa - grupo de investigación y desarrollo de software aeroespacial: Home. https://gidsa.unlam.edu.ar/.
2. Pablo Soligo and Jorge Salvador Ierache. Software de segmento terreno de próxima generación. In *XXIV Congreso Argentino de Ciencias de la Computación (La Plata, 2018).*, 2018.
3. Pablo Soligo and Jorge Salvador Ierache. Segmento terreno para misiones espaciales de próxima generación. *WICC 2019.*
4. Pablo Soligo, Jorge Salvador Ierache, and German Merkel. Telemetría de altas prestaciones sobre base de datos de serie de tiempos. 2020.
5. Unlam Ground Segment: Home unlam ground segment: Home. `https://ugs.unlam.edu.ar/`. Accessed: 2021-07-30.
6. Satnogs satnogs. `https://satnogs.org/`. Accessed: 2021-07-30.
7. Takehisa Yairi, Minoru Nakatsugawa, Koichi Hori, Shinichi Nakasuka, Kazuo Machida, and Naoki Ishihama. Adaptive limit checking for spacecraft telemetry data using regression tree learning. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, volume 6, pages 5130–5135. IEEE, 2004.
8. Pablo Soligo and Jorge Salvador Ierache. Arquitectura de segmento terreno satelital adaptada para el control de límites de telemetría dinámicos. 2019.
9. Charu Aggarwal. *An introduction to outlier analysis.* Springer New York, 1 edition, 2017.
10. Aaron Rosenbaum. Detecting credit card fraud with machine learning. 2019.
11. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of Statistical Learning.* Springer Series in Statistics. Springer, 2 edition, 2008.
12. Low Orbit Satellite Dataset: Home low orbit satellite dataset: Home. `https://gidsa.unlam.edu.ar/data/LowOrbitSatellite.csv`. Accessed: 2021-07-30.
13. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
14. Algoritmo Rolling Means rollingmeans. `https://gidsa.unlam.edu.ar/data/rolling.py`. Accessed: 2021-07-30.

ISBN 978–987–633–574–4

Mapyzer: una herramienta de carga y visualización de datos espacio-temporales

Gustavo Marcelo Nuñez[2], Markel Jaureguibehere[2], Carlos Buckle[1,2][0000−0003−0722−0949], Leo Ordinez[1,2][0000−0003−2237−812X], and Damián Barry[1,2]

[1] Laboratorio de Investigación en Informática (LINVI)
[2] https://linvi.unp.edu.ar
[3] Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco
Puerto Madryn, Argentina
{guscostaf,markeljaure2000,carlos.buckle,leo.ordinez,demian.barry}@gmail.com

Resumen La emergencia generada en distintos ámbitos de la sociedad a raíz de la pandemia por COVID-19 dio lugar a la necesidad de realizar acciones de planificación urgentes. Así, los datos, escasos, dispersos y en muchos casos parciales se volvieron preciados para la toma de decisiones de planificación urbana. Entre las múltiples demandas espontáneas atendidas por la Universidades Nacionales de Argentina, se encuentra el desarrollo de un software para la visualización de datos espacio-temporales, que permita la puesta en valor rápida de estos activos para organizaciones con bajo nivel de gestión y escasos recursos técnicos. Dicho desarrollo se realizó por un equipo de alumnos y docentes.

Keywords: COVID-19 · datos espacio-temporales · visualización dinámica

1. Introducción

La propuesta desarrollada fue llevada a cabo bajo el contexto de pandemia de COVID-19 originada a principios del 2020, y tiene como principal objetivo el aportar conocimiento a través de la representación de datos espacio-temporales, que asistan en el ordenamiento de datos y faciliten la toma de decisiones. El contexto sanitario mundial, también ha marcado los hábitos y las costumbres relacionadas con el acceso a la información, convirtiéndose los datos en un activo aún más preciado en niveles ejecutivos, gerenciales y decisorios.

En función de diversas demandas espontáneas, tanto intra como extra universitarias, surgidas en el contexto mencionado, se identificaron una serie de requerimientos simples, que dieron lugar a este trabajo. Esta situación de demanda fue generalizada a todo el sistema científico-académico argentino [1,2,3,7]. En este sentido, se elaboró un perfil de destinatario potencial de la propuesta aquí presentada. Dicho destinatario se caracterizaría por contar con datos espaciales (no necesariamente georreferenciados, sino con direcciones) organizados en un formato plano, no relacional, como una tabla u hoja de cálculo. A esto se

agrega la variabilidad en el tiempo de esos datos. En relación a ello, el principal problema detectado es que, las herramientas existentes que ofrecen representación de datos sobre mapas, no disponen de funcionalidades para la visualización del comportamiento de los datos a través del tiempo. A la vez, la carga masiva a partir de un formato como el de las hojas de cálculo, sin georreferenciación no resulta amigable a un usuario no avezado en Sistemas de Información Geográfica (SIG)[4][5]. Entonces, y bajo el contexto mencionado, resulta necesario contar con herramientas de este estilo, ya que contribuyen y aportan valor a partir del conocimiento que organizan y muestran.

El resto del trabajo se organiza de la siguiente manera: la Sección 2 expone los requerimientos desde el punto de vista del negocio, en términos de usuarios potenciales; en la Sección 3 se presenta la solución contruida; y finalmente en la Sección 4 se exponen las conclusiones y trabajos futuros.

2. Requerimientos del negocio

Las entidades u organizaciones que desarrollan proyectos con información geográfica requieren manejar, además de los datos propios de cada proyecto, datos espacio-temporales recolectados en trabajos de campo, aplicaciones, o investigaciones anteriores. Si bien cada proyecto maneja datos de un determinado dominio, todos tienen una necesidad común: gestionar, analizar y visualizar datos geográficos variables en el tiempo. Para un mejor aprovechamiento de los recursos, se visualiza la posibilidad de reuso de datos entre proyectos como el camino hacia la optimización y la potencial integración de información para la generación de nuevos conocimientos. En este marco, surgen las necesidades de pensar herramientas que sean capaces de manejar de forma integrada datos geo-temporales de diferentes dominios.

Las herramientas para atender estas necesidades encuadran dentro de los Sistemas de Información Espacial y Temporal. Un sub-conjunto de ellos son los Sistemas de Información Geográfica (SIG) con la capacidad de capturar, procesar y reportar información de espacial. Para ello manejan datos espaciales genéricos: *Localizaciones* (lugares) expresados a través de coordenadas, *Relaciones* (trayectos ó áreas) expresadas como colecciones de localizaciones y *Descripciones* (Clasificaciones, elementos visuales, atributos propios, etc). La referencia geográfica de estos elementos es la superficie de la tierra y sobre ellos se permiten superponer mapas o imágenes como así también establecer capas (*layers*) que los agrupen por criterios específicos del dominio que se esté representando. Los elementos del SIG residen en una base de datos de la cual se puede extraer información en formato tabular o geográfico y permiten carga masiva a través de archivos con formatos estándar como KML/KMZ, GeoJSON y otros.

Se encuentran disponibles muchos paquetes de software que cubren este conjunto de requerimientos comunes, pero las necesidades de este proyecto plantea-

[4] https://www.arcgis.com
[5] https://www.qgis.org

ban requisitos particulares que no fueron encontrados en los productos evaluados. Ellos son:

Carga masiva amigable de datos urbanos: Los proyectos que manejan información urbana o territorial tienen la necesidad de capturar, individual o masivamente, datos geográficos referidos a domicilios o lugares identificados en mapas por medio de rótulos y que el sistema se encargue de su geolocalización. Asimismo, hay muchos proveedores de datos que no cuentan con sistemas SIG y por ende no pueden entregar datos en formatos estándar. No obstante, sí tienen la posibilidad de generar rápidamente archivos XLS o CSV con elementos geográficos de los cuales sólo cuentan con un rótulo, una breve descripción y un domicilio o rótulo de ubicación.

Vigencia temporal de los elementos geográficos: Los datos utilizados en proyectos territoriales y urbanos suelen estar asociados a una línea de tiempo y es de suma importancia poder validar la vigencia temporal de los elementos como así también serializarlos en el tiempo.

Personalización de atributos de los elementos geográficos: Dado que cada proyecto maneja datos de diferentes dominios, los elementos geográficos de cada proyecto deben poder ser clasificados mediante clasificadores propios y descriptos por atributos propios del área de estudio.

Visualizaciones basadas en la dinámica temporal: Sumado a las capacidades de visualización por capas, es fundamental que se provean visualizaciones que dispongan de la posibilidad de desplegar elementos válidos en diferentes momentos del tiempo, para ello son necesarios componentes interactivos que "animen" o permitan desplazar una barra de tiempo sobre un mapa.

3. Solución propuesta

Como solución a las necesidades planteadas en la Sección 2, el equipo de desarrollo optó por construir una aplicación web basada en la arquitectura estilo RESTful API [4], que utiliza un modelo de diseño cliente-servidor para el soporte de la concurrencia de múltiples usuarios. A continuación, se brindan los detalles sobre las distintas perspectivas consideradas a lo largo del avance general del proyecto.

3.1. Requerimientos de Software

En base a las necesidades del negocio, la Figura 1 muestra cómo se agrupan los aspectos que se han tenido en cuenta a la hora de definir las distintas funcionalidades del sistema. El método utilizado fue el Mapeo de Historias de Usuarios (*User Story Mapping*, en inglés) [5,6]. Esto permitió al equipo de trabajo dividir y priorizar los distintos bloques del software, organizando y distribuyendo el

trabajo a realizar a lo largo del tiempo total del proyecto. Este diagrama no solo aporta una visión general del producto, sino que también permite visualizar el progreso de lo realizado durante cada *release*, separado por sus distintos módulos funcionales. En la figura se puede ver cómo se organizan las tres grandes funcionalidades del sistema (en las calles verticales) y su descomposición en Historias de Usuario de alto nivel (épicas), a lo largo de los diferentes *sprints* ejecutados (calles horizontales).

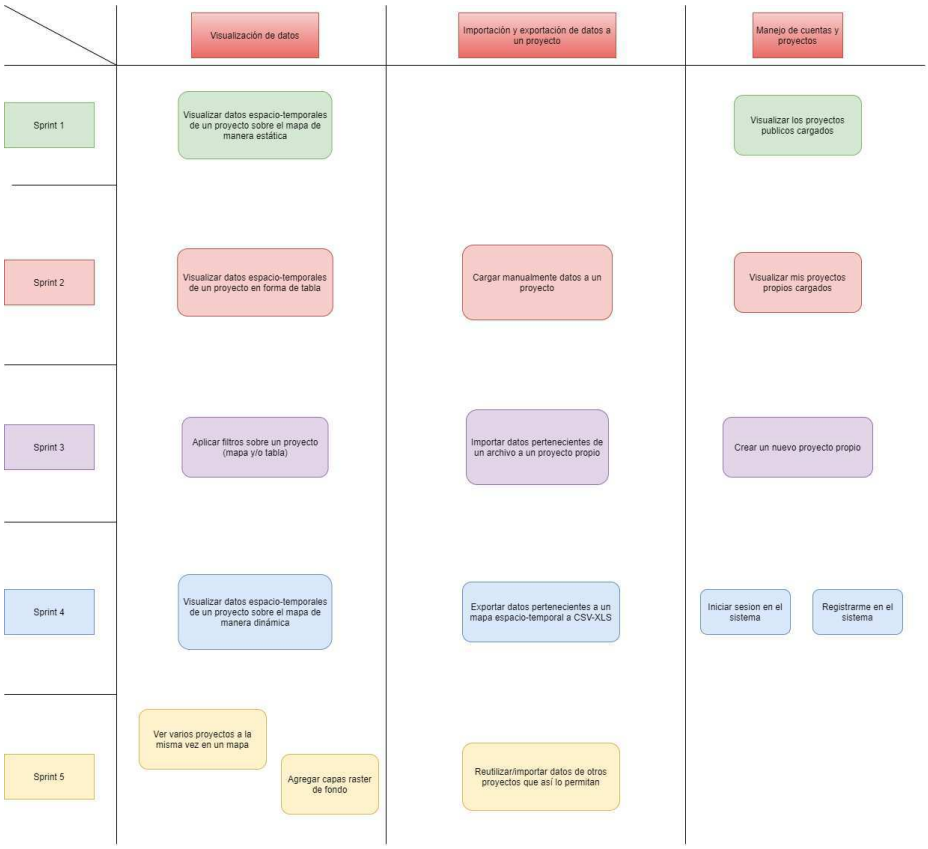


Figura 1: Mapeo de Historias de Usuario en formato visual.

El producto a construir va a proporcionar a los usuarios la capacidad de crear múltiples proyectos, en los cuales se podrán cargar datos que tengan una posición geográfica (sea un lugar, una zona o un trayecto), junto a un rango de tiempo en donde el mismo tiene validez (también puede ser un instante de tiempo). Sobre los mismos proyectos se brindara la posibilidad de ver sus datos en una tabla o en un mapa dinámico, que permita a los usuarios desplazarse en

el tiempo para ver como éstos varían en el tiempo y espacio, permitiendo así un análisis visual rápido y una representación gráfica agradable del proyecto.

En cuanto a la carga de datos, el sistema va a estar preparado para la importación de datos manual o masiva a partir de los tipos de archivos más comunes (CSV, XLS). A la vez, se dará soporte a distintos tipos de esquemas para que el usuario pueda cargar datos que tiene en distintos formatos, por ejemplo, tener un lugar por sus coordenadas o dada su dirección (calle, numero, y ciudad). Finalmente, brindará posibilidad de modificar los datos, exportarlos o reutilizarlos en otros proyectos desde la misma aplicación web.

A continuación se describen las principales entidades del dominio identificadas.

Lugar. Los lugares son representados a través de marcadores sobre el mapa. Este tipo de dato puede ingresarse de dos maneras; a partir de un par coordenadas (latitud y longitud), o a partir de una dirección. A su vez, cada lugar debe tener asignado un tipo de lugar.

Tipo de lugar. Los tipos de lugar son uno de los atributos correspondientes a los distintos lugares. El tipo de lugar asignado es representado sobre el mapa con su ícono correspondiente.

Zona. Las zonas son otro de los tipos de datos que permite representar la aplicación. Se registran en el sistema a partir del ingreso de coordenadas, y se ven visualizarán sobre el mapa con forma de polígono, donde cada vértice corresponde a una coordenada ingresada. Al igual que los lugares, cada zona debe tener asignada un tipo de zona.

Tipo de zona. Los tipos de zona son uno de los atributos correspondientes a cada zona en particular, esto tiene como principal objetivo el poder diferenciar qué tipo de zona se visualiza sobre el mapa.

Trayecto. El último de los tipos de datos representables son los trayectos. Este tipo de datos corresponde a una sucesión de coordenadas que, una vez ingresadas, se podrán visualizar sobre el mapa unidas entre sí. Al igual que sucede con los lugares y las zonas, los trayectos también deben tener asignado un tipo de trayecto.

Tipos de trayecto. Atributo correspondiente a los trayectos. El tipo de trayecto establece el grosor, el tipo de línea y el color de la misma. De esta manera, es posible distinguir a simple vista qué tipo de trayecto está representado sobre el mapa.

Usuario Invitado. Los usuarios invitados son aquellos usuarios que no se encuentran registrados en el sistema. Este tipo de usuarios solo podrá acceder y visualizar proyectos públicos. Para crear y/o editar sus propios proyectos, deberán registrarse en el sistema y ser validado por un administrador. Una vez registrado, el usuario invitado tendrá el rol de dueño de los datos.

Usuario Dueño de los datos. Los dueños de los datos son aquellos usuarios que han realizado el proceso de registro en el sistema y sido admitidos por un administrador. Estos usuarios, además de tener acceso a proyectos públicos, podrán crear y gestionar sus propios proyectos, además de sus datos correspondientes.

Usuario Administrador. Por último, el usuario de tipo administrador tendrá control sobre los usuarios registrados y sus roles. Además, es el único tipo de usuario que puede crear y/o modificar los tipos de lugar, tipos de zona y tipos de trayecto.

Proyectos. Los proyectos permiten mantener organizados los datos y poder visualizarlos de manera ordenada sobre una tabla, o representados sobre el mapa. Los proyectos pueden ser de tipo público o privado, lo cual determinará quién tiene acceso al proyecto y a su contenido. Un dueño de los datos puede crear varios proyectos. Sin embargo, un proyecto pertenece a un único dueño de los datos. Dentro de su proyecto, el dueño de los datos podrá agregar, modificar y/o eliminar los datos según su criterio.

3.2. Modelo de Datos

Una visión estática del dominio del problema se presenta en la Figura 2. Allí se muestra un diagrama Entidad-Relación del sistema. En el modelo se representan las entidades, sus relaciones y se indican aquellas que pueden *variar en el tiempo*, esto es, presentan un comportamiento dinámico.

En vista de la necesidad de manejar datos de tipo geográfico a nivel de base de datos, se optó por utilizar como motor PostgreSQL con su extensión PostGIS. El principal motivo de esta selección es el soporte de datos espaciales (puntos, polígonos, líneas) y su manejo a nivel de primitivas de la base de datos.

3.3. Arquitectura de *Backend*

Con respecto a la arquitectura del servidor, como la gran mayoría de modelos de negocio de este estilo, se ha implementado la arquitectura por capas, desarrollada en NodeJS. Esta distribución típica de aplicaciones web, permite desacoplar los distintos módulos según su funcionalidad (presentación, negocio, datos).

Como se puede ver en la Figura 3, el énfasis y la guía del diseño arquitectónico está puesto en el manejo de datos espaciales, siendo sus clases de objetos las que centralizan la atención.

3.4. Arquitectura de *Frontend*

Para el desarrollo del cliente de la aplicación, se utilizo el conocido Framework AngularJS, complementado con la librería Bootstrap para añadir componentes de diseño, y Leaflet para la visualización cartográfica de datos. La arquitectura del framework es la de *Model-View-Controller* (MVC), que separa los módulos e interfaces por medio de *componentes*, aunque se añaden algunas mejoras [6].

La Figura 4 ilustra de manera esquemática cómo un requerimiento del usuario, materializado en una visualización de geoespacial, atraviesa los distintos componentes del framework, realiza la solicitud correspondiente al servidor y recorre el camino de vuelta presentando los datos.

[6] https://v2.angular.io/docs/ts/latest/guide/architecture.html

ISBN 978–987–633–574–4

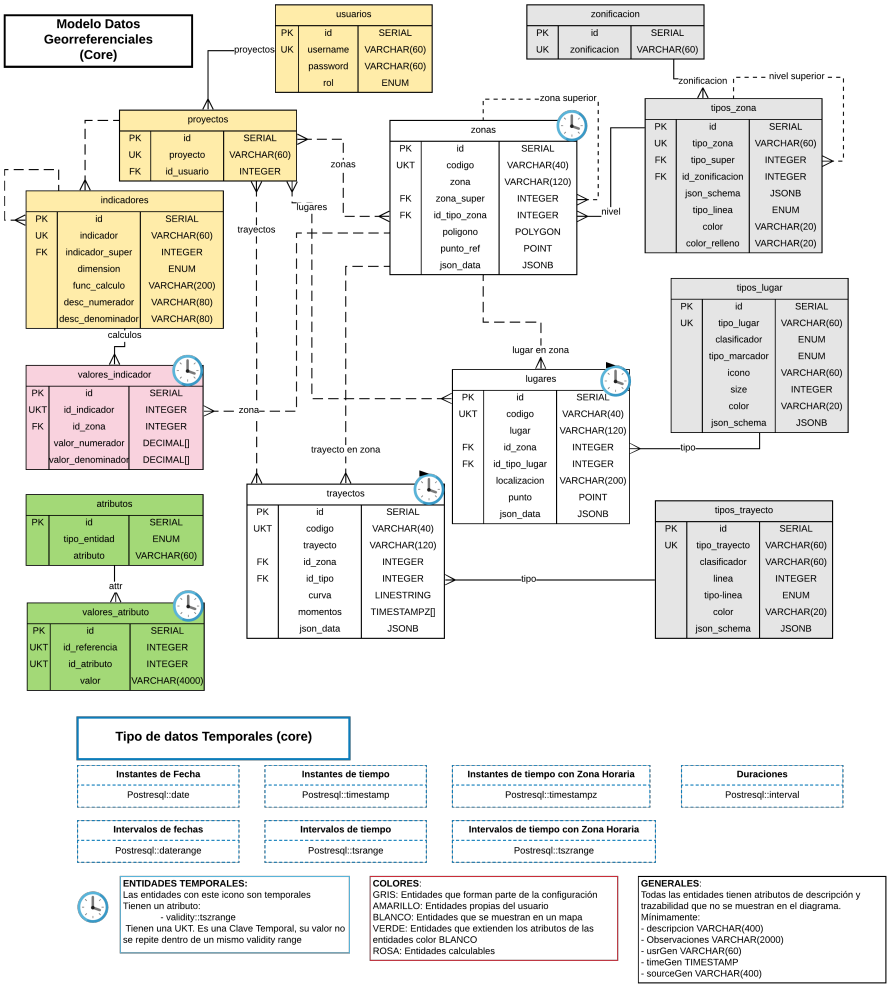


Figura 2: Modelo de datos del sistema.

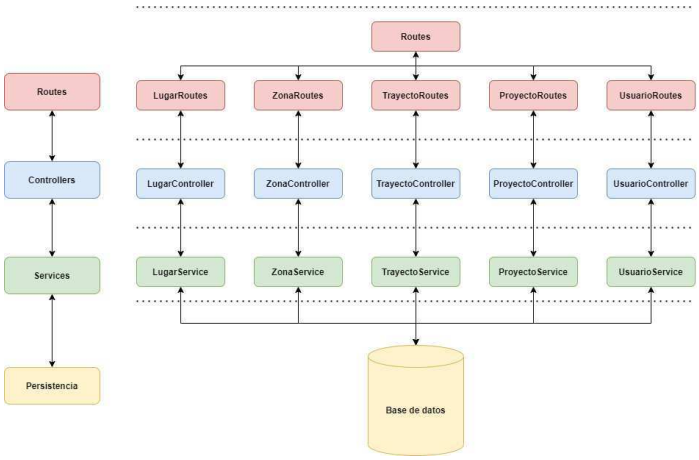


Figura 3: Arquitectura de software, servidor.

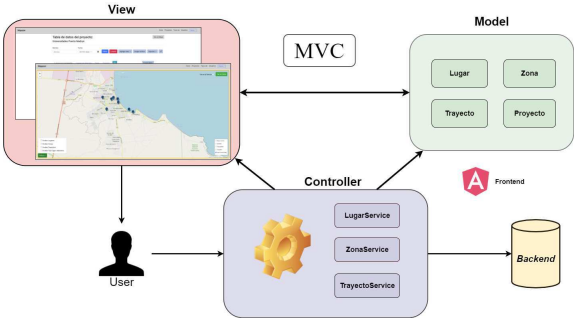


Figura 4: Arquitectura de software, cliente.

3.5. Representación de Datos Dinámicos

La Figura 5 muestra la funcionalidad principal de la solución desarrollada: su mapa dinámico y la visualización de datos espacio-temporales. Allí es posible visualizar los datos pertenecientes al/los proyecto/s seleccionado/s, esto es, superponer datos que así lo permitan de acuerdo a su visibilidad (público o privado).

En la interfaz de mapa de un proyecto, se pueden distinguir los lugares, zonas y trayectos correspondientes. A su vez, el mapa posee distintas opciones para interactuar con los datos que se muestran en pantalla, como por ejemplo:

Slider temporal: El usuario podrá visualizar los datos de manera temporal. Para ello, solo basta con hacer clic en la casilla `Ver en el tiempo`, ubicada en la esquina superior derecha. A continuación, se activará el slider temporal y podrá visualizar los datos de manera dinámica.

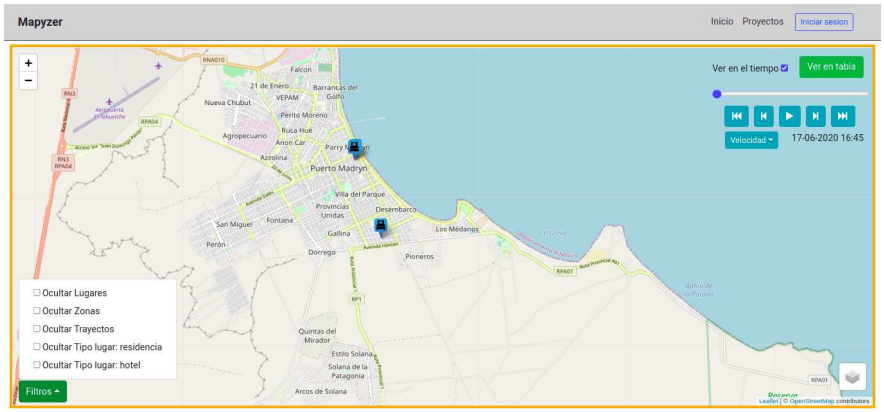


Figura 5: Visualización dinámica de datos.

Animación: Con la casilla `Ver en el tiempo` seleccionada, también se mostrarán en la pantalla una serie de botones de reproducción. Con ellos, el usuario podrá ver la evolución de los datos a través del tiempo de manera automática.

Filtros: El usuario podrá aplicar distintos filtros sobre el mapa. Para ello, deberá hacer clic sobre el botón Filtros ubicado en la esquina inferior izquierda. A través del menú desplegado, se seleccionan los filtros deseados según su interés.

Capas Raster: Esta opción le permite al usuario visualizar los datos a través de distintas capas. Para ello, deberá hacer clic sobre el botón ubicado en la esquina superior derecha, y seleccionar la capa deseada.

4. Conclusiones y Trabajos futuros

En este trabajo se sintetizó un desarrollo de software surgido de manera espontánea con el objetivo de aportar una contribución técnica a la emergencia dada por la pandemia de COVID-19. Entre los varios aspectos sociales, económicos y gubernamentales que afecta aún la situación sanitaria, se encuentra la necesidad de gestionar de manera eficiente y urgente distintos frentes de conflicto. En este sentido, se advirtió, a partir de demandantes reales que requirieron asistencia al grupo de trabajo, la necesidad de visualizar datos cartográficos y al mismo tiempo poder analizar su evolución en el tiempo. Así, el arribo de personas a una ciudad, contando con información sobre su origen; los puntos de asistencia alimentaria disponibles; los casos confirmados y sus ámbitos laborales o escolares; las zonas de circulación restringida; entre otras situaciones, constituyeron la motivación de este trabajo.

En términos del impacto, es importante destacar las vinculaciones que se dieron a partir de este proyecto y las capacidades que se generaron en virtud de

ello. El proyecto “Análisis Prospectivo Inteligente Del Impacto Social, Económico y Productivo Del COVID-19 En La Provincia De Chubut”, financiado por la convocatoria Programa De Articulación y Fortalecimiento Federal De Las Capacidades En Ciencia y Tecnología COVID-19, que dio marcó a este trabajo permitió vincular al grupo de informática, y principalmente a los estudiantes participantes del mismo con instituciones y profesionales tales como geógrafos/as, sociólogos/as, economistas, personal de salud, asociaciones vecinales, merenderos y funcionarios/as gubernamentales, entre otros.

En este sentido, los vínculos generados se verán fortalecidos en trabajos futuros enmarcados en diferentes iniciativas institucionales como son un Proyecto de Desarrollo Tecnológico Social (PDTS) de reciente aprobación y en el Laboratorio de Sistemas de Información Geográfica de la UNPSJB.

Referencias

1. Casali, A., Torres, D.: Impacto del covid-19 en docentes universitarios argentinos: cambio de prácticas, dificultades y aumento del estrés. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología (28), e53 (4 2021). https://doi.org/10.24215/18509959.28.e53, `https://teyet-revista.info.unlp.edu.ar/TEyET/article/view/1531`
2. Del Valle, D., Perrotta, D., Suasnábar, C.: La universidad argentina pre y post pandemia: acciones frente al covid-19 y los desafíos de una (posible) reforma. Integración y Conocimiento **10**(2) (2021)
3. García de Fanelli, A., Marquina, M., Rabossi, M.: Acción y reacción en época de pandemia: La universidad argentina ante la covid-19. Revista de Educación Superior en América Latina pp. 3–8 (07 2020). https://doi.org/10.14482/esal.8.378.82
4. Fielding, R.T., Taylor, R.N.: Architectural Styles and the Design of Network-Based Software Architectures. Ph.D. thesis (2000), aAI9980887
5. Milicic, A., El Kadiri, S., Perdikakis, A., Ivanov, P., Kiritsis, D.: Toward the definition of domain concepts and knowledge through the application of the user story mapping method. International Journal of Product Lifecycle Management **7**(1), 3–16 (2014)
6. Patton, J., Economy, P.: User story mapping: discover the whole story, build the right product. .ºReilly Media, Inc.”(2014)
7. Zelaya, M.: Las políticas públicas universitarias en el contexto de pandemia en la argentina. Revista de Educación Superior del Sur Global-RESUR (9-10), 172–200 (2020)

ISBN 978–987–633–574–4

Q2MGPS: Una librería para recolectar indicadores QoS sobre redes GPS en dispositivos móviles

Ariel Machini[1], Juan Enriquez[2], Sandra Casas[3],

GISP – Instituto de Tecnología Aplicada
Universidad Nacional de la Patagonia Austral – Unidad Académica de Río Gallegos
[1]arielmachini@protonmail.com, [2]jenriquez@unpa.edu.ar, [3]sicasas@uarg.unpa.edu.ar

Abstract. Durante el transcurso de las últimas décadas, el Sistema de Posicionamiento Global (GPS) se convirtió en una de las herramientas más utilizadas por dispositivos móviles a nivel mundial; Por esta razón, resulta de relevancia estudiar aquellos factores causantes del deterioro en la calidad de los datos provistos por el servicio de GPS. Actualmente, existen distintos trabajos que efectúan análisis sobre la calidad de la información provista por el GPS; sin embargo, estos estudios no son realizados desde los dispositivos móviles que tanto usan este sistema. Es a causa de esto que, en este trabajo, se presenta Q2MGPS: una librería para aplicaciones Android que permite recolectar indicadores sobre GPS. Para verificar el correcto funcionamiento de Q2MGPS, se ejecutaron diversos casos de estudio, cuyos resultados fueron organizados gráficamente. La información recopilada por la librería también se empleó para constatar el posible vínculo entre las condiciones climáticas y la calidad de los datos.

Keywords: Android, GPS, Smartphones, QoS, Clima

1 Introducción

Durante el transcurso de los últimos años, el Sistema de Posicionamiento Global (conocido por sus siglas en inglés GPS) pasó de ser sólo un sistema utilizado con fines militares a la herramienta más precisa, confiable y utilizada en diversos sectores [4].

Uno de los usos más comunes que se le da a esta herramienta actualmente es la navegación: En dispositivos móviles tales como smartphones, el GPS es utilizado para obtener la ubicación de los usuarios con diversos fines; En el caso de Google Maps, por ejemplo, se utiliza la ubicación del usuario para que este pueda acceder al servicio de navegación con el que cuenta dicha aplicación, es decir, para que obtenga (en tiempo real y en función de su ubicación actual) indicaciones útiles para la navegación por parte del dispositivo cuando tiene que desplazarse de un lugar a otro. Es por estas razones que resulta de gran importancia que, al momento en el que el dispositivo requiera utilizar el servicio, este funcione correctamente. Sin embargo, este no es siempre el caso, ya que debido a diversos factores que influyen en el GPS, la información que brinda en ocasiones puede ser incorrecta ([4] y [6]).

En la actualidad, existen múltiples trabajos avocados al análisis de la calidad de la información provista por el Sistema de Posicionamiento Global, no obstante, es escasa

la investigación realizada en torno a los factores de QoS relacionados con GPS desde el punto de vista de uno de los clientes más comunes de este sistema: los dispositivos móviles. A raíz de esto, a través del presente trabajo construimos una librería mediante la cual desarrolladores de aplicaciones (Android) podrán consultar diversos indicadores de calidad (QoS) relacionados con el GPS.

Este trabajo se encuentra organizado de la siguiente manera: En la Sección 2, se discute acerca de los trabajos relacionados; En la Sección 3, se detallan los indicadores y datos que recolecta la librería Q2MGPS; En la Sección 4, se describe la estructura y funcionalidades clave de la librería, así como también el formato que sigue el documento en el que se almacenan los indicadores que esta recolecta; En la Sección 5, se presenta información sobre los casos de prueba llevados a cabo junto con sus resultados y, en la Sección 6, se analizan los resultados obtenidos y se presentan las conclusiones.

2 Trabajos relacionados

Se encontraron aplicaciones, librerías y trabajos en los que se relevan datos sobre la ubicación de dispositivos móviles: *GPSLogger* (https://github.com/mendhak/gpslogger). GPSLogger es una aplicación Android que registra información GPS en varios formatos. Cuenta con una opción para subir dicha información a diversos sitios; *BasicAirData GPS Logger* (https://github.com/BasicAirData/GPSLogger). La aplicación Android GPS Logger de BasicAirData permite registrar recorridos así como la posición actual del dispositivo, junto con varios otros detalles relacionados con GPS. Permite exportar la información registrada en formato KML, GPX y TXT; *TrackLogger* (https://github.com/lyriarte/TrackLogger). Esta aplicación Android registra constantemente (mientras está en funcionamiento) datos sobre la ubicación del dispositivo y los guarda en formato GPX; *Librería Q2M* (https://github.com/gispunpauarg/Q2M) [1]. Esta librería, desarrollada en un proyecto de investigación anterior, computa métricas QoS y QoE en aplicaciones Android; *Nexo* (https://github.com/gispunpauarg/Nexo) [2]. Esta herramienta, desarrollada en un proyecto de investigación anterior, sirve para visualizar gráficamente indicadores QoS y QoE a fines de facilitar el análisis de las relaciones que puedan existir entre dichos indicadores; *Mobile QoE exploration: an unsupervised field study in an Argentine Patagonian city* [3]. En este trabajo, llevado a cabo durante el año 2020, se estudió la relación entre la Calidad de Servicio (QoS) y la Calidad de Experiencia (QoE) del usuario mediante la recolección de indicadores haciendo uso de una librería implantada en una aplicación Android llamada CovidInfo (https://github.com/gispunpauarg/CovidInfoUNPA); *Android QoS SDK* (https://github.com/RestComm/android-QoS). Provee métricas sobre QoS/QoE (geolocalización, voz, video, mensajería…) y analytics para dispositivos Android; *Android Network measures* (https://github.com/APISENSE/android-network-measures). Brinda herramientas para llevar a cabo mediciones de red (QoS) en Android mediante el uso de utilidades como ping, traceroute y descargas/subidas por TCP/UDP; *QoE Probe for Android* (https://github.com/farnazfotrousi/QoE-Probe-Android). Se

trata de una aplicación móvil Android que se puede integrar con otras aplicaciones, a fines de recolectar información relevante para la QoS y la QoE.

3 Indicadores considerados

Tras investigar qué indicadores QoS relacionados con GPS pueden obtenerse dentro del contexto de una aplicación Android, y las limitaciones impuestas por este sistema operativo, se decidió implementar (dentro de la librería) la recolección de los indicadores presentados en la Tabla 1. Muchos de los indicadores QoS descritos en dicha tabla son considerados de importancia en la calidad del servicio de GPS por otros trabajos de investigación ([5], [6] y [7]).

Tabla 1. Indicadores recolectados por la librería.

Indicador	Descripción
Latencia	Tiempo (en milisegundos) que se tarda en obtener la actualización de ubicación del dispositivo móvil.
Satélites utilizados	Cantidad de satélites que se utilizan para determinar la ubicación del dispositivo móvil.
Precisión	Precisión radial © (en metros) estimada de la actualización de ubicación obtenida. Mientras menor sea el valor, mejor será la precisión.
Nubosidad	Porcentaje de nubosidad (de nubes en el cielo) en donde se registran las actualizaciones de ubicación.
Presión	Presión atmosférica (en hPa) en donde se registran las actualizaciones de ubicación.

Además de indicadores QoS, la librería desarrollada también recolecta la latitud, longitud y datos climáticos correspondientes a la ubicación en la que se encuentra el usuario del dispositivo móvil. Esto se hace ya que, según algunas investigaciones ([8], [9] y [10]), puede que exista una relación entre las condiciones climáticas y la precisión de las ubicaciones obtenidas por el GPS.

4 Librería Q2MGPS

4.1 Diagrama funcional

En la Figura 1 se expone brevemente el funcionamiento de la librería Q2MGPS.

Figura 1. Diagrama funcional de la librería Q2MGPS.

- **1:** El desarrollador invoca, desde el código de su aplicación, las funcionalidades de la librería para recolectar uno o más indicadores.
- **2:** La librería recolecta indicadores haciendo uso del GPS del dispositivo móvil y la API del clima de OpenWeather.
- **3:** La librería registra los indicadores recolectados en un documento XML, que se guarda en el almacenamiento del dispositivo móvil.
- **4:** El desarrollador puede analizar los indicadores recolectados consultando el documento XML previamente mencionado.

Cabe mencionar que la librería Q2MGPS fue construida sobre la librería Q2M[1], la cual fue desarrollada en un trabajo de investigación anterior.

4.2 Funcionalidades

Como se muestra en el diagrama (Figura 2), la librería Q2MGPS está comprendida por múltiples clases, las cuales cumplen distintos roles dentro de la misma. Sin embargo, las dos clases más importantes dentro de la librería son *Indicadores* y *ConstructorXML*. En la presente sección, se describirán brevemente las funcionalidades de dichas clases.

Figura 2. Diagrama de clases de Q2MGPS.

Indicadores La clase *Indicadores* es la responsable de recolectar todos los indicadores descritos previamente en la Tabla 1. Para funcionar correctamente, requiere del contexto[2] de la aplicación en la que se implanta la librería. Una vez se crea la instancia de la clase *Indicadores*, se comienzan a registrar automáticamente las actualizaciones de ubicación y los datos climáticos con los parámetros especificados por el desarrollador (cantidad de actualizaciones y tiempo entre actualizaciones).

ConstructorXML La clase *ConstructorXML* cumple con la tarea de registrar de manera organizada todos los indicadores recolectados por la clase *Indicadores* en un documento XML en el dispositivo móvil del usuario. Su funcionalidad es invocada desde la clase *Indicadores*, en los métodos que se encargan de computar los indicadores.

4.3 Documento XML

El documento XML creado y manipulado por la librería Q2MGPS tiene una estructura como la que se presenta a continuación.

[1]https://github.com/gispunpauarg/Q2M
[2]https://developer.android.com/reference/android/content/Context

```
    <indicador nombre="ClimaNubosidad" fecha="2021-08-25 12:00:01"
lat="0.0" lon="0.0">25%</indicador>
    [...]
</indicadores>
```

Cada vez que se recolecta un nuevo valor para un indicador, este se registra en el documento XML previamente mencionado, dentro de etiquetas <*indicador*>. Cada una de estas etiquetas cuenta con varios atributos, que describen la información del valor recolectado: *nombre:* Nombre del indicador para el cual se recolectó el valor; *fecha:* Fecha (*yyyy-MM-dd HH:mm:ss*) en la que se recolectó el valor; *lat, lon:* Ubicación (latitud y longitud) a la cual corresponde el valor registrado.

5 Casos de prueba

Para poner a prueba la librería desarrollada en este trabajo y para recolectar datos que permitan llevar a cabo conclusiones, Q2MGPS se implantó en una aplicación llamada UNPA Runner (Figura 3). Esta aplicación funciona de manera similar a Google Fit[3]; resumidamente, permite registrar (a demanda) los recorridos del usuario en tiempo real y, al finalizar, los almacena en una base de datos local SQLite. La librería Q2MGPS se implantó en esta aplicación de manera que, cada vez que se inicia un nuevo recorrido, se recolecten los indicadores y datos meteorológicos de interés y se almacenen en un documento XML.

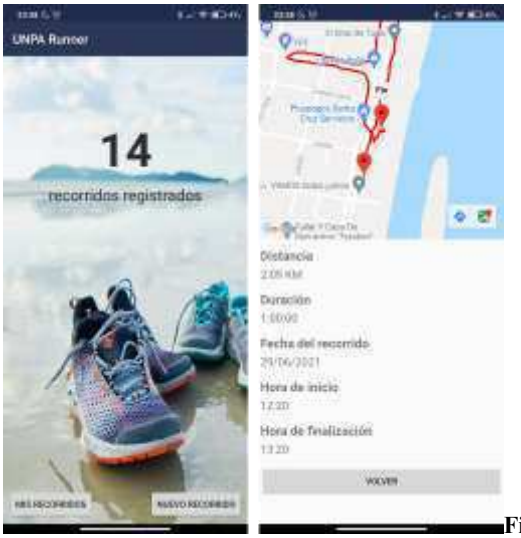


Figura 3. UNPA Runner.

[3]https://www.google.com/intl/es_es/fit

En total, se realizaron pruebas en diez dispositivos, con usuarios pertenecientes a tres ubicaciones diferentes (Río Gallegos, Comandante Luis Piedrabuena y Buenos Aires). Seis de los dispositivos son Samsung (modelos: Galaxy A01; Galaxy A30; Galaxy J7 Prime; Galaxy S8; J4 Core), tres Motorola (modelos: Moto E6 Plus; Moto G8; One Fusion) y uno Xiaomi (modelo: Redmi Note 8 Pro). De estos dispositivos, tres cuentan con Android 8 (Oreo), otros dos con Android 9 (Pie), cuatro con Android 10 (Q) y uno con Android 11 (R). Todas estas pruebas fueron llevadas a cabo entre Junio y Agosto del 2021.

5.1 Resultados

Una vez recuperados los datos generados por los usuarios detallados en la sección anterior, estos se sintetizaron (promediaron[4]) mediante la ejecución de un script PHP. Dicho script toma como entrada los documentos XML; los procesa y genera como salida la información resumida en formatos CSV y HTML.

Con la información generada por el script, se confeccionaron gráficos en LibreOffice[5] Calc para poder estudiar fácilmente los indicadores que recolectó la librería en los dispositivos partícipes de los casos de prueba. En las Figuras 4, 5 y 6, se presenta una comparación de indicadores GPS versus nubosidad[6]. Por otro lado, en las Figuras 7 y 8 se presenta una comparación de indicadores GPS versus presión atmosférica[7].

Figura 4. Gráfico para cielo soleado. **Figura 5.** Gráfico para cielo parcialmente nublado.

Figura 6. Gráfico para cielo nublado.

Figura 7. Gráfico para presión atmosférica alta.

Figura 8. Gráfico para presión atmosférica baja.

Cabe aclarar que los rangos (cielo *soleado*, *parcialmente nublado* y *nublado*; presión atmosférica *alta* y *baja*) utilizados para clasificar los indicadores fueron extraídos de [11], [12] y [13].

[4]En este caso, “promedio” es la *media aritmética*; la suma de los números dividida por cuántos números se promedian.
[5]https://www.libreoffice.org
[6]En el presente trabajo, un menor porcentaje de nubes opacas se considera como “mejor clima”.
[7]De acuerdo a [12] y [13], una presión atmosférica elevada equivale a buenas condiciones climáticas.

6 Conclusiones

Como se explicó en la Sección 3, en [8], [9] y [10] se sugiere una posible relación entre la calidad de los datos brindados por el servicio de GPS y las condiciones climáticas actuales en la ubicación en la que se encuentra el usuario del dispositivo. Observando la información recopilada por la librería Q2MGPS (presentada gráficamente en las Figuras 4, 5, 6, 7 y 8), se puede apreciar que las condiciones climáticas parecen no tener influencia alguna sobre los indicadores de calidad estudiados.

En lo que respecta a la *nubosidad*, cuando estuvo soleado (< 25% nubes), los valores promedio para la latencia, precisión radial y cantidad de satélites utilizados fueron 11.6 segundos, 39.2 metros y 7, respectivamente; cuando estuvo parcialmente nublado (25% ≤ nubes < 50%), 7.2 segundos, 15.1 metros y 12 y, cuando estuvo nublado (50% ≤ nubes), 7.1 segundos, 17.2 metros y 13. Estos datos sugieren que no existe una relación entre la cantidad de nubes en el cielo y la calidad del servicio de GPS.

En lo que respecta a la *presión atmosférica*, cuando fue mayor o igual a 1013 hectopascales, los valores promedio para la latencia, precisión radial y cantidad de satélites utilizados fueron 13.2 segundos, 34.3 metros y 10, respectivamente y, cuando fue menor a 1013 hectopascales, 7.3 segundos, 23.5 metros y 12 respectivamente. Al igual que con la nubosidad, estos datos no sugieren que exista relación entre la presión atmosférica y la calidad del servicio de GPS, dado a que, de hecho, se obtuvieron mejores resultados cuando las condiciones climáticas fueron peores.

También, es relevante destacar que se obtuvo una menor latencia y una mejor precisión radial cuando la cantidad de satélites utilizados para la fijación de la ubicación fue mayor.

Pese a la información obtenida mediante la realización de este trabajo, no se considera apropiado llevar a cabo afirmaciones, ya que puede que estos resultados se deban a que las pruebas fueron realizadas desde smartphones, y haciendo uso de las herramientas proporcionadas por el sistema operativo Android, las cuales tienen sus respectivas limitaciones. Por estas razones, se considera adecuado investigar esta temática más a fondo en trabajos futuros.

Referencias

1. Machini, A., Enriquez, J., & Casas, S. (2019). *Q2M, una librería para computar métricas de calidad en aplicaciones móviles*. Informes Científicos Técnicos-UNPA, 11(2), 1-17.
2. Machini, A., Enriquez, J., & Casas, S. (2020). *Nexo: Una herramienta para la visualización y análisis de indicadores QoS y QoE móviles*. Informes Científicos Técnicos-UNPA, 12(2), 47-62.
3. Garcia, A. C., & Casas, S. (2020). *Mobile QoE exploration: an unsupervised field study in an Argentine Patagonian city*. Presentado en 2020 39th International Conference of the Chilean Computer Science Society (SCCC) (pp. 1-7). IEEE.
4. Yeh, T. K., Liou, Y. A., Wang, C. S., & Chen, C. S. (2008). *Identifying the degraded environment and bad receivers setting by using the GPS data quality indices*. Metrologia, 45(5), 562.

5. Machaj, J., Brida, P., & Majer, N. (2012). *Novel criterion to evaluate QoS of localization based services*. Presentado en Asian Conference on Intelligent Information and Database Systems (pp. 381-390). Springer, Berlin, Heidelberg.
6. Filjar, R., Bušić, L., & Pikija, P. (2008). *Improving the LBS QoS through Implementation of QoS Negotiation Algorithm*. Croatia, Zagreb, 4.
7. Filjar, R., Bušić, L., Dešić, S., & Huljenić, D. (2008). *LBS position estimation by adaptive selection of positioning sensors based on requested QoS*. Presentado en International Conference on Next Generation Wired/Wireless Networking (pp. 101-109). Springer, Berlin, Heidelberg.
8. Álvarez Pacheco, J. G. (2019). *Analizar los efectos de la tropósfera sobre la señal de GPS y el impacto en la precisión en el posicionamiento de un receptor* (Master's thesis, Escuela Superior Politécnica de Chimborazo).
9. Solheim, F. S., Vivekanandan, J., Ware, R. H., & Rocken, C. (1999). *Propagation delays induced in GPS signals by dry air, water vapor, hydrometeors, and other particulates*. Journal of Geophysical Research: Atmospheres, 104(D8), 9663-9670.
10. Yeh, S. C., Hsu, W. H., Su, M. Y., Chen, C. H., & Liu, K. H. (2009). *A study on outdoor positioning technology using GPS and WiFi networks*. Presentado en 2009 International Conference on Networking, Sensing and Control (pp. 597-601). IEEE.
11. *Appendix A: Anatomy of a Zone Forecast* (p. 1). National Weather Service. Recuperado el 6 de Agosto del 2021, de https://www.weather.gov/media/pah/ServiceGuide/A-forecast.pdf.
12. *Las zonas de alta/baja presión*. meteoblue. Recuperado el 6 de Agosto del 2021, de https://content.meteoblue.com/es/meteoscool/el-clima-a-gran-escala-lsw/alta-baja-presion.
13. *Altas y bajas presiones*. Wikipedia. Recuperado el 6 de Agosto del 2021, de https://es.wikipedia.org/wiki/Altas_y_bajas_presiones.

ISBN 978-987-633-574-4

CACIC 2021

WORKSHOP PROCESAMIENTOS DE SEÑALES Y SISTEMAS DE TIEMPO REAL

COORDINADORES

Horacio Villagarcia Wanza (UNLP)
Emanuel Frati (UNdeC)
Jorge Ierache (UM)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

CACIC 2021 UNSa

Predicción del impacto de la vacunación. Una aproximación desde la simulación

Federico Montes de Oca[1], Diego Luparello[1], Diego Fretes[1], Julian Ifran[1], Román Bond[1], Martín Morales[1, 2], Diego Encinas[1, 3].

[1]SimHPC-TICAPPS. Universidad Nacional Arturo Jauretche. Florencio Varela, 1888, Argentina.
[2]Centro CodApli. FRLP. Universidad Tecnológica Nacional. La Plata, 1900, Argentina.
[3]Instituto de Investigación en Informática (III-LIDI). Facultad de Informática, Universidad Nacional de La Plata - Centro Asociado CIC. La Plata, 1900, Argentina.
{federicomdo97,compu.diego94,djoaquin.fretes,shuloifran}@gmail.com, {rbond,martin.morales,dencinas}@unaj.edu.ar

Resumen. El objetivo de este trabajo es modelar y simular escenarios epidemiológicos y analizar el impacto de la vacunación como medida anti-epidémica. Se toma como punto de partida un modelo base existente que ya contempla la propagación de una enfermedad. Se extiende el modelo original contemplando el porcentaje de efectividad de la vacuna aplicada a la población. Para lograr un análisis entre individuos y los cambios de estados que éstos sufren durante una epidemia se utiliza el Modelado y Simulación Basado en Agentes (ABMS; Agent-Based Modeling and Simulation). En esta aproximación se busca dar soporte a la toma de decisiones con predicciones acerca del impacto de vacunas, realizar pruebas con la enfermedad COVID-19 utilizando el mismo modelo para obtener predicciones y finalmente, concientizar a la población acerca de la importancia de la vacunación como prevención y barrera contra brotes epidemiológicos.

Palabras Clave: Simulación, Vacunación, ABMS, Epidemiología.

1 Introducción

Históricamente la humanidad se ha preguntado cómo prevenir las epidemias y la manera de detener su avance. Es complejo decidir políticas puntuales que las erradiquen. El contexto de cada población es distinto: demografía, cultura, recursos, tecnología, sistema sanitario, etc.

Aquí toman importancia los modelos computacionales de escenarios epidemiológicos. Con una simple parametrización de un simulador, se pueden analizar los impactos de políticas de Salud Pública (aislamiento, vacunación, extensión del sistema sanitario, entre otros) en la batalla contra la propagación de una enfermedad.

Simular es el proceso de diseñar un modelo computarizado de un sistema (o proceso) y realizar experimentos con este modelo con el propósito de comprender el comportamiento del sistema y/o de evaluar varias estrategias para el funcionamiento de este [1].

1.1 Objetivos

- Adaptar un modelo existente para que contemple el efecto de la vacunación a través del porcentaje de efectividad de la vacuna.
- Dar soporte a la toma de decisiones con predicciones acerca del impacto de vacunas.
- Poder realizar pruebas con la enfermedad COVID-19 utilizando el mismo modelo para obtener predicciones.
- Concientizar a la población acerca de la importancia de la vacunación como prevención y barrera contra brotes epidemiológicos.

1.2 Modelado y Simulación Basado en Agentes (ABMS)

Es un tipo de modelo computacional que posibilita la simulación de acciones e interacciones de individuos autónomos dentro de un entorno, y permite determinar qué efectos producen en el conjunto del sistema. Los modelos simulan las operaciones simultáneas de entidades múltiples (agentes) en un intento de recrear y predecir las acciones de fenómenos complejos [2].

1.3 Modelo SIR

Es un modelo matemático epidemiológico sencillo. Fue descrito en 1927 por el bioquímico William Ogilvy Kermack y el epidemiólogo Anderson Gray McKendrick [3]. El modelo recibe este nombre porque divide la población en 3 conjuntos: Susceptibles, Infectados y Recuperados.

La población inicial está compuesta por individuos susceptibles de contraer la enfermedad, los individuos infectados cambian de estado e infectan a otros. Al pasar los días, los infectados se irán recuperando y cambiarán de estado. El proceso continúa hasta que no quedan infectados y la población quedará dividida entre Susceptibles (que no se infectaron) y Recuperados [4].

Restricciones del modelo SIR.

- La epidemia transcurre tan rápido que no es significante tener en cuenta los nacimientos ni las defunciones
- Si un individuo sano se infecta, se vuelve infeccioso de inmediato.
- Si un individuo se recupera adquiere inmunidad y no puede volver a contraer el virus.

1.4 Modelo epiDEM Travel & Control.

Es un modelo basado en SIR e implementado en NetLogo [5]. Simula la propagación de una enfermedad infecciosa en una población semicerrada, pero con características adicionales como viajes, aislamiento, cuarentena, inoculación y vínculos entre individuos. Sin embargo, se continúa asumiendo que el virus no muta y que, tras la recuperación, una persona tendrá una inmunidad perfecta [6].

El número de reproducción R0.

Al final de la simulación, el R0 refleja la estimación del número de reproducción, la relación de tamaño final que indica si habrá una epidemia

$$R_0 = \frac{\beta \cdot S(0)}{\gamma} = N \cdot \frac{ln\left(\frac{S(0)}{S(t)}\right)}{N - S(t)} \qquad (1)$$

- N es la población total
- S(0) es el número inicial de susceptibles
- S(t) es el número total de susceptibles en el tiempo t.
- β es la tasa de contagio diaria
- γ es la inversa del tiempo de infección

En este modelo, la estimación R0 es el número de infecciones secundarias que surgen para un individuo infectado promedio durante el transcurso del período infectado de la persona.

2 Aportes del modelo

Aunque el modelo epiDEM contempla aspectos importantes como la propagación, el mismo no modela la mortalidad.

Teniendo en cuenta que el objetivo del trabajo es analizar el impacto de la vacunación contra epidemias, introducimos los siguientes aportes al modelo:

- *Initial-people-infected-chance*: permite aumentar el número de infectados iniciales.
- *Average-recovery-time-hospitalized*: permite decidir el promedio de tiempo que un individuo pasa hospitalizado.
- *Mortality-chance*: permite decidir la tasa de mortalidad de los individuos infectados.
- *Vaccine-efficacy*: permite decidir el porcentaje de eficacia de la vacuna.

***Initial-people-infected-chance*.**

El modelo epiDEM fija que un individuo tiene 5% de probabilidad de comenzar la simulación infectado. Se parametriza ese porcentaje para poder analizar escenarios donde la enfermedad ya está propagada en la población. La variable es un porcentaje que va desde el 1 hasta el 50. Por lo tanto, como máximo la mitad de la población podrá iniciar la simulación infectada.

Average-recovery-time-hospitalized.

EpiDEM asume que todo individuo infectado hospitalizado se recupera cinco veces más rápido que un individuo no hospitalizado.

A partir del análisis de publicaciones [7][8] acerca de la permanencia en hospitales de pacientes con Influenza A, se observa que no se puede afirmar que los hospitales aceleren la recuperación de los pacientes.

Esta variable es un número que representa el promedio de tiempo en el cual un individuo infectado y hospitalizado se recupera o muere. El rango de dicho promedio va desde 1 hasta 1000 unidades de tiempo.

Vaccine-efficacy.

La vacunación del modelo epiDEM resulta muy elemental: si una persona es vacunada, se vuelve inmune al 100%. Por esto, se parametriza la inmunidad adquirida por la vacuna a aplicar, introduciendo el parámetro *vaccine-efficacy* que es un porcentaje del 1 al 100.

De esta forma, se puede simular distintos escenarios utilizando varios porcentajes de eficacia de vacunas. En el uso, se calcula el promedio ponderado de eficacia de las vacunas a aplicar, ingresando el resultado como *vaccine-efficacy* en la simulación.

Mortality-chance.

Ni el modelo SIR ni el modelo epiDEM contemplan fallecimientos.

Dado que uno de los objetivos es concientizar a las personas, resulta importante poder modelar las muertes causadas por la enfermedad, que es la consecuencia que más relevancia tiene para la humanidad. De esta manera, las personas podrán visualizar el impacto positivo de la vacunación.

Se introduce el parámetro *mortality-chance*. Es un porcentaje del 1 al 100 que representa la letalidad de la enfermedad. Los individuos fallecidos se representan en color gris.

Dado que, en el modelo SIR, el conjunto de Recuperados representa a su vez a los recuperados y los muertos, *mortality-chance* generará un subconjunto D de muertos (*deaths*):

$$D = \{i \in R \mid i \notin S; i \notin I\} \quad (2)$$

$$|D| = (MortalityChance \div 100) \cdot |R| \quad (3)$$

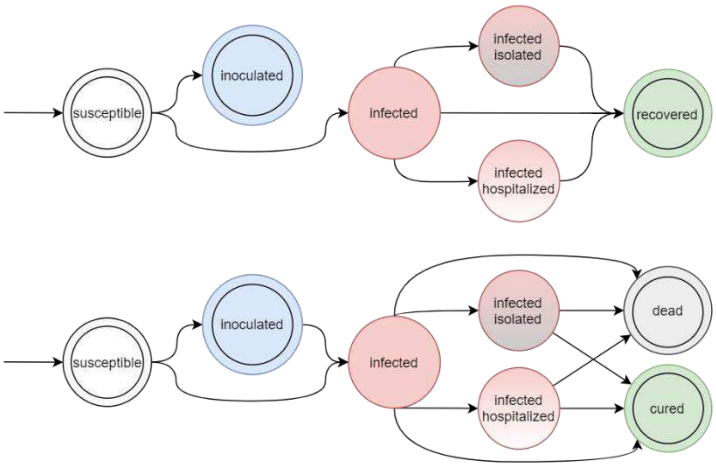


Fig. 1. Contraste de diagramas de estados del modelo epiDEM (primer diagrama) contra el modelo extendido (segundo diagrama)

Por un lado, se permite que los individuos inoculados se infecten según la eficacia de la vacuna.

Se "separa" el estado *recovered* en los estados *dead* y *cured*. En los cálculos finales, el modelo sigue tomando a estos dos grupos como uno solo, para mantener el mismo cálculo de R0.

2.1 Restricciones

El modelo resultante presenta las siguientes restricciones relevantes:

- No contempla nacimientos.
- Solo contempla defunciones relacionadas a la epidemia simulada.
- La eficacia de la vacuna solo afecta la posibilidad de ser infectado.
- El porcentaje de infectados inicial no puede ser mayor al 50%
- El porcentaje de inoculados inicial no puede ser mayor al 50%
- No existe la posibilidad de volver a infectarse con el virus una vez recuperado.

3 Simulación

3.1 Número de ejecuciones

Es necesario obtener la cantidad óptima de ejecuciones a realizar del escenario propuesto. Para esto se tiene en cuenta el teorema de Chebyshev [9] utilizando un intervalo de confianza del 95%. La distribución correspondiente es la estándar.

Entonces:

$$M = \frac{\sigma^2 \cdot \left(Z_{\alpha/2}\right)^2}{\kappa^2} = 15{,}3664 \simeq 15 \quad (4)$$

Es decir, se realizan 15 ejecuciones sobre el mismo escenario para obtener una salida estacionaria y así poder realizar un análisis correcto.

3.2 Escenarios

Para los escenarios de simulación, se fijan parámetros sustraídos de informes y estudios actuales sobre la enfermedad COVID-19 en el contexto del Territorio Nacional Argentino [10][11][12]. Sin embargo, algunos parámetros son fijados en base al análisis de estudios realizados en el exterior [13][14].

- *initial-people*: 1000. Terreno con máxima población posible.
- *initial-people-infected-chance*: 4%. En Buenos Aires, en la Semana Epidemiológica 20 2021 hubieron 225K casos [12], 33K casos en promedio al día, Si se está 5 días infeccioso [13], entonces, en promedio, hubieron 33K*5=165K personas contagiando el virus a la vez en esa semana. Siendo el 3.66% de la población argentina de 45M. Esto es una burda aproximación.
- *infection-chance*: 30%. Infecciosidad elevada. No existen muchos datos que aporten certeza acerca de este parámetro, por lo que se lo selecciona de forma adrede.
- *recovery chance*: 95% para que los individuos cambien de estado cerca de las fechas promedio de *recovery-time*.
- *average-recovery-time*: 5 días - 120 hs [13]
- *average-recovery-time-hospitalized*: 8 dias - 192 hs [11]
- *intra-mobility*: 0.8. Movilidad mínimamente reducida.
- *links*: Encendido. Se simulan contactos estrechos.
- *travel & travel-tendency*: Apagado. Se observa un comportamiento no esperado: los individuos se juntan demasiado en el límite al activar dichas variables.
- *average-isolation-tendency*: 25%. Se asume que la capacidad de diagnóstico temprano no es muy eficiente.
- *average-hospital-going-tendency*: 15% [14]
- *mortality-chance*: 3% (letalidad acumulada en la Pcia. de Bs. As. 2,64%) [10].
- *initial-ambulance*: 0. No hay relevancia.

Los parámetros *inoculation-chance* y *vaccine-efficacy* son variados luego para analizar los resultados.

Fig. 2. Configuración del escenario en el simulador

- **Escenario 1**: Escenario sin vacunación: *inoculation-chance* en 0.
- **Escenario 2**: Escenario con vacunación. El parámetro *vaccine-efficacy* se configura al 60% y el parámetro *inoculation-chance* se configura al 50% para que la mitad de la población esté vacunada. Análogo sobre la aplicación de una primera dosis.
- **Escenario 3**: Se incrementa la eficacia de la vacuna al 90%, manteniendo *inoculation-chance* en 50%. Análogo a la aplicación de una segunda dosis.

3.3 Resultados

Como se observa en la Fig. 3, a medida que la eficacia de la vacuna aumenta, suceden menos contagios. En el primer escenario, prácticamente toda la población contrajo la enfermedad, haciendo que la misma mate aproximadamente 3% de la población total.

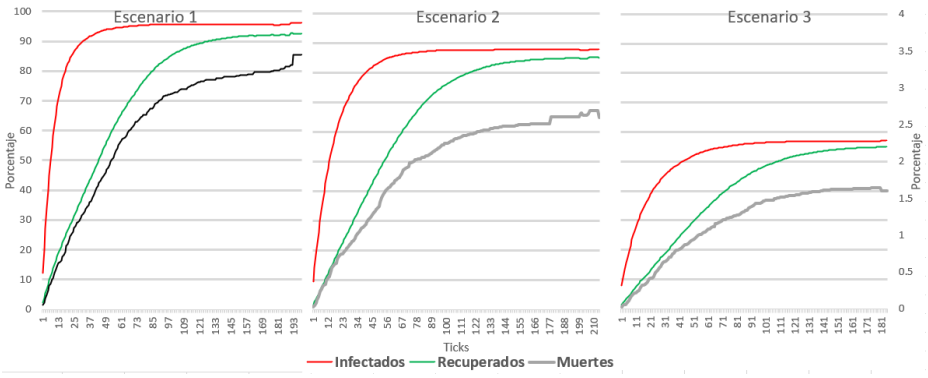


Fig. 3. Comparativo de Infectados, Recuperados y Muertos Acumulativos (Muertos en el eje vertical secundario)

En la Fig. 4, se observa cómo en el primer escenario la curva de infectados aumenta de forma exponencial. Esta suele ser una de las razones por las cuales los sistemas de

salud colapsan. En los escenarios 2 y 3, esta curva es aplanada. Además, es posible observar que menos personas se contagian al final de la simulación. La cantidad de muertos se reduce drásticamente.

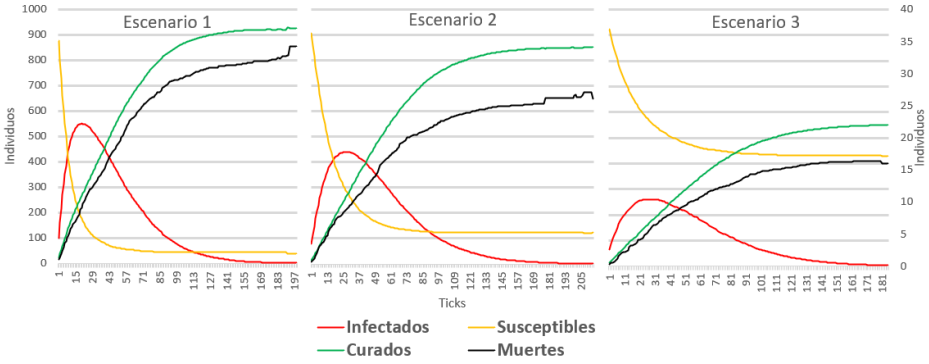


Fig. 4. Comparativo de estados de los individuos (Muertes en el eje vertical secundario)

En la Fig. 5, se observa cómo la Tasa de Infección se reduce drásticamente a medida que la eficacia de la vacuna aumenta.

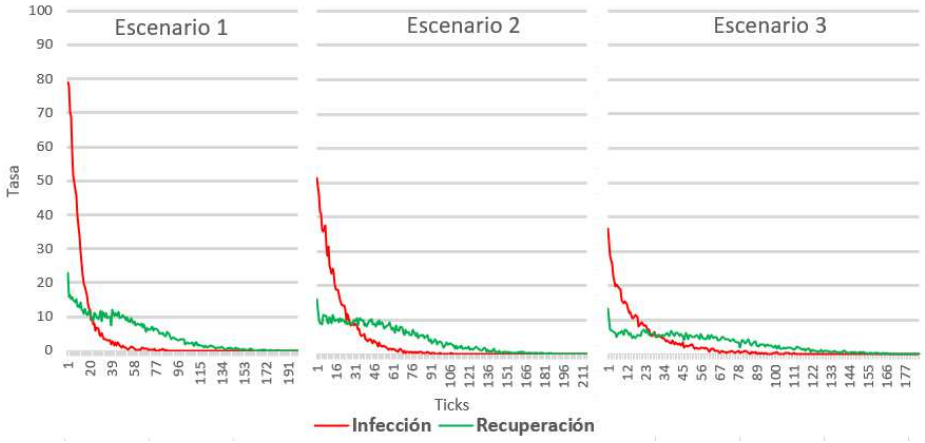


Fig. 5. Comparativo de Tasas de Infección y de Recuperación

4 Conclusiones

En base a la investigación, análisis y desarrollo realizado, se muestra la importancia de generar modelos y simuladores como herramienta para medir el impacto de los planes de vacunación. Además, se consigue parametrizar una amplia gama de variables relevantes para llegar a una predicción.

Luego, en el caso particular de las pruebas realizadas para los escenarios propuestos, se comprueba que, la relación entre la velocidad de propagación de una enfermedad

puede ser drásticamente desacelerada con un plan progresivo de vacunación. Esto sin tener en cuenta otros tipos de medidas anti epidémicas. Además, se comprueba que, gracias a la inoculación, se puede lograr que una gran parte de la población nunca contraiga la enfermedad. Esto es importante ya que muchas enfermedades pueden llegar a dejar secuelas.

Cómo futuros trabajos, se propone mejorar la precisión del impacto de la vacunación sobre la mortalidad incluyendo más modificadores. Si un individuo se infecta a pesar de estar vacunado, un modificador podría reducir la posibilidad de muerte.

También, se podría disminuir la posibilidad de hospitalización en aquellos individuos que hayan sido vacunados.

Luego, podría aplicarse esto mismo para que la vacuna también afecte a la capacidad de infectar a otros de un individuo que, a pesar de haber sido vacunado, contrajo la enfermedad.

También se debería poder flexibilizar el parámetro de población vacunada al inicio de la simulación. El parámetro está seteado a un máximo de 50%, esto es para no generar conflicto con el parámetro *Initial-people-infected-chance* que también va hasta el 50% de la población.

5 Bibliografía

1. Simulation Modeling and Methodology. Robert E. Shannon. The University of Alabama Hunstville, Hunstville Alabama. Winter Simulation Conference. 1976.
2. Bonabeau, E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences, 99(Supplement 3), pp.7280-7287.
3. On the Mathematical Interpretation of Epidemics by Kermack and McKendrick.Raúl Isea and Karl E. Lonngren. Gen. Math. Notes, Vol. 19, No. 2, December, 2013, pp. 83-87 ISSN 2219-7184; Copyright © ICSRS Publication, 2013.
4. The SIR model and the Foundations of Public Health. Howard (Howie) Weiss. MATerials MATemàtics, Volum 2013, treball no. 3, 17 pp. ISSN: 1887-1097
5. Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Centro para el aprendizaje conectado y el modelado basado en computadora, Northwestern University, Evanston, IL.
6. Yang, C. y Wilensky, U. (2011). Modelo de viaje y control NetLogo epiDEM. http://ccl.northwestern.edu/netlogo/models/epiDEMTravelandControl. Centro para el aprendizaje conectado y el modelado basado en computadora, Northwestern University, Evanston, IL.
7. Análisis descriptivo de los casos de Gripe A (h1n1) notificados durante la pandemia de 2009 en la Región Sanitaria de la Provincia de Buenos Aires, Argentina. Silvina Busto - Fernanda Bonet - Adriana Alberti. Rev. Argentina Salud Pública, Vol 1, N°3 Junio 2010
8. Fajardo-Dolci, G., Hernández-Torres, F., Santacruz-Varela, J., Rodríguez-Suárez, J., Lamy, P., Arboleya-Casanova, H., Gutiérrez-Vega, R., Manuell-Lee, G. and

Córdova-Villalobos, J., 2009. Perfil epidemiológico de la mortalidad por influenza humana A (H1N1) en México. Salud Pública de México, 51(5).
9. Shannon, R.E., Systems Simulation: The Art and Science, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.
10. Portal-coronavirus.gba.gob.ar. 2021. Sala de Situación | EMERGENCIA SANITARIA. [online] Available at: <https://portal-coronavirus.gba.gob.ar/es/sala-de-situacion> [Accedido en Agosto 2021].
11. Gba.gob.ar. 2021. Informe de Rendimiento Hospitalario UCI Adultos | Provincia de Buenos Aires. [online] Available at: <https://www.gba.gob.ar/saludprovincia/informes_de_gestion/informe_de_rendimiento_hospitalario_uci_adultos> [Accedido en Agosto 2021].
12. Argentina.gob.ar. 2021. Sala de situación. [online] Available at: <https://www.argentina.gob.ar/coronavirus/informes-diarios/sala-de-situacion> [Accedido en Agosto 2021].
13. Cevik, M., Tate, M., Lloyd, O., Maraolo, A., Schafers, J. and Ho, A., 2021. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. The Lancet Microbe, 2(1), pp.e13-e22.
14. Gomez, J., Du-Fay-de-Lavallaz, J., Fugar, S., Sarau, A., Simmons, J., Clark, B., Sanghani, R., Aggarwal, N., Williams, K., Doukky, R. and Volgman, A., 2021. Sex Differences in COVID-19 Hospitalization and Mortality. Journal of Women's Health, 30(5), pp.646-653.

ISBN 978–987–633–574–4

Prototipo de controlador MIDI biomecánico para uso en sintetizadores virtuales

Fernando Andrés Ares [1], Matías Presso[1, 2, 3], Claudio Aciti[1, 2]
[1] Universidad Nacional de Tres de Febrero, Sede Caseros I.
Valentín Gómez 4828, Caseros (B1678ABJ), Buenos Aires, Argentina
[2] Universidad Nacional del Centro de la Provincia de Buenos Aires.
Pinto 399, Tandil (7000), Buenos Aires, Argentina
[3] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires
Calle 526 entre 10 y 11, La Plata (1900), Buenos Aires, Argentina
{ferares4@gmail.com, matiaspresso@gmail.com, caciti@exa.unicen.edu.ar}

Resumen. El presente trabajo consiste en el diseño, desarrollo e implementación de un prototipo de dispositivo biomecánico de bajo costo capaz de controlar, con movimientos corporales, distintos tipos de sintetizadores o instrumentos virtuales que existen actualmente en la industria musical utilizando el protocolo MIDI. Durante el desarrollo del trabajo se presentan distintos aspectos que involucran, la elección de componentes de hardware, la fabricación de una placa de circuito impreso y de una carcasa para el montaje de la misma, así como también el desarrollo de una aplicación para el testeo del dispositivo a través de una PC.

Palabras Clave: Controlador MIDI, Biomecánica, movimiento corporal, sintetizador, instrumentos musicales virtuales.

1 Introducción

Desde de la década del ‘80 hasta la actualidad, se han fabricado todo tipo de instrumentos y controladores que utilizan el protocolo MIDI (Musical Instrument Digital Interface) [1] para diversas funcionalidades. Sin embargo, la gran mayoría de estos dispositivos son desarrollados en formato de pianos u otro tipo de instrumento con teclas, botones, deslizadores u otros mecanismos. No abundan en el mercado actual dispositivos de estas características que sean accionados exclusivamente con movimientos corporales.
El protocolo MIDI es un estándar de la industria musical que se utiliza para conectar instrumentos musicales digitales, computadoras y diversos dispositivos móviles.
Los inicios del protocolo MIDI, tal como lo conocemos hoy, son atribuidos a Dave Smith y Chet Wood [2] en colaboración con distintas empresas del mercado en aquel momento, entre las que se menciona a Roland, Yamaha, Korg, Kawai entre otras.
Esta interfaz, fue una solución a un problema común surgido a partir del auge de los sintetizadores analógicos que pretendía generar un lenguaje en común para la

interconexión de este tipo de instrumentos, que hasta el momento, eran producidos con distintos estándares de acuerdo al fabricante.
La primera versión del protocolo MIDI fue anunciada al público en el año 1982, y fue recién a fines del mismo año en que apareció el primer instrumento con una implementación de dicho protocolo.
A lo largo de los años siguientes, el protocolo ha ido evolucionando ya sea con mejoras en su especificación o nuevas funcionalidades para adaptarse a las tecnologías contemporáneas hasta llegar al día de hoy, donde dicho protocolo no solo se utiliza para la interconexión de dispositivos sino que representa un lenguaje en común para la creación, interpretación y comunicación musical de forma digital.
Existen empresas como Holonic Systems [3], y proyectos como "Making music with your muscles!" [4], "A MIDI Controller based on Human Motion Capture"[5], "Wireless Midi Controller Glove" [6], "Glove MIDI Controller" [7] y "KAiKU Glove Wearable MIDI Controller - S" [8] que abordan la temática pero con una perspectiva y propósitos diferentes al del presente trabajo.

2 Motivación, Objetivos y Alcances

En general resulta difícil encontrar un controlador MIDI que funcione exclusivamente con movimientos corporales en el mercado actual, y los productos existentes son de difícil alcance ya sea por costo, flexibilidad, o por dificultades en su adquisición. La motivación de este trabajo es construir un dispositivo de bajo costo con materiales o componentes de fácil acceso.
El objetivo del trabajo es desarrollar un dispositivo que permita generar nuevas formas de interpretación y creación musical a través del uso de la mecánica corporal. Se busca de esta manera enriquecer el espectro de la música actual a través de un cambio de paradigma en la ejecución de un instrumento virtual. Para ello se proponen los siguientes objetivos específicos:

1. Diseñar un dispositivo de hardware que cuente con un sensor de movimiento en tres dimensiones
2. Desarrollar un mecanismo para traducir los movimientos capturados por el dispositivo de hardware al protocolo MIDI
3. Desarrollar un mecanismo para transmitir la señal MIDI a través de un medio inalámbrico hacia una PC
4. Diseñar una carcasa para montar el dispositivo al reverso de la mano
5. Testear el dispositivo utilizando software de apoyo que simplifique la conexión inalámbrica entre el mismo y una PC

El alcance del trabajo está delimitado al diseño del dispositivo y el desarrollo de funcionalidades básicas con ciertos parámetros definidos que puedan ser testeados en un prototipo. No incluye el desarrollo de un driver específico para el manejo del hardware desde una PC así como tampoco el manejo de puertos virtuales para su uso desde cualquier aplicación MIDI. Las condiciones de evaluación del prototipo fue en un entorno de trabajo con PC y sistema operativo Windows 10, intensidad alta de la señal, y sensores dentro de un entorno libre de interferencia.

3 MIDI

MIDI es un protocolo estándar de comunicación en serie de la industria musical diseñado para la interconexión de instrumentos y dispositivos musicales. La información se transmite en forma de mensajes codificados de manera binaria, que pueden considerarse como instrucciones que indican a un dispositivo, como ejecutar una pieza musical con ciertos parámetros definidos por el mismo formato.
Los dispositivos MIDI pueden transmitir, recibir y retransmitir mensajes de acuerdo a la naturaleza y función del dispositivo en particular. Algunos dispositivos, están diseñados únicamente para crear y transmitir instrucciones MIDI y puedan conectarse a otro dispositivo esclavo, con capacidad de reproducir sonido, para que reciba estos mensajes y si pueda generarse el sonido. A este tipo de dispositivos que emiten órdenes para la ejecución de instrucciones se los conoce como controladores MIDI.

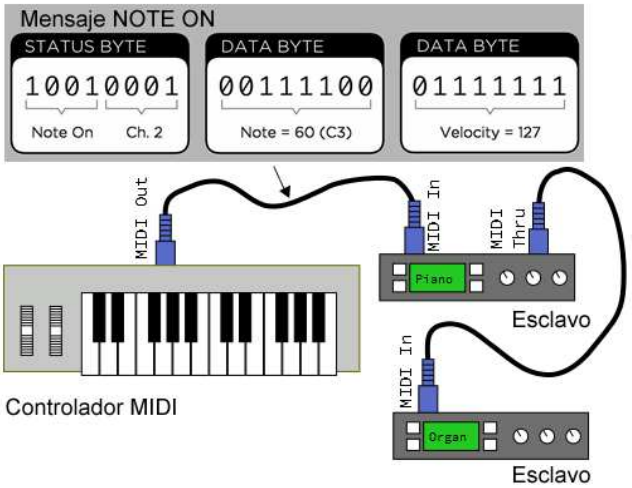


Fig. 1. Conexiones, controlador y mensajes MIDI

4 Descripción del Problema

El problema a abordar en este trabajo es el de construir un dispositivo capaz de traducir los movimientos de una mano en sonidos, a través de un sintetizador virtual instalado en una computadora con sistema operativo Microsoft Windows. El mismo debe ser construido utilizando materiales de bajo costo y de fácil acceso en el mercado. Además debe ser capaz de conectarse de forma inalámbrica a una computadora para transmitir los datos asociados a cada movimiento, será imprescindible que esa comunicación se lleve a cabo implementado el estándar MIDI para ser soportado por cualquier sintetizador virtual disponible en la actualidad.

Se deberán contemplar tres aspectos para el diseño y la construcción del dispositivo. En primera instancia deberá diseñarse y testearse una solución de hardware para interconectar todos los componentes necesarios para que el dispositivo funcione. Deberá también diseñarse un soporte para que este pueda ser montado sobre una mano. Finalmente deberá desarrollarse una solución de software capaz de traducir los movimientos de la mano en mensajes del tipo MIDI.
El software deberá incluir un mecanismo de calibración para asegurarse que los datos obtenidos sean precisos, y se envíen los mensajes MIDI en tiempos válidos en un contexto de tiempo real para aplicaciones de audio.
Al ser un dispositivo inalámbrico deberá alimentarse con baterías, las cuales deben brindar una autonomía razonable y la posibilidad de optar por baterías del tipo recargables.

5 Solución Propuesta

La solución consiste en tres aspectos fundamentales, el primero está centrado en el diseño y la construcción de una placa de montaje para todos los componentes de hardware y su respectivo circuito de alimentación. El segundo aspecto será diseñar y fabricar una carcasa para atornillar la placa y montar sobre la mano, debiendo contener también la batería y cualquier otro componente que se integre al dispositivo.
Finalmente se desarrollará una aplicación que traduzca el movimiento de la mano en mensajes del tipo MIDI que luego serán interpretadas por un instrumento virtual que se esté ejecutando en una PC.

5.1 Componentes del Sistema

El diagrama de bloques en la Fig. 2. muestra una visión de conjunto de cómo se interconectan los distintos componentes del dispositivo, así como también, la conexión externa a una PC. A su vez se detallan los componentes de hardware del dispositivo prototipo que fueron utilizadas y las herramientas de software empleadas en la computadora.

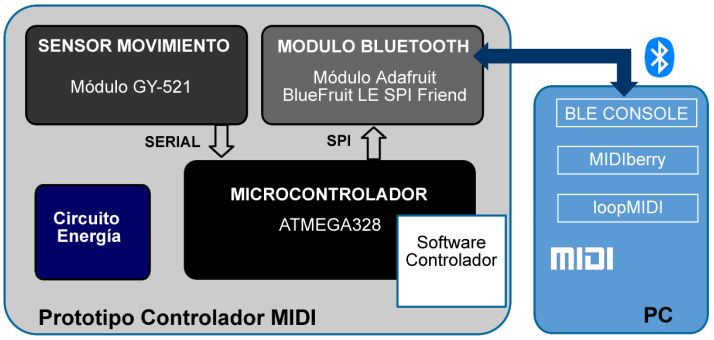


Fig. 2. Diagrama de bloques del sistema

En cuanto al hardware del prototipo del dispositivo controlador, el diseño funcional se divide en cuatro bloques principales cada uno con un propósito específico: detección de movimiento, manejo de energía, comunicación y procesamiento.

Para poder capturar los movimientos en tres dimensiones de una mano, será imprescindible contar con un sensor capaz de generar en tiempo real, una representación de estas variaciones en relación a la posición o a la rotación de dicha extremidad. Entre las variantes que se pueden encontrar hoy en día, resulta imprescindible para el propósito de este trabajo, elegir un sensor de bajo costo y lo suficientemente confiable para evitar el ruido en la medición. El sensor MPU-6050 [9] de InvenSense, combina las características mencionadas. Está compuesto por un acelerómetro en tres ejes que sensa aceleración gravitacional, un giróscopo en tres ejes que mide velocidad rotacional junto con un Procesador Digital de Movimiento (DMP) integrado, el cual simplifica el manejo de cálculos para combinar los datos medidos en tiempo real. Entre los módulos más convencionales el elegido para el prototipo fue el GY-521 [10].

Un aspecto fundamental para este dispositivo será la comunicación entre éste y una computadora. Para una mayor comodidad y portabilidad, es ideal que la comunicación se establezca de forma inalámbrica, para ello la tecnología Bluetooth, resulta un estándar global adecuado. Una de las mejoras más recientes de esta tecnología es el denominado Bluetooth de baja energía o BLE. Existen diversos tipos de módulos de comunicación inalámbrica que utilizan esta tecnología, se evaluaron distintos modelos para este proyecto y el elegido fue el módulo Adafruit Bluefruit SPI [11], que establece una comunicación sincrónica y serial para la transmisión de los datos. Posee ventajas sustanciales debido a que el firmware está íntegramente desarrollado para este módulo, lo cual lo vuelve más confiable, robusto y fácil de configurar, y además permite comunicarse a través del protocolo SPI.

Para el procesamiento y ejecución del programa del dispositivo se eligió el microcontrolador ATMEGA328P [12] por ser un chip de bajo consumo, de bajo costo y perfectamente accesible en el mercado para este proyecto. Cuenta con la cantidad de entradas y salidas compatible con los sensores y para futuras expansiones del dispositivo. Además posee herramientas avanzadas para su programación.
En cuanto a la energía, se diseñó un circuito de alimentación, utilizando una batería externa lo más pequeña posible y con una razonable autonomía. Se realizaron mediciones de consumo del dispositivo con un uso normal, las cuales mostraron oscilaciones de corriente en un rango de 49 y 51 mA. Para una batería convencional de 9 volts, y una capacidad aproximada de 500 mAh, se estima una autonomía aproximada de 10 horas.
La conexión entre el dispositivo y una PC requiere de software de apoyo. Se emplearon tres herramientas de software con distintos propósitos: BLE Console [13] para conectar el módulo bluetooth Bluetooth Adafruit BlueFruit LE SPI Friend al sistema operativo, loopMIDI [14] para crear un puerto MIDI virtual y MIDIBerry [15] para mapear Inputs y Outputs desde el módulo al puerto virtual.

5.2 Captura del movimiento

Para poder sensar los movimientos de una mano en tres dimensiones, se utiliza la metodología yaw, pitch, roll, o ypr, mostrada en la Fig. 3.

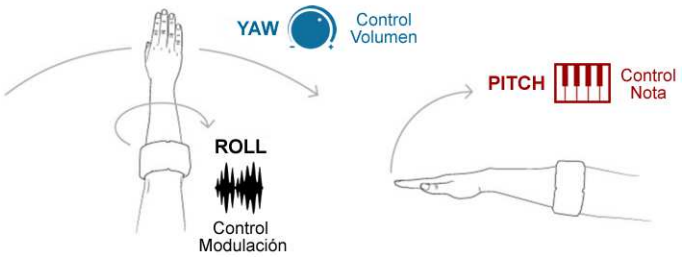


Fig. 3. Movimientos de una mano y parámetros asociados del controlador virtual MIDI Representación YPR [16]

Debido a que los movimientos del brazo humano están restringidos, habrá que limitar el radio de medición de cada uno de los ejes a un rango de 0 a 180 grados. Cada una de las variaciones respecto de los ejes, está asociada a un determinado evento que será traducido en un mensaje MIDI. Para este trabajo se proponen tres parámetros variables. El parámetro de las notas será asociado al movimiento vertical de la mano, definido por la variable pitch, en este caso teniendo en cuenta que el rango de notas de un Grand Piano se encuentra entre los valores 21 y 108, se realiza una transformación

lineal que ajuste dinámicamente los valores de 0 a 180 al rango de notas previamente mencionado. De acuerdo a la especificación MIDI, este parámetro se transmite utilizando un mensaje del tipo Note on/off con el respectivo valor de la nota.
Para el caso del volumen, la especificación MIDI precisa un rango dinámico representado entre 0 y 127, de manera similar se realiza una transformación lineal. Este parámetro será asociado al movimiento horizontal del brazo, o yaw. Este mensaje se transmite a través del parámetro MIDI velocity, y se agrega como parte de los datos del mensaje Note on/off.
La modulación de la nota está asociada a la rotación de la mano, el parámetro MIDI a utilizar será por especificación, un Control Change que utiliza el rango de 0 a 127, en este caso también se hará una transformación lineal.

6 Software del controlador

El software del prototipo controlador MIDI, consta de tres etapas, una inicial de configuración e inicialización de variables del microcontrolador y todos los sensores, otra de calibración del dispositivo, y finalmente un bucle principal para sensar los movimientos y enviar mensajes de forma inalámbrica.
En la etapa de configuración, se realizan los ajustes preliminares para la ejecución del programa, como la inicialización del puerto serial, para un correcto funcionamiento de la aplicación. Se inicializa la librería Wire para capturar los datos del sensor y se realiza una primera calibración del dispositivo. Finalmente se inicializa el módulo bluetooth y se verifica la conexión antes de empezar a transmitir
La calibración se realiza con el dispositivo en reposo durante la inicialización considerando una serie de muestras, parametrizables, que se promedian para luego sustraer esos valores a cada medición en tiempo real.
El bucle principal del programa consta de la secuencia: Lectura de registros de acelerómetro y giróscopo, Cálculos de ángulos en cada eje, Filtrado complementario, Transformación Lineal para adaptación de valores, Transmisión de mensaje MIDI.

Lectura de registros de acelerómetro y giróscopo: Los registros del sensor de movimiento se acceden directamente por medio de la librería Wire [17] de esta manera se evita cualquier factor externo que pueda incidir en el intervalo de tiempo total de la aplicación. Los registros se leen de forma secuencial en cada ciclo del programa.
Cálculos de ángulos: De acuerdo a la información del sensor, y utilizando la configuración de fábrica, al dividir el valor medido por la unidad de sensibilidad podemos obtener la variación de la aceleración en cada eje.
Filtrado complementario: Tanto los datos del acelerómetro como el del giroscopio son propensos a errores sistemáticos. Para resolver este problema se utiliza un método estándar, en el que se combinan ambas mediciones utilizando filtros complementarios para obtener un único valor.
Transformación Lineal: La adaptación de escala en cada eje, tomando ángulos de 0° a 180° a cada uno de los parámetros MIDI, se realiza con una transformación lineal

para comprimir o expandir dichos valores de acuerdo a la nueva escala, utilizando la función map() de la librería Math [18].

```
int note = map(pitch,-90,90,21,108);
int velocity = map(roll,-90,90,127,0);
int modulation = map(yaw,-90,90,0,127);
```

Transmisión de mensajes MIDI: Se utiliza una librería provista por el módulo Bluefruit, Adafruit_BLEMIDI [19], para instanciar el servicio MIDI y enviar mensajes. Estos envío se realizan invocando a uno de sus métodos:

```
// note on
midi.send(0x90, note, velocity);

// note off
midi.send(0x80, note, velocity);

// control change
midi.send(0xB, contoller_number, ccvalue);
```

7 Prototipo Final

Para arribar al prototipo final, previamente se realizó un modelado en tres dimensiones y un prototipo preliminar evaluando aspectos funcionales y ergonómicos El prototipo final se construyó con una placa de circuito impreso de fibra de vidrio tipo FR4, simple faz, con máscara antisoldante de 50,8 mm x 39,4 mm, reduciendo el tamaño total del dispositivo. Finalmente la placa fue atornillada a un modelo impreso en tres dimensiones y montado sobre la mano utilizando dos cintas de velcro, Fig. 4-5.

Fig. 4. Vista aérea de dispositivo

Fig. 5. Dispositivo montado sobre mano

8 Conclusiones

De acuerdo a los resultados del trabajo se puede concluir que se ha cumplido el objetivo principal que fue desarrollar un dispositivo capaz de traducir la mecánica corporal en sonido, utilizando distintos tipos de herramientas estándar en la industria musical actual.
En lo que respecta a los objetivos específicos podemos concluir lo siguiente:
Se ha diseñado y construido un dispositivo de hardware capaz de capturar los movimientos de una mano en tres dimensiones utilizando un sensor de movimiento de bajo costo y confiable con corrección a efectos ruidosos inherentes al sensor utilizado.
Se desarrolló una solución de software para transformar los datos obtenidos en mensajes del tipo MIDI.
Se ha implementado un mecanismo de transmisión inalámbrica utilizando tecnología Bluetooth de baja energía para optimizar la utilización de recursos.
Se ha construido una carcasa utilizando tecnología de impresión 3D para el montaje del dispositivo a la mano.
Se ha testeado el dispositivo en distintas iteraciones agregando valor al proceso de diseño y mejorando el producto final.
A lo largo del mismo se ha experimentado el proceso íntegro de fabricación de un producto desde todas sus etapas generando un resultado tangible y apto para su evaluación como tal, ya sea desde el ámbito académico, como comercial.
El prototipo final se ha construido íntegramente con componentes y materiales económicos y de fácil acceso, concluyendo que dicha construcción se puede realizar a un costo bajo y en un periodo corto de tiempo. Además se puede destacar la experimentación con nuevas tecnologías como el diseño y la fabricación de modelos en tres dimensiones, lo cual contribuye al resultado final con un foco puesto en materia de innovación
Finalmente, se destaca que el dispositivo puede extender su uso en otro tipo de aplicaciones no relacionadas con la música, o cualquier otro propósito en el que pueda inspirarse desde este trabajo. Entre ellas pueden mencionarse aplicaciones para asistencia a personas con movilidad reducida, aplicaciones de entretenimiento, aplicaciones de realidad aumentada, aplicaciones de comando a distancia que resulten útiles en el contexto sanitario actual.

9 Trabajos Futuros

En cuanto a hardware se pueden mencionar mejoras como: incluir controles manuales, para regular parámetros del software; diseñar una placa de montaje superficial y usar batería de litio para reducir el tamaño del dispositivo. Referente al software se puede: desarrollar una funcionalidad de arpegiador, para generar distintos patrones a través de una nota, que podrían ser acordes o secuencias; incorporar el parámetro MIDI Program Change para poder cambiar el tipo de instrumento desde el dispositivo.

10 Agradecimientos

Los autores agradecen a la CIC PBA, donde M. Presso pertenece a la carrera de Profesional de Apoyo a la Investigación.

Referencias

1. The Complete MIDI 1.0 Detailed Specification, The MIDI Manufacturers Association, 1996
2. The 'USI', or Universal Synthesizer Interface, AES, 1981
3. Holonic Systems - https://www.holonic.systems/
4. ARDUINO TEAM, 2018 - https://blog.arduino.cc/2018/06/04/making-music-with-your-muscles/ , Arduino - https://www.arduino.cc/reference/en/ , MIDI - https://ccrma.stanford.edu/~craig/articles/linuxmidi/misc/essenmidi.html
5. A MIDI Controller based on Human Motion Capture, Maurice Velte, 2012 - https://www.researchgate.net/profile/Maurice_Velte/publication/264562371_A_MIDI_Controller_based_on_Human_Motion_Capture_Institute_of_Visual_Computing_Department_of_Computer_Science_Bonn-Rhein-Sieg_University_of_Applied_Sciences/links/53e876030cf2fb7487241a8d.pdf
6. Wireless Midi Controller Globe, Michael Brady, Sarah Palecki, and Allan Belfort, 2018 - https://courses.engr.illinois.edu/ece445/getfile.asp?id=12401
7. Glove MIDI Controller, Anson Dorsey, Eric Gunther, Jonathon Smythe, - https://people.ece.cornell.edu/land/courses/ece4760/FinalProjects/s2010/ecg35_ajd53_jps93/ecg35_ajd53_jps93/index.html
8. The Music Glove, Kaiku https://www.kaikumusicglove.com/
9. MPU6050 - https://invensense.tdk.com/products/motion-tracking/6-axis/mpu-6050/ , https://invensense.tdk.com/wp-content/uploads/2015/02/MPU-6500-Register-Map2.pdf , https://playground.arduino.cc/Main/MPU-6050/
10. Tutorial: How to use the GY-521 module (MPU-6050 breakout board) with the Arduino Uno, Michael Schoeffler, Consultado Enero 2020 - https://www.mschoeffler.de/2017/10/05/tutorial-how-to-use-the-gy-521-module-mpu-6050-breakout-board-with-the-arduino-uno/
11. Introducing the Adafruit Bluefruit LE SPI Friend, Kevin Townsend, 2020 - https://cdn-learn.adafruit.com/downloads/pdf/introducing-the-adafruit-bluefruit-spi-breakout.pdf?timestamp=1594981255
12. ATMEGA328 – https://www.microchip.com/wwwproducts/en/ATmega328P , https://drive.google.com/file/d/1ydLJbmDPw5O1KUB8RgsAxUxedzD1G0qI/view
13. Ble Console - https://sensboston.github.io/BLEConsole/
14. virtualMIDI, Tobias Erichsen - http://www.tobias-erichsen.de/software/virtualmidi.html
15. How to use MIDIberry with DAW - http://newbodyfresher.linclip.com/how-to-use-with-daw
16. GUI without the G: Going Beyond the Screen with the Myo™ Armband - https://developerblog.myo.com/gui-without-g-going-beyond-screen-myotm-armband/
17. Wire Library - https://www.arduino.cc/en/reference/wire
18. map(), arduino - https://www.arduino.cc/reference/en/language/functions/math/map/
19. Adafruit BLE Library Documentation, Dan Halbert, June 2020 - https://readthedocs.org/projects/adafruit-circuitpython-ble/downloads/pdf/latest/

ISBN 978–987–633–574–4

Control Activo de Ruido Impulsivo Basado en la Correntropía del Error con Ancho de Kernel Variable

Patricia N. Baldini[1],

[1] Departamento de Ingeniería Electrónica, Facultad Regional Bahía Blanca, Universidad Tecnológica Nacional, 11 de Abril 461
8000 Bahía Blanca, Argentina
pnbaldi@frbb.utn.edu.ar

Abstract. Active control is a methodology based on the waves destructive interference that has proven to be effective for attenuating noise in the low frequency audible spectral range. However, the case of impulsive type noise sources, as harmful as frequent in industrial environments, represents a challenge to the convergence of the control algorithm that is still a matter of study. Outliers in the measured signals cause overcorrections in adaptive adjustment of filter weights which can produce instability. This paper presents the results of applying a new robust methodology to attenuate impulsive noise in a single-channel system. The proposed algorithm based on the maximum correntropy criterion with recursively adjusted kernel size, does not require prior statistical information on noise. The convergence properties and the effectiveness of the control indices are verified by simulation in different conditions of noise environments. Impulsive noise is represented by the non-gaussian model proposed in the bibliography.

Keywords: Active noise control, Maximum correntropy algorithm, Adaptative kernel size, Impulsive noise.

1 Introduction

El ruido es uno de los contaminantes ambientales más extendido en la actualidad. El concepto de contaminación acústica hace referencia a los niveles excesivos de ruido y vibraciones provocados por la actividad humana, que se constituyen en causa de una gran variedad de efectos nocivos para las personas y su entorno. Tradicionalmente, la gestión ambiental del ruido se realiza mediante las denominadas técnicas pasivas. Estas técnicas consisten en la introducción de barreras físicas para bloquear la propagación o absorber el ruido directo y el reverberante, sin aporte de energía. Sin embargo, a mayores longitudes de onda de la señal a silenciar, el control pasivo se torna ineficiente con incremento significativo en volumen y costo. En todo caso, resultan sistemas poco flexibles que no contemplan cambios del entorno acústico.

La alternativa para el rango acústico audible de bajas frecuencias es el control activo (CAR) que trata de transformar favorablemente el campo sonoro mediante

dispositivos electroacústicos. El uso de fuentes secundarias permite generar de forma controlada nuevas ondas de sonido (*anti-ruido*), que se superponen al campo ruidoso original de modo de producir interferencia destructiva. Se crea una zona de silencio o, al menos, un campo resultante del menor nivel sonoro posible en regiones del espacio tanto más grandes cuanto mayores sean las longitudes de onda del ruido a cancelar.

Si bien el avance tecnológico de los procesadores digitales de señal sumado al desarrollo de algoritmos de procesamiento adaptativo permitieron la implementación de sistemas de CAR eficaces en distintas aplicaciones, el caso de ruido de impacto o impulsivo representa aún un desafío que sigue siendo motivo de estudio. Este tipo de ruido se caracteriza por un número significativo de perturbaciones de gran amplitud que ocurren al azar con una baja probabilidad y no puede describirse mediante un modelo gaussiano. La presencia de valores atípicos, ya sea en el ruido a cancelar o en la señal de error, compromete la convergencia del algoritmo adaptativo pudiendo causar inestabilidad [1]. Los métodos propuestos inicialmente para atenuarlo pueden clasificarse en tres categorías. La primera incluye a los algoritmos adaptativos que utilizan como información el ruido primario y la señal de error residual recortados a un umbral conveniente para suavizar el efecto de los valores atípicos sobre la actualización del controlador [2]. La segunda categoría incluye a los algoritmos basados en la minimización del momento fraccional de orden p ($p<2$) del error residual, teniendo en cuanta que no existe el de segundo orden para la descripción estadística del ruido impulsivo, [3]. Una tercera categoría, engloba algoritmos que emplean transformaciones no lineales del error, con crecimiento acotado [4], [5].

Recientemente han adquirido relevancia algoritmos de control que pueden encuadrarse en una nueva categoría asociada al aprendizaje basado en la Teoría de la Información [6]-[14], que no requieren información a priori de las características estadísticas del ruido. En particular la maximización de la correntropía es uno de los criterios de optimización más populares debido a su simplicidad y robustez, que ha sido aplicado con éxito en casos de ruido no Gaussiano e impulsivo [10]-[13]. En estos métodos la selección del ancho del kernel afecta significativamente la eficacia del filtrado. Si bien algunas alternativas han sido propuestas contemplando un ancho de kernel adaptativo [11], [12], [15],[16], éstas no han sido verificadas para los casos de modelos de ambientes ruidosos de fase no mínima con alta impulsividad.

En este trabajo se propone un algoritmo de filtrado adaptativo inspirado en el de máxima correntropía (MCC) con actualización del ancho del kernel, que presenta buenas características de convergencia y error de estado estacionario en las simulaciones para sistemas tanto de fase mínima como no mínima. Los resultados se analizan en base a los índices de comportamiento usuales y el ruido impulsivo se modela mediante una distribución alfa-estable simétrica (SαS) [2]-[4],[9],[12],[16].

1.1 Configuración Básica del Sistema de CAR

El CAR se basa en el principio de interferencia destructiva entre ondas acústicas. Esencialmente, el ruido se cancela en una determinada región del espacio al superponerle otro en contrafase generando en forma controlado. Un sistema de CAR de un solo canal de tipo feedforward (Fig. 1), comprende: un sensor de referencia para

captar el ruido fuente o primario, $x(n)$; un parlante que actúa como transductor electroacústico para propagar la señal de cancelación, $y(n)$, que es generada por el filtro adaptativo con función transferencia $H(z)$, y un micrófono de error para detectar el nivel de ruido residual, $e(n)$, en la zona de silencio predeterminada [1].

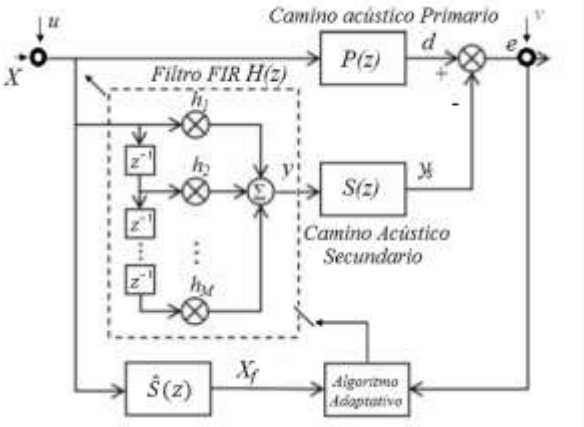


Fig. 1. Esquema del sistema de CAR monocanal de tipo *feedforward.*

El sistema adaptativo depende de cuatro elementos básicos: la estructura del filtro; los parámetros a ajustar que son los coeficientes que definen la función transferencia que modela el filtro; las señales que se procesan, y el algoritmo adaptativo que describe la actualización de los parámetros en cada instante de tiempo, k. Este algoritmo optimiza el control en base a una función de objetivo. El error cuadrático medio ha sido la opción más utilizada bajo la suposición implícita de que el error resultante es una variable aleatoria de tipo gaussiano, justificada por el teorema del límite central.

Para una estructura de filtro con respuesta al impulso finita (FIR) de longitud L, con vector de salida y, $\boldsymbol{y}(k) = [y(k) \quad y(k-1) \cdots y(k-L+1)]^T$, donde $[.]^T$ denota transposición, el error residual queda definido por

$$\begin{aligned} e(k) &= d(k) - \boldsymbol{S}^T \boldsymbol{y}(k) + v(k) \\ y(k) &= \boldsymbol{H}^T(k)\boldsymbol{x}(k) \quad , \quad d(k) = \boldsymbol{P}(k)^T \boldsymbol{x}(k) \end{aligned} \tag{1}$$

$\hat{S}$, S y P son los vectores de las respuestas al impulso estimada y real del camino acústico secundario, y real del camino acústico primario, modelados por las funciones transferencias $\hat{S}(z)$, $S(z)$ y $P(z)$, respectivamente. Los procesos aleatorios $\boldsymbol{u}(k)$ y $v(k)$ representan, respectivamente, ruido de medida de la señal de entrada y del error residual y $x_f(k) = \hat{\boldsymbol{S}}^T(k)(\boldsymbol{x}(k) + \boldsymbol{u}(k))$ es la señal de entrada filtrada por la estimación (fuera de línea) de la respuesta impulsiva del camino secundario.

De todos modos, el ruido de tipo impulsivo representa un desafío a los métodos adaptativos convencionales debido a que la gran amplitud ocasional en las señales

medidas produce una actualización repentina significativa de coeficientes del filtro que puede comprometer la convergencia e inestabilizar al sistema. Si bien se han propuesto distintas estrategias para superar estas limitaciones, el obstáculo común para implementarlas en la práctica es la complejidad computacional asociada y la respuesta insatisfactoria para sistemas de fase no mínima. [2]-[8].

1.2 Algoritmos basados en la Maximización de la Correntropía

La correntropía, que puede pensarse como una correlación generalizada, se emplea como medida no lineal de la similitud entre dos variables aleatorias en una vecindad del espacio conjunto dependiente del ancho del kernel. La robustez frente a *outliers* que se logra reduciendo este parámetro, la convierte en una función objetivo adecuada para sistemas adaptativos frente a ruido impulsivo con distribuciones *heavy-tailed.*

La correntropía se define mediante la expresión (2):

$$V(d,y) = E[\kappa(d-y)] = \int \kappa(d-y)dF_{dy}(d,y) \qquad (2)$$

donde κ denota un kernel de Mercer invariante al desplazamiento y F_{dy} es la función de distribución de probabilidad conjunta de las variables aleatorias *d, y* ([14]). El kernel comúnmente adoptado es el Gaussiano con ancho σ (>0),

$$\kappa(d-y) = G_\sigma(d-y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(d-y^2}{2\sigma^2}\right). \qquad (3)$$

Como la función distribución conjunta es desconocida, el operador esperanza se reemplaza por el estimador muestral que tiene en cuenta la ventana temporal de *N* pares de datos disponibles (d_k,y_k) (k=1,2..,N) [9] de modo que

$$V(d,y) \approx \frac{1}{N}\sum_{k=1}^{N} G_\sigma(d_k - y_k) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(d_k-y_k)^2}{2\sigma^2}\right) \qquad (4)$$

Desde el punto de vista de CAR, el vector de coeficientes del filtro FIR adaptativo con longitud L, $\boldsymbol{H}(k)$, se actualiza maximizando la correntropía entre las señales de salida de los caminos acústicos primario y secundario, $d(k)$, $y_s(k) = \boldsymbol{S}^T\boldsymbol{y}(k)$, mediante el método de máximo descenso del gradiente.

Si $e_k = e(k) = d(k) - \boldsymbol{S}^T\boldsymbol{y}(k) + v(k)$ es el error de predicción en la iteración *k-ésima* , $\boldsymbol{X}(k)$ el vector de entradas, $x_f(k) = \hat{\boldsymbol{S}}^T(\boldsymbol{X}(k) + \boldsymbol{v}(k))$ y $\boldsymbol{X_f}$=[x_f(k)… x_f(k-L_s)]^T el vector de entradas filtradas por $\hat{\boldsymbol{S}}$, la regla de actualización puede escribirse

$$\boldsymbol{H}(k+1) = \boldsymbol{H}(k) + \mu\sum_{n=1}^{k} exp\left(-\frac{e_n^2}{2\sigma^2}\right)e_n\boldsymbol{X}_f(n) \rightarrow \qquad (5)$$

$$\boldsymbol{H}(k+1) = \boldsymbol{H}(k) + \mu\, exp\left(-\frac{e_k^2}{2\sigma^2}\right) e_k \boldsymbol{X}_f(k)$$

Es evidente que el efecto del error sobre los coeficientes del filtro se hace despreciable a medida que su magnitud crece lo que garantiza la robustez frente a ruido impulsivo. De todos modos, al considerar solo las muestras actuales del error y la entrada, el desajuste estacionario resulta significativo. Por otro lado, las grandes fluctuaciones de la entrada pueden afectar negativamente la velocidad de convergencia [13]. Una alternativa posible es aproximar el vector gradiente considerando el promedio de las últimas *N* muestras, a expensas de aumentar tanto el costo de almacenamiento como de cómputo.

Otro inconveniente del método es la selección adecuada del ancho fijo del kernel, σ^2.Un valor grande proporciona buena velocidad de convergencia con un desajuste estacionario importante. De otro modo, si el valor es pequeño, se corrige el valor estacionario pero se pierde rapidez de convergencia. Para superar esta dificultad se propusieron algunas estrategias de variación recursiva del ancho del kernel [11],[12],[15]-[17] aunque no siempre resultan robustas para ruido fuertemente impulsivo o ambientes acústicos caracterizados por modelos de fase no mínima, presentan importante desajuste en estado estacionario o incrementan el costo de almacenamiento. Por ejemplo, la combinación convexa de dos filtros MCC con diferentes anchos de kernel propuesto en [18] duplica la complejidad de cómputo.

2 Algoritmo para CAR Impulsivo Propuesto

En este trabajo se propone usar un filtro adaptativo basado en MCC para ajustar los coeficientes del filtro que pretende mejorar la estabilidad y el comportamiento estacionario en ambientes de ruido desfavorables, sin incrementar costo de almacenamiento. El algoritmo de máxima correntropía recursiva con filtrado-x y actualización adaptativa del ancho de un kernel de tipo Gaussiano, se describe mediante el conjunto de ecuaciones (6) a (8),

$$\boldsymbol{P}(k+1) = \lambda \boldsymbol{P}(k) + (1-\lambda) exp\left(-\frac{e_{k+1}^2}{2\sigma_{k+1}^2}\right) e_{k+1} \boldsymbol{X_f}(k+1) \qquad \textbf{(6)}$$

donde $\boldsymbol{P}(.)$ es el promedio móvil ponderado exponencial del vector gradiente de MCC con un factor de suavizado λ ($0 << \lambda < 1$) que se utiliza para aproximar al vector gradiente de la esperanza de la función objetivo. La operación de promediado tiene el efecto de un filtrado pasabajos que tiende a reducir las oscilaciones en rangos cortos de tiempo, estabilizando el comportamiento de $\boldsymbol{P}$ [13]**.**

En lo que se refiere al ancho del kernel, se busca que sea grande al inicio de la adaptación para una buena velocidad de convergencia y se reduzca cuando el vector de coeficientes del filtro se acerca su valor óptimo para disminuir el desajuste estacionario. Se opta por promedio móvil ponderado de la magnitud del error,

$$\sigma_{k+1} = \eta\sigma_k \ + (1-\eta)|e_k\ | \qquad \textbf{(7)}$$

donde $0<\eta<1$ es un factor de olvido. Finalmente, la expresión de actualización del vector de coeficientes queda expresado por (8):

$$\boldsymbol{H}(k+1) = \boldsymbol{H}(k) + \mu \boldsymbol{P}(k) \tag{8}$$

donde $\mu>0$ es el paso de actualización. El método propuesto aumenta la robustez a costa de un incremento leve incremento de multiplicaciones y adiciones en relación al MCC clásico. De todos modos, la complejidad es comparable o menor a la de otras opciones de ancho de kernel variable. En la Tabla 1 se sintetiza el procedimiento, donde se incluye la normalización del vector $\boldsymbol{X}_f$ para acotar la variación por ruido impulsivo en la entrada.

Table 1. Algoritmo propuesto.

Input: η, λ ,μ , $\widehat{S}$

Initialize: σ^2_0 , $\boldsymbol{P}(0)$, $\boldsymbol{H}(0)$

while { $x(k)$,$e(k)=e_k$ } available

$\boldsymbol{x_f}(k) = \widehat{S}^{\mathrm{T}}\boldsymbol{X}(k)$

$\boldsymbol{H}(k+1) = \boldsymbol{H}(k) + \mu \boldsymbol{P}(k)$

$\sigma_{k+1} = \eta\sigma_k + (1-\eta)|e_k|$

$\boldsymbol{P}(k+1) = \lambda \boldsymbol{P}(k) + (1-\lambda) exp\left(-\frac{e_k^2}{2\sigma_k^2}\right) e_k \frac{\boldsymbol{X_{fk}}}{(\boldsymbol{X_{fk}}^T \boldsymbol{X_{fk}} + \boldsymbol{\varepsilon})}$

end while

Output: $\boldsymbol{H}^*=\boldsymbol{H}(k+1)$

2.1 Modelo Estadístico de Ruido Impulsivo

El ruido impulsivo se caracteriza por la aparición de muestras de gran valor con baja probabilidad de ocurrencia. En la literatura reciente sobre CAR impulsivo, es una práctica habitual modelar este tipo de ruido no gaussiano usando una distribución alfa-estable simétrica (SαS), [2], [3], [9], [12], [16].

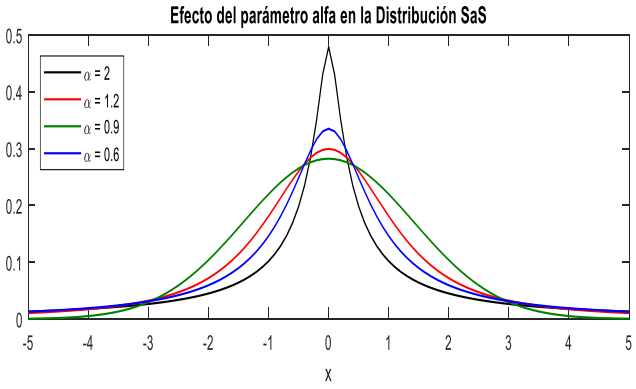


Fig. 2. Función distribución de probabilidad SαS en función de α.

Un proceso aleatorio es SαS, si su función característica se puede expresar como $\varphi(x) = e^{-\gamma|x|^{\alpha}}$, donde γ (>0) es el parámetro de dispersión y α (1< α ≤ 2) es el exponente característico. En particular, se considera que la fuente de ruido se modela mediante una distribución SαS estándar (γ=1), donde el grado de impulsividad será tanto mayor cuanto menor sea el exponente característico α (Fig. 2). El caso particular de α=2 corresponde a la distribución Gaussiana que es la única que admite momento de segundo orden finito. En el resto de los casos de (α < 2) solo existen momentos de orden fraccional $p < \alpha$. Como consecuencia de este hecho, el CAR clásico basado en el algoritmo de mínimos cuadrados (FxLMS) resulta inapropiado.

2.2 Experimentos de Simulación

Los resultados de aplicar el algoritmo propuesto fueron analizados, en una primera instancia, mediante simulación computacional utilizando la plataforma Matlab. Se consideraron distintas experiencias variando el grado de impulsividad del ruido primario mediante el parámetro α incluyendo casos de su variación en el tiempo.

El índice de desempeño usado para medir la velocidad de convergencia fue la reducción de ruido promedio, RRP, siguiendo la bibliografía, [2], [3], [9], [12], definida por las ecuaciones (9) y (10) con $\lambda = 0.99$, promediando una serie de 20 conjuntos independientes de datos de entrenamiento generados aleatoriamente.

$$RRP(n) = 20 log_{10}\left(\frac{A_e(n)}{A_d(n)}\right) \quad \textbf{(9)}$$

donde

$$\begin{aligned} A_e(n) &= \lambda A_e(n-1) + (1-\lambda)|e(n)| \\ A_d(n) &= \lambda A_d(n-1) + (1-\lambda)|d(n)| \end{aligned} \quad \textbf{(10)}$$

La estimación del camino secundario fuera de línea, previa a la aplicación del control y la longitud del filtro FIR de control se fija en L=16.

2.1.2 Primer Experimento

En esta experiencia se considera un ruido altamente impulsivo, con $\alpha = 1.2$ y funciones transferencia (FT) de los caminos acústicos primario y secundario de fase mínima (FM) (Fig. 3 y 5 *a*) y de fase no mínima (FNM) (Fig. 4y 5 *b*).

En ambos casos, las figuras muestran el ruido primario, la señal acústica de control o cancelación el ruido (anti ruido), el error residual y la evolución de la RRP [dB] con las iteraciones.

La reducción de ruido en el caso de FT de FM es aproximadamente de 60dB mientras que en el caso de FNM, si bien se reduce a 25dB, es un valor altamente satisfactorio teniendo en cuenta que la dificultad implicada.

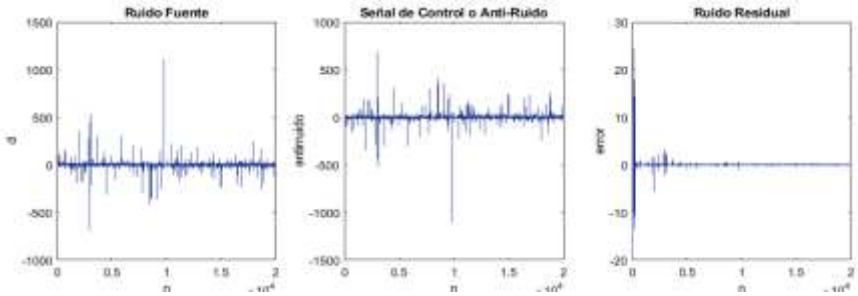


Fig. 3. Señales de salida de ruido camino primario, anti ruido y error residual caso FM

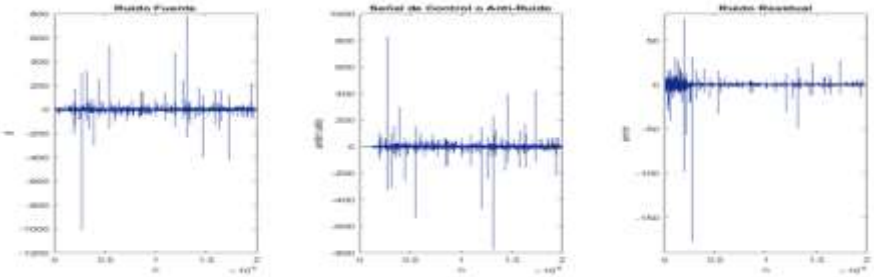


Fig. 4. Señales de salida de ruido camino primario, anti ruido y error residual caso FNM

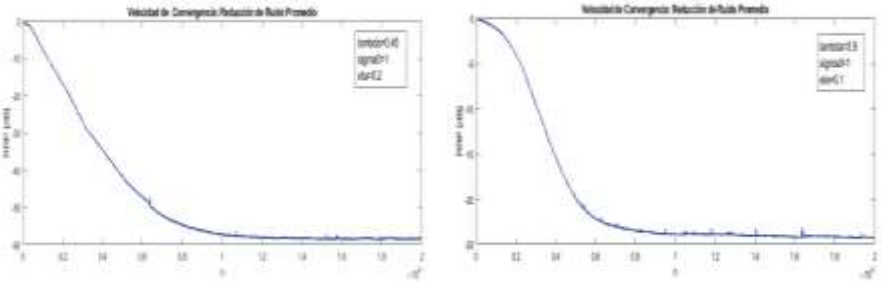


Fig. 5. Evolución del RRP : a) caso FM y b) caso FNM

2.1.3 Segundo Experimento VE:

En esta experiencia se incluye ruido impulsivo con características probabilísticas variables en el tiempo [3]. Se considera un cambio abrupto del grado de impulsividad luego de un cierto tiempo, según se indica en la expresión *a)* de (11), para modelar una modificación de la probabilidad de ocurrencia de *outliers* (VE con FT FM y FNM: Fig. 6 *a*). También se analiza una variación suave de tipo senoidal según la expresión *b)* de (11) (VS con FT FM y FNM: Fig. 6 *b*)).

$$a)\begin{cases}\alpha = 1.8 & ,0 \le n \le 6000 \\ \alpha = 1.4 & ,6000 < n \le 13000 \\ \alpha = 1.6 & ,13000 < n \le 20000\end{cases} \quad , \quad b) \quad \alpha(n) = 1.6 + 0.3\,sen\left(\frac{2\pi}{5\times 10^4} n\right) \qquad \textbf{(11)}$$

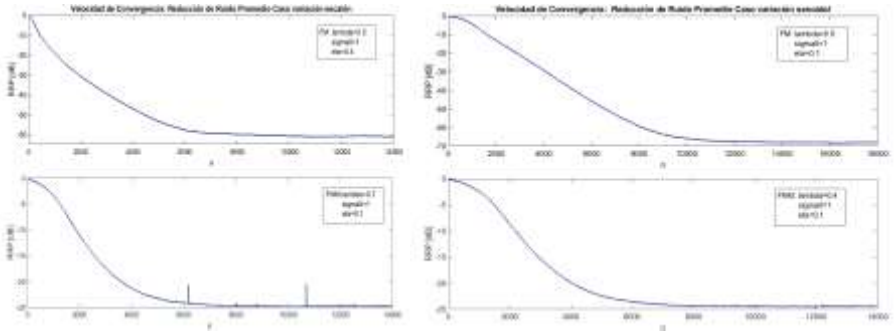


Fig. 6. Evolución del RRP : a) caso VE y b) caso VS.

2.1.3 Tercer Experimento

En este caso se compara mediante la RRP el algoritmo propuesto con el AMCC propuesto en [16] (parámetros μ=0.001, σ_0^2=4 y μ=0.0005, σ_0^2=4) para el caso de ruido altamente impulsivo (α=1.2) y FT de FM (Fig.7 *a*) y FNM (Fig. 7 *b*). En AMCC el ancho del kernel varía según $\sigma^2=\sigma_0^2+e_k^2$. La superioridad de la nueva metodología es notable en ambos casos, especialmente en cuanto al error residual.

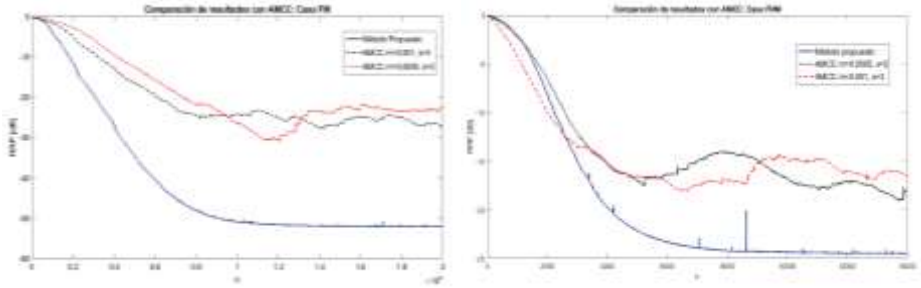


Fig. 7. Comparación del RRP : a) caso FM y b) caso FNM

3 Conclusiones y Trabajo Futuro

En este trabajo se analiza la performance de una estrategia de control adaptativo para la atenuación de ruido impulsivo unidimensional, inspirada en la optimización de una función objetivo definida como la correntropía del error residual. Se recurre a una aproximación del gradiente de la correntropía, empleando un promedio móvil ponderado no lineal, que mejora la robustez frente a ruido altamente impulsivo en particular en ambientes acústicos modelados con funciones transferencia de fase no mínima. La metodología introducida presenta buena velocidad de convergencia con un bajo desajuste en estado estacionario gracias a la modificación recursiva del ancho del kernel en base a la magnitud del error. Los resultados de las simulaciones muestran una atenuación del ruido de salida mucho mayor que otros métodos existentes (AMCC), sin añadir complejidad computacional considerable. En una siguiente etapa se buscará corroborar los resultados en un prototipo de laboratorio.

Agradecimiento. Se expresa el agradecimiento a la Secretaría de Ciencia y Tecnología de las Universidad Tecnológica Nacional por la financiación del proyecto de investigación en el marco del cual se desarrolló este trabajo.

References

1. Elliott, S.J.: Signal Processing for Active Control. Academic Press, San Diego, USA, 2001.
2. Akhtar, M.T., Mitsuhashi,W.: Impruving performance of FxLMS algorithm for active noise control of impulsive noise. Journal of Sound and Vibration, vol. 327, pp. 647-656, (2009).
3. Bergamasco, M., Della Rossa, F., Piroddi,l.: Active noise control with on-line estimation of non-Gaussian noise characteristics. Jou. Sound and Vibration, vol. 331, pp. 27-40 (2012).
4. Akhtar, M.T.: An adaptive algorithm, based on modified tanh non-linearity and fractional processing, for impulsive active noise control systems. Jou. Low Frequency Noise, Vibration and Active Control, vol. 37, issue 3, pp. 1–14, 2017.
5. Liang, ,T.,Li, Y., Zakharov , Y.V., Xue, W., Qi, J.: Constrained least lncosh adaptive filtering algorithm. Signal Processing. 183 (2021) 108044
6. Song, P., Zhao, H., Zhu, Y.: Filtered-s normalized maximum mixture correntropy criterion algorithm for nonlinear active noise control. In: Proceedings Volume 11719, Twelfth International Conference on Signal Processing Systems; 1171911 (2021)
7. Qian, G., Ning, X., Wang, S.: Recursive Constrained Maximum Correntropy Criterion Algorithm for Adaptive Filtering. IEEE Tran. Circuits and Systems-II: Express Briefs, vol. 67, no. 10, pp. 2229-2233 (2020)
8. Radhika, S., Chandrasekar, A.: Convergence analysis of Maximum Correntropy Criteria based adaptive filtering algorithm based on white input. In: 11th Int. Conference on Advanced Computing (ICoAC), 2019, pp. 158-16
9. Kurian, N.C., Patel, K., George, N.V.: Robust active noise control: An information theoretic learning approach. Applied Acoustics. 117, pp. 180--184 (2017)
10. Zhu, Y., Zhao, H., Zeng, X., Chen, B.: Robust Generalized Maximum Correntropy Criterion Algorithms for Active Noise Control. IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1282-1292 (2020).
11. Wang, W., Zhao, J., Qu, H., Chen, B.: A correntropy inspired variable step-size sign algorithm against impulsive noises. Signal Processing, 141, pp. 168-175 (2017).
12. Lu, L., Zhao, H.: Active impulsive noise control using maximum correntropy with adaptive kernel size. Mechanical Systems and Signal Processing, vol. 87, Part A, pp.180-191(2017)
13. Qu, H., Shi, Y., Zhao, J.: A Smoothed Algorithm with Convergence Analysis under Generalized Maximum Correntropy Criteria in Impulsive Interference. Entropy 21, (2019)
14. Principe, J.C.: Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives. Springer, New York (2010)
15. Huang, F., Zhang, J., Zhang, S.: Adaptive Filtering Under a Variable Kernel Width Maximum Correntropy Criterion. IEEE Tran. Circuits and Systems—II: Express Briefs, vol. 64, no.10, pp. 1247-1251 (2017)
16. Wang,W., Zhao, J., Qu, H.,Chen, B., Principe, J.C.: An adaptive kernel width update method of correntropy for channel estimation. In: 2015 IEEE International Conference on Digital Signal Processing (DSP), 2015, pp. 916-920. IEEE Press (2015)
17. Shi, Y., Zhao, H., Zakharov, Y.: An Improved Variable Kernel Width for Maximum Correntropy Criterion Algorithm. IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 67, no. 7, pp. 1339-1343 (2020)
18. Fontes, A.I.R., Linhares, L.L.S., Guimarães, J.P.F., Silveira, L.F.: An adaptive kernel width convex combination method for maximum correntropy criterion. Journal of the Brazilian Computer Society, 27:7 (2021).

ISBN 978-987-633-574-4

CACIC 2021

WORKSHOP INNOVACION EN EDUCACION EN INFORMATICA

COORDINADORES

Cecilia Sanz (UNLP)
Beatriz Depetris (UNDTF)
Marcelo De Vincenzi (UAI)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

Analíticas de aprendizaje en el contexto de un curso de Ingeniería de la UNLP

Di Domenicantonio Rossana[1], González Alejandro[2] y Hasperué Waldo[2,3],

[1] IMApEC: Facultad de Ingeniería de la Universidad Nacional de La Plata,
[2] Instituto de Investigación en Informática LIDI. Facultad de Informática. Universidad Nacional de La Plata,
[3] Investigador asociado - Comisión de Investigaciones Científicas (CIC-PBA)
1900 La Plata, Bs. As., Argentina
rossanadido@ing.unlp.edu.ar, agonzalez@lidi.info.unlp.edu.ar, whasperue@lidi.info.unlp.edu.ar

Abstract. Este trabajo relata la utilización de analíticas de aprendizaje en un curso de matemática para ingresantes a carreras de Ingeniería de la UNLP. Se realiza una breve historia del arte del tema, se enumeran los objetivos del trabajo y la aplicación de esta herramienta a tres cohortes de alumnos ingresantes a las diferentes carreras de Ingeniería que realizaron la materia "Matemática para Ingeniería" a distancia. Este trabajo forma parte de una tesis de maestría sobre tecnología informática aplicada en educación de la Facultad de Informática de la UNLP. Se esbozan los primeros resultados del trabajo de campo y se extraen conclusiones.

Keywords: 1-Analiticas de aprendizaje 2- Matemática 3- Ingresantes 4- Educación a distancia

1 Introducción

La minería de datos aplicada en contextos educativos provee diversos indicadores que pueden ser utilizados tanto en una materia de matemática ofrecida con modalidad a distancia o con acompañamiento virtual como en otras materias que utilizan plataformas educativas que constituyen el ambiente virtual de enseñanza y aprendizaje del alumno con el fin de detectar patrones que motiven a docentes y alumnos a implementar estrategias de retención y mejoras en el proceso de enseñanza y aprendizaje.

En coincidencia con M. Maggio, los registros de las prácticas de enseñanza y los procesos de aprendizaje en ambientes con alta disposición tecnológica reflejan una mayor complejidad y profundidad que en otros modelos con menor acceso a la tecnología [1]. En este sentido es un desafío plantear una propuesta donde observar, revisar, entender y construir un modelo y su validación que enriquezca el modelo actual que tiene esta modalidad de la materia.

"Matemática para ingeniería" es la primera materia del plan de estudio de las trece carreras de ingeniería de la Facultad de ingeniería de la UNLP. Es la materia que equivale a lo que antes del 2016 correspondía a la nivelación en matemática de los alumnos ingresantes a carreras de Ingeniería.

LMS "Learning Managment System" son las plataformas de formación virtual más utilizadas en instituciones de Educación Superior, para la gestión de acciones de formación con tecnología, tanto para educación a distancia completa o en combinación de la enseñanza presencial o b-learning. Se describirán

los factores o variables que pueden determinar patrones y características para un aprendizaje eficaz en los LMS.

Los objetivos principales del presente trabajo son:

- Describir las características y principales autores sobre LA “Learning Analytics” y su utilización en este trabajo.
- Definir factores, variables, patrones y características que determinen un aprendizaje eficaz en los LMS.
- Asociar esos factores al curso de matemática de la FI donde se realiza el estudio.
- Extraer primeros resultados cuantitativos y posibles reflexiones sobre esas mediciones.

2 Matemática para ingeniería

Esta materia es la primera con la que los alumnos ingresantes a las trece carreras de ingeniería de la UNLP deben enfrentarse para iniciar sus estudios universitarios. Una de esas carreras (Ingeniería en Computación) además requiere de una nivelación en informática IAI para los alumnos ingresantes que es dictada por la Facultad de Informática al mismo tiempo que hacen la materia de matemática. En matemática para ingeniería se nivelan, repasan y profundizan los conceptos y herramientas matemáticas que los alumnos deben conocer para realizar el resto de las materias de matemática que la suceden y tienen correlatividad en el plan de estudios. Esta materia se dictaba hasta el inicio de la pandemia por Covid-2019 de manera presencial y en forma a distancia para los alumnos ingresantes que residan a más de 60 kilómetros de la ciudad de La Plata donde tiene sede la Facultad. Existe desde hace unos años una modalidad especial para estos alumnos que se dictaba de manera virtual y sobre la que se tomaron datos y mediciones para hacer el análisis.

De manera presencial la materia tiene una modalidad de aula taller donde los alumnos aprenden haciendo y compartiendo entre alumnos y docentes el contenido, procedimientos y procesos que redunden en un mejor aprendizaje. La metodología de aula taller fomenta la integración de la teoría con la práctica; según Ander-Egg la teoría surge como una necesidad para la práctica, tanto para interpretar la problemática a resolver como para orientar las estrategias que se llevarán a cabo para ello [2]. Esta forma de aprender, a partir de la resolución de problemas, de una manera cooperativa y grupal, no es compatible con la enseñanza tradicional, sobre todo en cuanto a matemática se refiere, con un profesor transmitiendo su conocimiento de manera expositiva y un grupo pasivo de alumnos. En esta metodología de aula taller de acuerdo con Pasel y Asborno el profesor es el coordinador de las actividades en el aula y el evaluador de los procesos de aprendizaje que va realizando cada estudiante [3]. En concordancia con García el alumno debe construir su propio conocimiento y esto dependerá de su voluntad para aprender, de los aportes propios y de sus compañeros y de la interrelación entre los integrantes del grupo entre sí y con sus docentes [4].

En el curso de Matemática para ingeniería que se dictó para los alumnos que residían a más de 60 kilómetros de distancia se intentó replicar esta modalidad y por ello hay diferentes estrategias que se utilizaban con el fin de brindar a todos los ingresantes la misma preparación ya que además todos realizaban las mismas evaluaciones para la promoción de la materia. Para el seguimiento de la cursada con modalidad a distancia se diseñó el curso en una de las LMS más utilizadas que es la plataforma Moodle, y además es la adoptada por la Institución dentro del cual se realiza este estudio. Esta plataforma permite obtener buenos resultados académicos como afirman Cabanillas et al. y los estudiantes presentan buenas percepciones hacia ella [5]. Cabe destacar que el estudio se realizó con ingresantes de cohortes anteriores a COVID-19 y en general no todos los estudiantes ni docentes estaban familiarizados con la educación a distancia. Se analizaron datos referidos a cursos de ingresantes de las cohortes 2017, 2018 y 2019 que realizaban la

modalidad de curso a distancia y eran alumnos que estaban realizando el último semestre del colegio secundario en su lugar de residencia (Tabla 1).

Según Scorzo, Favieri y Williner el trabajo en una cátedra numerosa y la utilización de herramientas web y software matemático, es un desafío que requiere creatividad de los docentes, trabajo de indagación de las herramientas, tiempo de aprendizaje de estas y una constante actualización [6]. En este sentido en la cátedra de Matemática para ingeniería se venía trabajando con los alumnos ingresantes que no residían en la ciudad con una modalidad a distancia, aunque sin hacer uso de encuentros sincrónicos que luego de la pandemia por el Covid-19 se hicieron mucho más frecuentes. Según Bartolomé habrá un cambio de paradigma post pandemia en la didáctica y la alfabetización digital tanto de alumnos como de los docentes y las Tics tomarán un rol más activo y protagónico [7]. Es importante destacar, como afirman Di Domenicantonio y Langoni, que en materias masivas y de ingresantes a carreras científico-tecnológicas es deseable que se adquieran nuevas estrategias y formación profesional muy importante en los docentes para que no se reproduzca la presencialidad en la virtualidad [8].

Tabla 1. Cantidad de alumnos de la muestra estudiada de las tres cohortes de ingresantes

Alumnos	Inscriptos	Rindieron	Promocionaron	Relación de promocionados
Cohorte 2017	109	82	32	39%
Cohorte 2018	147	112	49	44%
Cohorte 2019	193	123	49	40%

3 Analíticas de aprendizaje

Daremos cita a diferentes autores que han realizado estudios sobre “Learning Analytics” (LA) como la herramienta que tiene por objetivo utilizar la enorme cantidad de datos que generalmente se dispone cuando se utilizan entornos virtuales en los procesos de formación y no siempre son aprovechados correctamente para realizar predicciones que permitan tomar decisiones oportunas. De esta manera trataremos de abordar a una definición de LA.

Según M. Zapata-Ros “ahora hay una nueva perspectiva: La analítica masiva de datos personalizados”. Los algoritmos adecuadamente orientados por las teorías del aprendizaje personalizado, por técnicas pedagógicas y de diseño instruccional pueden, junto con los avances en minería de datos, obtener información para ajustar mejor la intervención educativa, para mejorar el rendimiento de los alumnos, y el del programa educativo. Según el mismo autor, un desafío de naturaleza prioritaria consiste en utilizar la analítica para detectar indicadores de abandono precoz en estudios en línea [9].

En concordancia con Baker, Jaramillo & Paz, la analítica de aprendizaje se puede definir como el proceso de determinar, evaluar e interpretar el significado de grandes volúmenes de datos educacionales; utilizando para ello algoritmos matemáticos [10] [11].

La Society for Learning Analytics Research define la LA como un campo de estudio y lo describe como “la medición, recolección, análisis y presentación de datos sobre los alumnos y sus contextos, con el propósito de comprender y optimizar el aprendizaje y los entornos en los que se produce” [12].

Márquez Vera ha estudiado en su tesis doctoral datos reales de una escuela de nivel medio superior de la ciudad de Zacatecas, México, con la finalidad de predecir el resultado de los estudiantes al final del curso. Propuso la utilización de un algoritmo genético para obtener un modelo de clasificación que proporcione

reglas de inducción fácilmente comprensibles. Se ha demostrado que usar técnicas como la selección de mejores atributos, el rebalanceo de datos y clasificación pueden ser utilizadas exitosamente para mejorar la precisión de la clasificación [13].

Huang&Fang aplicaron cuatro modelos matemáticos para predecir el rendimiento académico de estudiantes de un curso de ingeniería utilizando para ello las calificaciones finales de los mismos. Los resultados finales mostraron que los puntajes de los exámenes finales de los estudiantes eran predecibles con un 88% de precisión en base a ocho variables recopiladas de un sistema de gestión de aprendizaje (LMS) [14].

Hu&Shih desarrollaron un sistema de alerta temprana basado en árboles de decisión para predecir si los estudiantes aprobaran o no. El modelo fue construido utilizando datos de 300 estudiantes y 13 variables recogidos mediante analítica en línea. Los resultados revelaron un 95% de precisión [15].

Salgado Reyes et al sostienen que la minería de datos educativos desarrolla modelos y métodos para explorar los datos recopilados de los entornos de aprendizajes educativos mediante analíticas de aprendizajes con el fin de detectar patrones que permitan predecir variables de interés en instituciones educativas universitarias [16].

Ye & Biswas, citado por Salgado Reyes et al. afirman que pesar de que el rendimiento académico es una variable multifactorial, muchos de los estudios alrededor de la misma incluyen solo factores personales y socioeconómicos; sin embargo, el surgimiento y la aplicación de las nuevas tecnologías de enseñanza sobre todo el uso de las plataformas virtuales, permiten a las universidades recolectar una gran cantidad de información en tiempo real. Estos cuantiosos datos electrónicos generados, proporcionan un abordaje multivariante en el estudio del rendimiento académico [16].

Adoptamos como definición para este trabajo que Analítica de Aprendizaje es el estudio y procesamiento de los datos obtenidos de un ambiente educativo enriquecido con tecnología con el fin de clasificar, determinar, evaluar e interpretar el significado de grandes volúmenes de datos educacionales, utilizando para ello algoritmos matemáticos, la observación detallada de patrones propios del contexto de estudio y la obtención de conclusiones para la toma de decisiones.

Cabero Almenara et al afirman que se deben definir variables favorecedoras de crear acciones para la formación virtual de calidad. Y en ese sentido es muy importante la definición de las variables a estudiar para luego aplicar una técnica estadística que logre mostrar la tendencia en el estudio realizado, teniendo en cuenta el contexto de estudio, la confiabilidad de los datos y la significación de estos [17].

4 Primeros análisis de datos

Es importante para la calidad de la enseñanza virtual, no solo tener en cuenta las variables de carácter tecnológicos como la usabilidad, funcionalidad o la facilidad de manejo de las diferentes herramientas tecnológicas, sino que es importante analizar como lo hacen Ruiz y Dávila las interacciones entre los diferentes actores que intervienen en el proceso de enseñanza y aprendizaje, para que los resultados del proceso sean satisfactorios, entre los estudiantes, docentes, contenidos, evaluaciones, tecnología y la institución donde se realiza el proceso [18].

Entre los factores que inciden en la calidad de la docencia universitaria virtual, diversos estudios señalan el rol destacado del docente por varios motivos: su función en la orientación y supervisión del proceso de aprendizaje, por su acción tutorial enfocada a ayudar a los estudiantes en las dificultades de aprendizaje y por motivar una reflexión crítica sobre el proceso de enseñanza y aprendizaje. Según Cabero Almenara et al la función más efectiva del docente para promover el aprendizaje auténtico en los entornos LMS y situar al estudiante en el centro del proceso de enseñanza y aprendizaje es su capacidad de construir y utilizar

recursos didácticos que con una consistente combinación entre las diferentes herramientas que brinde el LMS y la variedad de actividades que puedan realizarse tanto en el aula como fuera de ella [17]. Como señala Silva "La formulación, diseño e implementación de e-actividades son parte del diseño instruccional online, pueden responder a diferentes finalidades como: la motivación inicial hacia la materia; las formativas orientadas a la consecución de objetivos; competencias o resultados de aprendizajes; las evaluativas, que permiten constatar el nivel de progreso de los estudiantes" [19].

Según Salinas un entorno virtual de aprendizaje se presenta como un ámbito para promover la enseñanza y el aprendizaje a partir de procesos de comunicación multidireccionales (docente/alumno - alumno/docente y alumnos entre sí). Se trata de un ambiente de trabajo compartido para la construcción del conocimiento en base a la participación activa y la cooperación de todos los miembros del grupo, promoviendo el aprendizaje colaborativo y minimizando las barreras temporales y espaciales [20].

Las actividades que el docente propone en el ambiente de aprendizaje pueden ser o no significativas para el estudiante, de acuerdo con diferentes variables: tipo de materia, materia inicial o avanzada en el plan de estudios, contexto en el que se presenta, valor practico y motivacional, entre otras características. Los materiales didácticos que el docente incorpore a la plataforma deben fomentar y promover una diversidad de actividades en el estudiante para promover la adquisición de los conocimientos y en particular que favorezcan el aprendizaje colaborativo.

La interactividad es una característica fundamental para los entornos de enseñanza y aprendizaje virtuales y según Bartolomé Pina "es la posibilidad de que el emisor y el receptor permuten sus respectivos roles e intercambien mensajes" citado por Moya [21].

En este trabajo, se realizó un análisis del significado de los datos obtenidos de la plataforma Moodle utilizada en la Facultad de ingeniería como entorno propio de LMS y se contrasta junto a las notas finales de los alumnos de las tres cohortes intervinientes ya mencionadas. Se evaluaron patrones, características y variables que puedan incidir en evitar casos de deserción de estudiantes y contribuir a mejorar el rendimiento académico de alumnos de esta modalidad de cursada especialmente para alumnos ingresantes que estaban realizando la materia desde sus lugares de residencia previo al COVID-19. Se definieron variables de mayor impacto y las que sean menos relevantes que incluso fueron utilizadas como insumos cuando se debió migrar a la virtualidad por la pandemia.

Surgieron preguntas y planteamientos iniciales como, por ejemplo:

¿Existe correlación entre la actividad e interacción de los estudiantes en la plataforma virtual y el rendimiento académico final? ¿Se puede anticipar algún comportamiento de los alumnos en la plataforma virtual, que los docentes puedan guiar y mejorar, para evitar deserciones tempranas en cursos de matemática inicial?

Estos interrogantes son parte de las hipótesis planteadas en la propuesta del trabajo de tesis de Magister, son parte de muchos planteamientos surgidos en este periodo de pandemia que estamos atravesando.

4.1 Metodología de investigación

Desde el punto de vista de la investigación se analizaron los resultados obtenidos y las categorías encontradas en la aplicación de las analíticas de aprendizaje (Fig.1). Para esto se trabajó, en coincidencia con Giraldo Ocampo, con los siguientes enfoques [22].

El enfoque descriptivo: se corresponde con el análisis inicial de los datos, para contextualizar el dominio en el que se trabaja, los datos con que se cuenta y cómo se obtienen.

El enfoque diagnóstico: tiene por objeto conseguir una serie de resultados iniciales, puede estar representado en la ejecución de consultas relacionales o multidimensionales sobre los datos. Las consultas

relacionales corresponden a la extracción de conocimiento de almacenamiento de datos de tipo de relacional. El lenguaje de programación usado para la generación de este tipo de consultas es el SQL.

El enfoque predictivo: Este se concentra en tratar de mostrar lo que podría suceder a partir del análisis de los datos que se tienen y por medio de la aplicación de técnicas más complejas, como por ejemplo analíticas de aprendizaje.

El enfoque prescriptivo: hace referencia a las recomendaciones sobre qué se debe hacer después de los hallazgos obtenidos por medio de la adopción de estrategias que permitan atacar los puntos identificados como débiles o incentivar los aspectos positivos.

Validación del modelo: Según los objetivos planteados se estudiarán los registros de datos de tres cohortes de alumnos ingresantes.

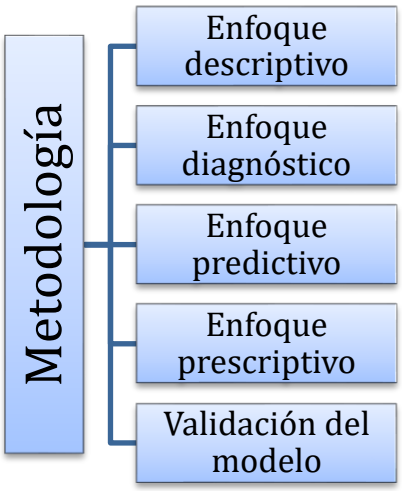


Fig. 1. Aspectos tenidos en cuenta en la metodología de investigación utilizada para el análisis de los datos

4.2 Resultados

En el presente trabajo se expondrán algunos datos preliminares del análisis correspondiente a la tesis que está en proceso de redacción. Las muestras se obtuvieron relevando los datos necesarios de las tablas que Moodle provee. Esto se realizó con un dump de la base de datos y luego los registros suministrados se sometieron a una limpieza y contrastación de estos de una manera aleatoria para su validación de calidad y luego de este proceso se definieron aquellos que se utilizarían en el procesamiento y análisis.

Después de realizar un importante proceso de recolección, depuración, normalización y procesamiento de los datos obtenidos de los tres cursos de ingresantes respecto de la participación en cada uno de los Foros de Preguntas y Respuestas habilitados en cada curso de la materia, respecto de los cuatro capítulos de los temas abordados en Matemática para ingeniería, se realizaron tablas de comparación para comparar y analizar lo más relevante.

Un foro, en coincidencia con Sanz y Zangara, es un espacio virtual comunicativo y/o colaborativo en el que todo un grupo toma parte en un debate sobre un tema que sea de interés general. Un objetivo básico es lograr la participación de los integrantes en el debate de temas específicos, reflexionar, y compartir informaciones desde y a todo el grupo. Otros objetivos, no menos importantes, son: socializar las

producciones entre todos los participantes, debatir sobre diversos aspectos planteados tanto por los docentes como por los propios estudiantes, y estimular el pensamiento creativo en la solución de problemas [23].

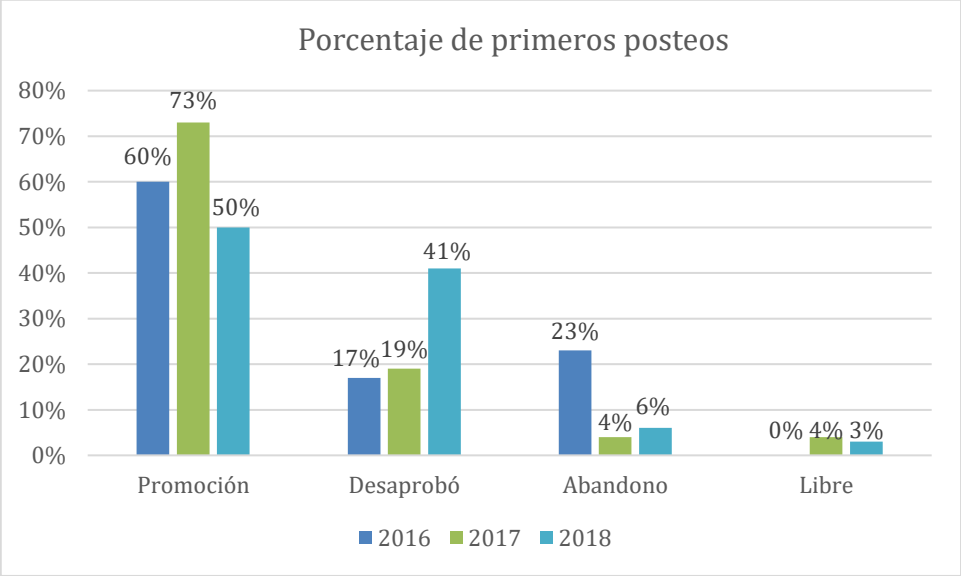
A continuación, se resumen los primeros datos y se observan algunos patrones.

Tabla 2. Análisis cuantitativo de los resultados académicos y la participación en los foros

	Cantidad de foros	Cantidad de "primeros" posteos	Cantidad total de posteos
2016	**59**	**148**	**420**
Abandonó	8	34	89
Desaprobó	10	25	64
Libre	0	0	0
Promocionó	41	89	267
2017	**12**	**167**	**306**
Abandonó	1	6	12
Desaprobó	4	32	58
Libre	1	7	17
Promocionó	6	122	219
2018	**281**	**333**	**1053**
Abandonó	44	20	72
Desaprobó	70	135	448
Libre	55	10	24
Promocionó	112	168	509

Se analizaron la cantidad de posteos o intervenciones de los alumnos en los foros y en particular los “primeros posteos” o sea aquellos posteos que inician un tema en cada uno de los foros por considerar que son aquellas intervenciones de alumnos más activos, más emprendedores y cuestionadores. Este proceso se considera similar en cierta medida con el comportamiento en un aula de clases presencial de alumnos activos cuando un profesor plantea un desafío o explicación en el pizarrón intentando recolectar la opinión y participación de los alumnos presentes. Según Vygotsky citado por Archundia-Sierra E., et al y la teoría constructivista, el aprendizaje es el resultado de la interacción del individuo con el entorno. En este contexto, el proceso de aprendizaje se espera que se convierta en un proceso activo, y pase de una simple memorización pasiva de información que se recibe a un proceso de reconstrucción de esta, por tanto, la nueva información se integra y correlaciona con el conocimiento ya existente. La motivación, la intervención, el trabajo colaborativo, las emociones, las actitudes, la interacción y la cooperación entre pares generan compromiso para el desarrollo de nuevas ideas, que inducen a la innovación en el aula [24].

Tabla 3. Análisis porcentual de los primeros posteos y su relación con el resultado académico de los alumnos



Se puede observar que en todos los cursos los alumnos que realizan los primeros posteos de los foros, mayoritariamente son alumnos que promocionan la materia en su gran mayoría. Se puede considerar que este comportamiento de los alumnos muestra una actitud y característica importante académicamente al momento de caracterizarlos.

En este tipo de cursos donde el trabajo en los foros fue muy importante para el desarrollo de los contenidos, la participación en ellos fue fundamental. Además, esta materia concibe la participación autónoma y activa de los alumnos frente a la matemática. Este aspecto en un curso a distancia es difícil de promover y de sostener en el tiempo, más aún con alumnos que son futuros ingresantes a la Facultad, pero aún no cursaron ninguna materia universitaria.

Cabe destacar que en estos cursos analizados no hubo nunca un encuentro sincrónico con los alumnos que realizaban la materia. Esta situación fue clara al inicio del curso de modalidad a distancia y entendiendo que los alumnos que realizaban estos cursos eran alumnos que estaban realizando el colegio secundario.

Es notorio que los alumnos con condición final de libre son los que menos participación realizaron.

Después de destacar el porcentaje de participación de los alumnos promocionados, se puede visualizar que continúan en cantidad los alumnos que desaprueban la materia, o sea aquellos alumnos que ocupan todas las instancias de evaluación, pero aun así no promocionan. Esto se considera comprensible desde el punto de vista que son alumnos que mantuvieron mayor actividad en la plataforma que los alumnos que abandonan, ya que estos últimos son alumnos que solo ocuparon al menos una instancia de evaluación y luego dejaron de participar en la plataforma y en la materia.

5 Reflexiones y acciones a futuro

Una vez realizado el primer análisis de los datos recolectados se concluye que, para la modalidad de trabajo impartida en “matemática para Ingeniería”, la participación en los foros de los alumnos participantes es

fundamental. Por supuesto siempre hay alumnos que son muy independientes y pueden estudiar y aprender casi solos, pero en su gran mayoría necesitan de la guía y acompañamiento docente y de sus pares para avanzar en el contenido de la materia. Es importante destacar que los estudiantes son aspirantes a una carrera de ingeniería y que, al realizar la primera materia en la universidad, tienen una preparación desigual en matemática según provienen de diferentes tipos de colegios secundarios, de diferentes ciudades y diferentes provincias. Estos patrones detectados serán evaluados y considerados en futuros análisis con los datos de las tres cohortes mencionadas. También es importante destacar que el tipo de materia y contenidos abordados en ella, complejizan en un alumno ingresante su trabajo autónomo y su comunicación a través de la plataforma por las fórmulas matemáticas, los procesos de modelización y la abstracción de las herramientas estudiadas. A pesar de ello, fue de fundamental importancia este trabajo realizado durante los años relatados y la experiencia transmitida a los docentes de la materia al inicio de la cuarentena por la pandemia, ya que no todos los docentes habían utilizado plataformas educativas virtuales ni tenían estudios o experiencias con cursos en modalidad a distancia.

Referencias

1. Mariana Maggio: Enriquecer la enseñanza. Los ambientes con alta disposición tecnológica como oportunidad. Buenos Aires, Argentina, Editorial Paidós (2016)
2. Ander-Egg, E.: El taller. Una alternativa de renovación pedagógica. Editorial Magisterio del Río de La Plata (1991)
3. Pasel, S.; Asborno, S.: Aula-Taller. Aique (1993)
4. García, Mabel M: Propuesta de aula taller en el Curso de Nivelación para el ingreso a Ingeniería. Facultad de Humanidades y Ciencias de la Educación de la UNLP (2017) http://www.memoria.fahce.unlp.edu.ar/tesis/te.1532/te.1532.pdf
5. Cabanillas, J.L., Luengo, R. y Torres, J.L: Diferencias de actitud hacia las TIC en la formación profesional en entornos presenciales y virtuales. Píxel-Bit. Revista de Medios y Educación, 55, pp. 37-55 (2019)
6. R.Scorzo, A.Favieri, B.Williner: Desarrollo de un espacio de enseñanza aprendizaje para realizar actividades con uso de software en una cátedra numerosa. TE&ET. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, n.º 21 pp. 77-83, (2018)
7. A. Bartolomé Pina: 15´ Notas & Entrevistas: Haría falta una formación muy intensa para el cambio de paradigma [Online] https://ladaga.net/notas/ (2020, April 22)
8. R. Di Domenicantonio y L. Langoni: Coordinación de materias masivas de Matemática en la Facultad de Ingeniería de la UNLP durante la pandemia COVID-19, TEyET, n.º 28, p. e20, (2021)
9. Zapata-Ros, M.: Analítica de aprendizaje y personalización. Recuperado de http://uajournals.com/ojs/index.php/campusvirtuales/article/view/41 (2013)
10. Baker, R: Educational data mining: An advance for intelligent system in education. IEEE Intelligent Systems, pp.78-82 (2014)
11. Jaramillo, A. y Paz Arias, H: Aplicación de técnicas de minería de datos para determinar las interacciones de los estudiantes en un entorno virtual de aprendizaje. Revista Tecnológica ESPOL, 28, 1, pp. 64-90 (2015)
12. Long, P. y Siemens, G.: Penetrating the Fog: Analytics in Learning and Education. Recuperado de https://bit.ly/2UvNXuA (2011)
13. Márquez Vera, C: Predicción del fracaso y el abandono escolar mediante técnicas de minería de datos. Directores: Ventura Soto y Romero Morales, Programa II de la UAPUAZ, Zacatecas, México, Universidad de Córdoba, pp. 39-63 (2015)
14. Huang, S. & Fang, N.: Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers & Education, 61, pp. 133–145 (2013)
15. Hu, Y., Lo, C. & Shih, S.: Developing early warning systems to predict students online learning performance. Computers in Human Behavior, 36, pp. 469–478 (2014)

16. Salgado Reyes, N., Beltrán Morales, J., Guaña Moya, J., Escobar Teran, Ch., Nicolalde Rodríguez, D. y Chafla Altamirano, G.: Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. RISTI, 17, 258-266 (2018)
17. Cabero, J., Del Prete, A.C., Aranciabia, M.L.: Modelo para determinar acciones de calidad en la formación virtual Modelo para determinar acciones de calidad en la formación virtual Digital Education Review. 37. 323-342. 10.1344/der.2020.37.323-342 (2020)
18. Ruiz, C. y Dávila, A.A.: Propuesta de buenas prácticas de educación virtual en el context universitario. RED-Revista de Educación a Distancia, 49 (12). DOI: 10.6018/red/49/12 (2016)
19. Silva Silva, J.: Un modelo pedagógico virtual centrado en las E-actividades. RED. Revista de Educación a Distancia, 53(10), 1-20, pp.7-8. DOI: 10.6018/red/53/10 (2017).
20. Salinas, M.: Entornos virtuales de aprendizaje en la escuela: tipos, modelo didáctico y rol del docente. Recuperado de http://www.uca.edu.ar/uca/common/grupo82/files/educacion-EVA-en-la-escuela_web-Depto.pdfGiraldo
21. Moya, M.M: La utilización de los foros en la enseñanza de la matemática mediada por tecnología digital (2008)
22. Ocampo, M.: Descubrimiento de patrones en interacciones entre estudiantes y plataformas virtuales de educación mediante el uso de analíticas de aprendizaje. Tesis de Doctorado, Universidad Nacional de Colombia, Sede Medellín (2016)
23. Sanz, C.V. y Zangara, M.A.: Los foros como espacios comunicacionales-didácticos en un curso a distancia. Una propuesta metodológica para aprovechar sus potencialidades. http://sedici.unlp.edu.ar/handle/10915/22535 (2006)
24. Archundia-Sierra E., et al: Redes de aprendizaje digital en nodos colaborativos. 2020, pp. 64 - 72.

ISBN 978–987–633–574–4

Estrategia Metodológica para la Comprensión de Textos Científicos en Programación Numérica

Lorena Elizabeth Del Moral Sachetti

Facultad de Ciencias Exactas. Sede Regional Orán. Universidad Nacional de Salta
lorena.dms.7@gmail.com

Resumen. Uno de los cuestionamientos que se realizan sobre la formación universitaria, es la pobre utilización del pensamiento crítico, reflexivo e investigativo a lo largo del trayecto universitario y posterior ámbito profesional. Es por ello que se debe fomentar la capacidad de lectura crítica, comprensión y reflexión. De esta manera, serán capaces de reflexionar sobre las ideas y los hechos, descubrir intenciones e ideologías y adoptar puntos de vista, y así construir conocimientos específicos propios de una comunidad científico disciplinar. Frente a estas circunstancias se pretende aportar con el presente artículo una estrategia metodológica para la lectura y comprensión de textos científicos y académicos, utilizando los mapas conceptuales como herramienta para plasmar los conceptos, descripciones y relaciones de saberes, extraídos, analizados e interpretados por los alumnos. Dicha estrategia se puso en práctica con alumnos de Programación Numérica de la carrera Licenciatura en Análisis de Sistemas, de la Universidad Nacional de Salta.

Palabras Claves: Estrategia Metodológica. Lectura crítica. Mapas conceptuales. Programación Numérica.

1 Introducción

El presente trabajo se lleva a cabo en el marco del proyecto de investigación Nro. 2.536/19 “Rediseño educativo para el aprendizaje del cálculo numérico”, aprobado en el año 2.019 por el Consejo de Investigación de la Universidad Nacional de Salta.

El proyecto se centra en el rediseño de las estrategias de enseñanza y aprendizajes tradicionales, en el que se busca un proceso que permita organizar y desarrollar nuevas actividades pedagógicas que satisfaga las necesidades formativas de los estudiantes en el nuevo mundo de la Sociedad de la Información. El rediseño del proceso de formación se realiza en el contexto de las asignaturas Programación Numérica y Cálculo Numérico que se dicta en el segundo año de las carreras Licenciatura en Análisis de Sistemas y Licenciatura en Matemática, ambas en la Universidad Nacional de Salta.

En este contexto Actualmente, uno de los cuestionamientos más relevantes, que se hacen sobre la formación universitaria, es la enseñanza y la utilización del pensamiento crítico, reflexivo e investigativo a lo largo de toda la formación y

posterior ámbito profesional. Una de las causas de esta problemática es la falta de lectura de textos académicos y científicos.

2 Problemática Observada

Durante algunos años, se viene observando en los alumnos diferentes falencias en relación a lo discursivo del ambiente científico de la Programación Numérica como consecuencia de la poca o nula lectura por parte de los alumnos de textos académicos y científicos.

Los alumnos leen cada vez menos, y se acostumbran a estudiar solo de los apuntes que toman de las clases de sus profesores. Si bien los recursos bibliográficos siempre estuvieron disponibles, ya sea en la biblioteca de la sede, como en archivos digitales en la plataforma o en fotocopias que ofrecía el profesor, estos recursos casi nunca fueron utilizados.

La resolución de los trabajos prácticos se apoya solamente en los ejercicios explicados por los docentes de clases prácticas.

Se observaba que el conocimiento adquirido, acerca de los temas del programa eran bastante recortados (acotado, limitado, deficiente). Además de que los alumnos no cultivaban el saber científico e investigativo de la materia, por lo que carecían de un propio pensamiento reflexivo (indagatorio) y analítico acerca de los temas de la materia. Aceptaban el conocimiento tal y como se los enseñaban en clase.

Reconociendo la problemática descripta, nos propusimos pensar alguna metodología (estrategia) para enseñar los modos específicos sobre cómo encarar los textos científicos y académicos de la Programación Numérica, pensada como disciplina académica o campo de estudio. Es decir, como un gran saber científico, investigativo y en creciente desarrollo y no (acotado) solo como un recorte acorde a lo que puede enseñarse en un cuatrimestre.

Es por ello, que se decidió insistir en la lectura y comprensión de textos científicos y académicos. Sin embargo, se han observado algunos obstáculos que dificultan en los alumnos la comprensión de la bibliografía. A continuación, se exponen algunos:

- Los alumnos están familiarizados con textos del nivel secundario. Estos escritos carecen de argumentos y justificaciones científicas, posturas de diferentes autores, controversias, entre otras limitaciones. Es decir, que estos textos solo se limitan a exponer el saber listo para ser memorizado. Quitándoles a los alumnos la posibilidad de razonar y reflexionar sobre lo que han leído, debatir acerca de lo que han comprendido e interpretar de manera diferente los contenidos expuestos. Al respecto Carli (2010), expone en una de sus investigaciones que los textos de secundaria: "…tratan al conocimiento como histórico, anónimo, único absoluto y definitivo. La cultura lectora de la educación secundaria exige aprender que dicen los textos, restándole importancia al porqué lo dicen y como lo justifican".
- Los textos científicos están dirigidos a personas que conocen y entienden acerca del saber que se está exponiendo en el texto. Es decir, estos textos están dirigidos a colegas, que comparten conocimientos,

modos de pensamientos, formas de argumentar y exponer, métodos para justificar el saber, etc. Y los alumnos desconocen por completo estos "códigos" del saber científico. Autores y lectores comparten, por su formación, gran parte del conocimiento que en estos textos se da por sabido (Sinclair,1993) Comparten el conocimiento de otros autores que estos textos mencionan al pasar, comparten el conocimiento de las corrientes más amplias a las que pertenecen ciertas posturas que aparecen sólo esbozadas. Como el escritor está inmerso en una discusión y un debate compartidos, no necesita poner de manifiesto sus ideas más allá de lo imprescindible dentro de su comunidad. (Fernández et al., 2002).

- Los estudiantes se enfrentan a textos que no desarrollan todos los conocimientos que exponen. Estos textos dan por supuesto muchos saberes que los alumnos no recuerdan (en el mejor de los casos, si corresponden a conceptos estudiados en materias anteriores) o no disponen (cuando los autores hacen referencia a posturas de otros autores sin explicarlas). Es decir, estos textos no explican esas "otras" cosas, ya que constituyen un marco conceptual dado por sabido.
- Los alumnos no saben qué esperan los docentes que ellos hagan cuando se encuentran frente a la bibliografía (Boise State Vardi, 2000) Son exigencias por parte de los docentes que se dan por sabido. Los docentes esperan que actúen de acuerdo a un específico modelo de lector que no tienen internalizado todavía. (Carli, 2000)

Estas cuestiones descriptas, son algunas de las que su pudieron observar entre los alumnos de Programación Numérica. Sin embargo, se pueden seguir exponiendo un sinfín de cuestiones que obstaculizan la lectura comprensiva de los alumnos en la universidad, sobre ello hay varias investigaciones llevadas a cabo por docentes de Argentina, que se pueden leer si se quiere interiorizarse aún más en el tema.

Es por ello, que se debe tratar de enseñar junto a los contenidos que se imparten, a leer como miembros de la comunidad disciplinar de la Programación Numérica.

3 Estrategia Metodológica

A continuación, se numeran los principales objetivos que se persiguen con la implementación de la metodología:

1) Favorecer en los alumnos la lectura en el nivel universitario.
2) Ser conscientes de las dificultades de lectura de los universitarios.
3) Propiciar la lectura compartida en las clases, ayudando a comprender lo que los textos dan por sobreentendido.
4) Contribuir el aprendizaje y apropiación del conocimiento disciplinar propio del nivel universitario, especialmente de la Programación Numérica.

La Metodología de trabajo se estructura en 5 etapas bien diferenciadas: 1) Antes de la lectura, 2) Durante la lectura, 3) Después de la lectura, 4) Diseño del mapa conceptual, y 5) Presentación del mapa conceptual. Cada una de estas tiene actividades propias, que se detallan más adelante.

Las actividades de las diferentes etapas, fueron puestas en acto luego de finalizar el estudio formal de cada unidad del programa. Es decir que, una vez que los contenidos teóricos y prácticos fueron brindados, los alumnos resolvieron el practico correspondiente y realizaron un coloquio. Esto fue necesario para brindarles las bases principales del conocimiento científico. Los alumnos, debían haber entendido de antemano los principales conceptos de la unidad, los procesos, los fundamentos, la nomenclatura comúnmente usada, etc. Todo ello iba a servir como bagaje de ideas previas para enfrentarse a la lectura de un material científico. Además, ya lo expresa Carlino: "…leer es un proceso de resolución de problemas. Esto significa que lo que un lector obtiene de la lectura depende de sus conocimientos previos".

Cabe destacar que todas las actividades se han realizado primero con el docente y el grupo-clase. Y en una segunda instancia, lo han realizado solo los alumnos, pero con la guía y ayuda del docente.

A continuación de describen brevemente las actividades de las etapas de la metodología llevada a cabo:

1) Antes de la lectura

a) Brindar al alumno precisiones sobre cómo analizar el texto. Es decir, proporcionar y explicar las "categorías de análisis", con las cuales se espera que operen sobre el texto. Es evidente que cada texto, tendrá sus propias categorías de análisis, sin embargo, a modo de ejemplo se presentan algunas que pueden ser usadas de manera general.

- *¿Puedes describir la estructura del texto, identificando títulos, subtítulos, gráficos, ejemplos, etc.?*
- *¿Cuáles son los conceptos y categorizaciones que se desarrollan?*
- *¿Cuáles fórmulas y modelos matemáticos están presentes? ¿Puedes interpretarlos? ¿Existen analogías con otras notaciones que conozcas? ¿Cuales?*
- *¿Cuáles ejemplos utiliza el autor para mostrar o describir una técnica particular? Trata de desarrollarlo y explicarlo.*
- *¿Puedes identificar referencias a otros textos o autores? ¿Cuáles? ¿Con que propósito se los utiliza?*
- *¿Con que textos leídos previamente, puedes relacionar algunos conceptos expuestos?*

b) Hacer circular entre los estudiantes los libros originales (de texto), para que puedan indagar acerca de los capítulos precedentes y posteriores, el índice, el prólogo, la introducción, las solapas que presentan a los autores, las contratapas que comenten el texto y otra información que pueda desprenderse, de tan solo la manipulación del ejemplar. Esto fue posible, ya que en la mayoría de los casos se trabajó con libros que estaban disponibles en biblioteca o con libros propios de los docentes. Cuando no se puede trabajar con el ejemplar original, se trabaja con material fotocopiado de

calidad, es decir legibles, en los cuales no se dificulta la visualización de lo impreso. Además, resulta necesario que el docente agregue todos los datos bibliográficos del material (autor, año, editorial, capítulo, etc.) de manera que el lector pueda ubicarse dentro de lo que lee. Es decir que, el lector debe familiarizarse con lo que se va a leer.

c) Activar el conocimiento previo sobre el tema del texto. Buena parte de esta actividad, ya se realiza con la actividad anterior, ya que los alumnos tienen la posibilidad de manipular los textos y leer títulos y subtítulos, con lo cual ya saben sobre que van a leer. También se puede fomentar un dialogo con los estudiantes acerca de lo que recuerdan sobre el tema que van a leer.

2) Durante la lectura

a) Desenvolver las ideas que en los textos están condensadas (Carlino). Es decir, tratar de disgregar partes, elementos o conceptos, para luego entender el todo del que forman parte y de qué manera están relacionados.

b) Motivar al alumno la búsqueda en otras bibliografías (leídas o no), en diccionarios específicos o en apuntes o clases de la cátedra los "conceptos confusos o difíciles", que se han encontrado en el texto que están leyendo. El docente, puede ayudar a los alumnos, de la misma manera que se espera que ellos actúen. Es decir, buscando y explicando estos "conceptos confusos o difíciles" haciendo referencia a la bibliografía. Se trata de no explicar directamente en la pizarra, sino de buscar en la bibliografía o en apuntes, (seguramente ya sabemos dónde). Esto es de suma importancia, ya que le brindamos al alumno todo un contexto (o marco) científico en el cual nos desenvolvemos. La importancia de esta actividad también radica en la relación de conceptos que puede encontrarse en diferentes líneas de investigación, puntos de vista de los autores, o libros académicos, que ayudan a completar, comprender y a tener una visión más general del concepto que estamos buscando.

c) Propiciar la realización de anotaciones y/o resaltados sobre lo que se va leyendo. Pueden ser conceptos que no se llegan a entender, interrogantes sobre el tema, palabras o conceptos claves, etc.

d) Desarrollar los ejemplos presentes en los textos, que sirven para la comprensión de un procedimiento, técnica o fórmula.

3) Después de la Lectura

a) Parafrasear las ideas principales de cada párrafo del texto. Es decir, el alumno debe ser capaz de explicar con sus propias palabras lo que ha entendido del texto. Para ello, puede usar las anotaciones que se han realizado al costado de la hoja. La importancia de esta actividad radica no solo en la posibilidad de que el alumno muestre lo que ha sido capaz de comprender y el uso del lenguaje técnico, sino también en la retroalimentación que puede realizarse, si es que hay conceptos erróneos, confusos o incompletos.

b) Propiciar discusiones colectivas, acerca de conceptos que no han quedado claro, y contando con la posibilidad de revisar nuevamente el texto para

corroborar datos, confirmar, validar o rectificar lo que no fue interpretado correctamente.

4) Diseño del mapa conceptual

Para el desarrollo del mapa conceptual se puede usar alguna herramienta de software específica como cmaptools, alguna herramienta de uso general como Word o Power Point, o simplemente puede desarrollarse con lápiz y papel. Lo importante en esta etapa es que el alumno sea capaz de diseñar una estructura de conceptos relacionados de manera coherente. En esta etapa el alumno vuelve a releer el texto, poniendo en juego actividades cognoscitivas superiores. Ya que debe estructurar y relacionar conceptos, términos, descripciones, fórmulas, etc. ordenadas gráficamente.

Se exige a los alumnos, que, una vez finalizado el mapa, lo revisen y lean, sin acudir al texto científico. De esta manera, se pretende saber si entienden lo que trataron de explicar.

5) Presentación del mapa conceptual

En esta última etapa el alumno debe presentar su mapa conceptual al resto del grupo-clase. Con esta actividad, el alumno pone en juego su capacidad de explicar el tema a los demás, apoyado en un recurso visual desarrollado por él mismo. Se ponen de manifiesto también la oralidad con el uso del lenguaje técnico adquirido durante el proceso de lectura.

4 Análisis de Resultados

En líneas generales, la metodología implementada arrojó buenos resultados, en cuanto a la calidad de los aprendizajes adquiridos por los alumnos, y también en cuanto a las capacidades comunicativas desarrolladas por estos, y ello teniendo en cuenta que se trata de alumnos de segundo año (en este periodo, suelen ser todavía bastante tímidos). Los docentes culminaron la cursada realmente satisfechos de haber participado de esta experiencia docente tan enriquecedoras. Algunos de los aspectos más relevantes, es que los alumnos lograron:

- Apropiarse del lenguaje técnico específico de cada uno de los temas de la materia.
- Favorecer la conexión entre diferentes conceptos.
- Propiciar el crecimiento de su bagaje de conocimientos relacionados a la materia.
- Mejorar gradualmente la expresión escrita y oral.
- Facilitar la explicación de un concepto mediante una herramienta gráfica, dejando de lado lo memorístico.

Desde la óptica de los alumnos, al principio se resistían un poco a las actividades de lectura, debido a que sostenían, que eran textos complicados a los que no estaban acostumbrados a leer. La mayoría tenía la creencia de que no iban a poder entender o comprender. Sin embargo, en el transcurso del cuatrimestre, esta visión fue cambiando, porque se dieron cuenta de que, no iban a trabajar solos, sino que

contaban todo el tiempo con la guía del docente. De manera que, el grado de compromiso de todos fue aumentando. Finalmente, finalizaron la cursada, muy contentos con la forma de trabajo, ya que no solo les era más fácil aprender los conceptos, sino que eran capaces de enfrentarse a los textos científicos desde otra posición.

Desde la posición de los docentes, si bien fue un trabajo satisfactorio, debido a los resultados que se lograron, lo complicado de esta forma de trabajo, es que se requiere invertir más tiempo, en la selección de los textos, la adquisición de los libros, la preparación de las categorías de análisis, etc. Todo ello se realizaba con cada texto que se seleccionaba. También hubo que programar (y reprogramar) las clases dedicadas a la lectura, y las exposiciones orales de los alumnos.

5 Conclusiones y Acciones Futuras

La metodología descripta se ha puesto en práctica durante el segundo cuatrimestre de los años 2019 y 2020. De allí, que se persigue como objetivo inmediato, refinar algunas actividades, de manera de mejorar la estrategia en pos de brindar una mejor calidad educativa.

Se pretende también, implementar la metodología en conjunto con otras cátedras. Para llevar a cabo acciones conjuntas que posibiliten a los alumnos de la carrera una apropiación a los textos científicos de las diferentes materias y el desarrollo del pensamiento científico reflexivo y crítico, en todo el trayecto de formación universitaria.

6 Referencias

1. Barberis A. R. and Del Moral L. E. (2016). Scrum como Herramienta Metodológica en el Entrenamiento Cooperativo de la Programación: De la Teoría a la Práctica. Proceedings of XI Congreso de Tecnología en Educación y Educación en Tecnología 2016 (TE&ET 2016). pp. 365-374. Red UNCI. Universidad de Morón, Argentina.

2. Carlino, Paula, "Leer textos científicos y académicos en la educación superior: obstáculos y bienvenidas a una cultura nueva", 6º Congreso Internacional de Promoción de la Lectura y el Libro, mayo 2.003.

3. Steiman, J. y Melone, C. (2000) Algunos recursos didácticos para el trabajo con textos en la educación superior. Ficha de cátedra: Didáctica IV, Facultad de Ciencias Sociales, Universidad Nacional de Lomas de Zamora.

4. Olson, D. (1998) El mundo sobre el papel. El impacto de la lectura y la escritura sobre la estructura del conocimiento. Barcelona: Gedisa. Edición original en inglés de 1994.

5. Marucco, M. (2001) “La enseñanza de la lectura y la escritura en el aula universitaria”. Ponencia presentada en las I Jornadas sobre La lectura y la escritura como prácticas académicas universitarias, organizadas por el Departamento de Educación de la Universidad Nacional de Luján, Buenos Aires, junio de 2001. Disponible en Internet en: www.unlu.edu.ar/~redecom/

6. Fernández, G., Izuzquiza, M. V. y Laxalt, I. (2002) “¿Eneseñanza de prácticas de lectura en la universidad?”. Ponencia presentada en el Tercer encuentro: La universidad como objeto de investigación. La Plata, 24 y 25 de octubre de 2002, Fac.de Humanidades y Ciencias de la Educación, UNLP.

7. Ferreiro, E., Castorina, J. A., Goldin, D. y Torres, R. M. (1999) Cultura escrita y educación. Conversaciones con Emilia Ferreiro, México, Fondo de Cultura Económica.

8. Carlino, P. (2003 a) “Alfabetización académica: Un cambio necesario, algunas alternativas posibles”. Educere, Revista Venezolana de Educación, Vol. 6 Nº 20 (ISSN 1316-4910). Universidad de Los Andes, Mérida, enero-febrero-marzo de 2003, 409-420. Disponible también en Internet en: http://www.saber.ula.ve/db/saber/Edocs/pubelectronicas/educere/vol6num20/articul7.pdf

9. Arnoux, E., Di Stefano, M. y Pereira, C. (2002) La lectura y la escritura en la universidad. Buenos Aires, Eudeba.

10. Benvegnú, M. A., Galaburri, M. L., Pasquale, R. y Dorronzoro, M. I. (2001) “La lectura y escritura como prácticas de la comunidad académica”. Ponencia presentada en las I Jornadas sobre La lectura y la escritura como prácticas académicas universitarias, organizadas por el Departamento de Educación de la Universidad Nacional de Luján, Buenos Aires, junio de 2001. Disponible en Internet en: www.unlu.edu.ar/~redecom/

ISBN 978–987–633–574–4

Propuesta didáctica para el aprendizaje de la especificación de requisitos

Lía G. Rico[1], María Fernanda Villarrubia[1], Laura R. Villarrubia[1]

[1] Calle Ítalo Palanca N°10 - Cátedra de Sistemas de Información - Facultad de Ingeniería - UNJu - Jujuy

ricogalia@gmail.com, marvillisi@gmail.com, l.r.villarrubia@gmail.com

Resumen. Este trabajo presenta una propuesta didáctica que incluye la herramienta de medición benchmarking para identificar los requisitos funcionales de los sistemas de información. La misma se llevó a cabo en la asignatura Sistemas de Información, correspondiente al 4to. año de las carreras Ingeniería Informática, Licenciatura en Sistemas e Ingeniería Industrial de la Facultad de Ingeniería de la Universidad Nacional de Jujuy. El desarrollo presenta el marco conceptual, la descripción de las actividades de la propuesta didáctica, la experiencia de cátedra y los resultados obtenidos. Finalmente, en las conclusiones se realiza una valoración de la experiencia, en base a los resultados observados y se proyecta el trabajo a futuro.

Palabras clave: requisitos funcionales, benchmarking, cadena de valor

1 Introducción

La asignatura en la que se llevó a cabo esta experiencia, se denomina Sistemas de Información. Entre las actividades que realizan los estudiantes, se encuentra al inicio y, como base del estudio, el análisis de una organización, con el fin de determinar el alcance de los sistemas de información que brindarán soporte a sus procesos, a partir de identificar los requisitos funcionales.

Para especificar los requisitos, el estándar IEEE 830-1998 [1] menciona al cliente como el actor fundamental que define los servicios que debe proporcionar un sistema. Ante la situación de pandemia COVID-19 disponer del cliente para el relevamiento de datos no fue posible, y se replantea esta actividad a través de una nueva propuesta didáctica. La misma se sostiene sobre dos ejes, el primero consiste en delimitar el ámbito de estudio focalizando el análisis de una actividad primaria de la cadena de valor de la organización, y el segundo en incorporar conceptos del proceso estructurado benchmarking.

Esta propuesta pretende ser una contribución al proceso de especificación de requisitos funcionales para el desarrollo de un sistema informático, con el fin de que

el estudiante visualice necesidades funcionales claras y consistentes de la actividad en estudio.

2 Marco teórico

2.1 Requisitos funcionales

"Los requisitos funcionales son declaraciones de los servicios que debe proporcionar el sistema, de la manera en que este debe reaccionar a entradas particulares y de cómo se debe comportar en situaciones particulares. En algunos casos, los requisitos funcionales también pueden declarar explícitamente lo que el sistema no debe hacer" [2]

Los requisitos funcionales deben ser: [3]

- Precisos: la ambigüedad a la hora de definir los requisitos puede conducir a dobles interpretaciones por parte de los desarrolladores y los clientes, lo que se acaba traduciendo en problemas.
- Completos: deben incluir la descripción de todos los servicios y características requeridas.
- Consistentes: no puede haber contradicciones en la descripción de los servicios y características del sistema.

"Los Requisitos No Funcionales son restricciones de los servicios o funciones ofrecidas por el sistema. Incluyen restricciones de tiempo, sobre el proceso de desarrollo y estándares. Los Requisitos No Funcionales, a diferencia de los Requisitos Funcionales, se caracterizan por no estar de forma directa vinculados a las funciones del sistema, sino a las propiedades de este y a determinadas restricciones" [3].

2.2 Cadena de Valor

Se denomina cadena de valor a las principales actividades de una empresa comparadas éstas con los eslabones de una cadena; las actividades van añadiendo valor al producto a medida que éste pasa por cada una de ellas. Todas las empresas, cualquiera sea su rubro, cuentan con una cadena de valor, conformada por actividades, que van desde el diseño del producto y la obtención de insumos hasta la distribución del producto y los servicios de post venta.

Esta herramienta clasifica las actividades generadoras de valor de una empresa en dos: ***Actividades primarias o de línea***: Son aquellas actividades que están directamente relacionadas con la producción y comercialización del producto, son:

- Logística interior (de entrada): actividades relacionadas con la recepción, almacenaje y distribución de los insumos necesarios para fabricar el producto.

- Operaciones: actividades relacionadas con la transformación de los insumos en el producto final.
- Logística exterior (de salida): actividades relacionadas con el almacenamiento del producto terminado, y la distribución de éste hacia el consumidor.
- Mercadotecnia y ventas: actividades relacionadas con el acto de dar a conocer, promocionar y vender el producto.
- Servicios: actividades relacionadas con la provisión de servicios complementarios al producto tales como la instalación, reparación y mantenimiento del mismo.

Actividades de apoyo o de soporte: Son aquellas actividades que agregan valor al producto pero que no están directamente relacionadas con la producción y comercialización de éste, sino que más bien sirven de apoyo a las actividades primarias.

- Infraestructura de la empresa: actividades que prestan apoyo a toda la empresa, tales como la planeación, las finanzas y la contabilidad.
- Gestión de recursos humanos: actividades relacionadas con la búsqueda, contratación, entrenamiento y desarrollo del personal.
- Desarrollo de la tecnología: actividades relacionadas con la investigación y desarrollo de la tecnología necesaria para apoyar a las demás actividades.
- Aprovisionamiento: actividades relacionadas con el proceso de compras.

La cadena de valor permite identificar las fortalezas y debilidades de una empresa, compararla con empresas competidoras, detectar fuentes potenciales de ventajas competitivas, y comprender mejor el comportamiento de los costos [4]

2.3 Benchmarking

"Tague (2005) define benchmarking como un proceso estructurado que permite comparar las mejores prácticas de las organizaciones, de manera que se pueden incorporar aquellas que no se desarrollan o mejorar las que se desarrollan a la propia organización, o a los procesos de la organización."

Las fases para desarrollar un benchmarking es el siguiente:

1. Planificar:
    a. Definir los objetivos del estudio, aquellos que sean críticos para el éxito organizacional.
    b. Formar un equipo multidisciplinar
    c. Estudiar los procesos de la organización
    d. Identificar los profesionales de la organización que podrían desarrollar las mejores prácticas
2. Recopilar datos: haciendo uso de las técnicas y prácticas del relevamiento de información.
3. Analizar: Comparar los procesos de la organización con los de otras organizaciones, determinar las diferencias en sus medidas de rendimiento, e identificar las prácticas que provocan dichas diferencias.
4. Adaptar: Desarrollar los objetivos de los procesos de la organización, desarrollar planes de acción para conseguir esos objetivos, implementarlos y controlarlos. [5]

3 Experiencia

El objetivo de la propuesta es que el estudiante considere el proceso de benchmarking como marco de referencia, para la especificación de requerimientos.

La propuesta incluye dos fases:

1- ***Organización para el desarrollo del trabajo práctico*.**

1.1 Definición del caso de estudio: El caso describe el objetivo, estrategia a mediano plazo y actividades generales de una empresa que produce y distribuye productos lácteos. Es una organización de mediana envergadura. No se comentan los detalles de la operación ni organización de la misma; puesto que los alumnos, al aplicar la guía que se propone, podrán consultar páginas y libros para conocer los procesos con mayor profundidad.

1.2 Conformación de grupos de trabajo: Los grupos son interdisciplinarios y se conforman con estudiantes de las tres carreras que cursan esta asignatura. La cátedra viene trabajando de esta manera porque obtuvo buenos resultados. Los estudiantes aprenden a colaborar, a valorar las otras carreras y profesiones, identifican mejor el rol que les corresponde en el juego de los sistemas informáticos, al diferenciarse y compararse con sus pares.

1.3 Configuración y organización de salas virtuales para el trabajo en grupo: Se realizó usando la aplicación BigBlueButton, provista por la plataforma virtual Moodle de la Universidad; la misma permite gestionar salas de reuniones de manera eficiente. [6]

1.4 Asignación de la actividad primaria de la cadena de valor, a los grupos, para su estudio:

Figura 1. Asignación de actividades primarias por grupos.

Las actividades primarias de la cadena de valor que se asignaron a los grupos son Operación, Logística Externa y Marketing/Ventas como se visualiza en la figura 1. La

actividad Logística Interna se adoptó por los docentes para explicar y ejemplificar la guía de actividades solicitada a los estudiantes que se detalla en el siguiente punto.

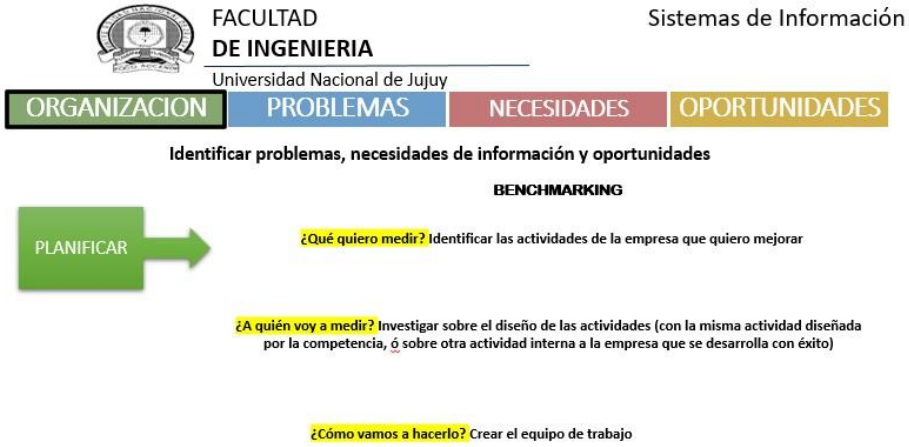


Figura 2. Pasos para aplicar benchmarking.

Para resolver cada actividad propuesta en el segundo eje, los grupos deberán investigar las empresas del rubro relacionado con el caso dado, aplicando los pasos que se observan en la figura 2. A continuación se describe en detalle esta actividad.

2- ***La guía de actividades*** **para especificar los requisitos funcionales de los sistemas de información para la organización descrita en el caso. Cada actividad se ejemplifica con una muestra extraída de los trabajos prácticos de los estudiantes.**

Actividad 1: **Definir el objetivo y buenas prácticas de la actividad de la cadena de valor asignada.**

El primer paso para iniciar el estudio de la actividad primaria es, a través de investigación, conocer con claridad el fin o meta que persigue y las acciones que benefician, tanto a dicha actividad, como a las entidades relacionadas con la misma.

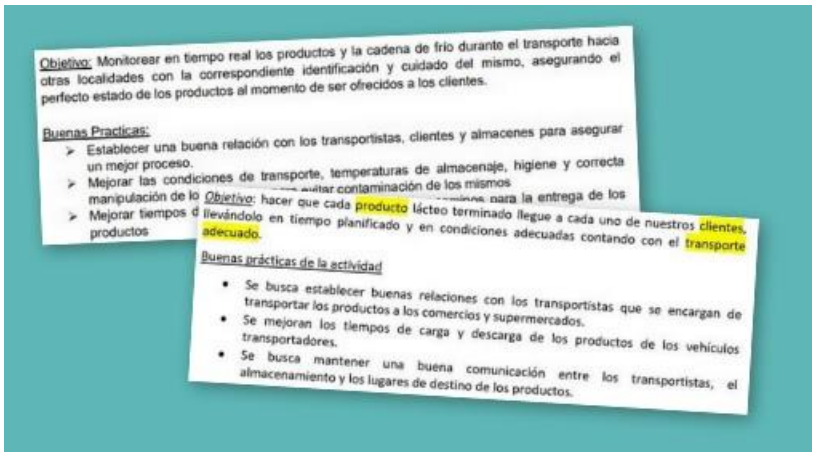


Figura 3. Objetivo y buenas prácticas. Ejemplos desarrollados por alumnos.

Actividad 2: **Identificar los procesos funcionales de la actividad primaria**

Luego de entender el fin y las buenas prácticas, se procede a investigar los procesos funcionales que componen la actividad utilizando la técnica SADT (Structured Analysis and Design Technique) [7]

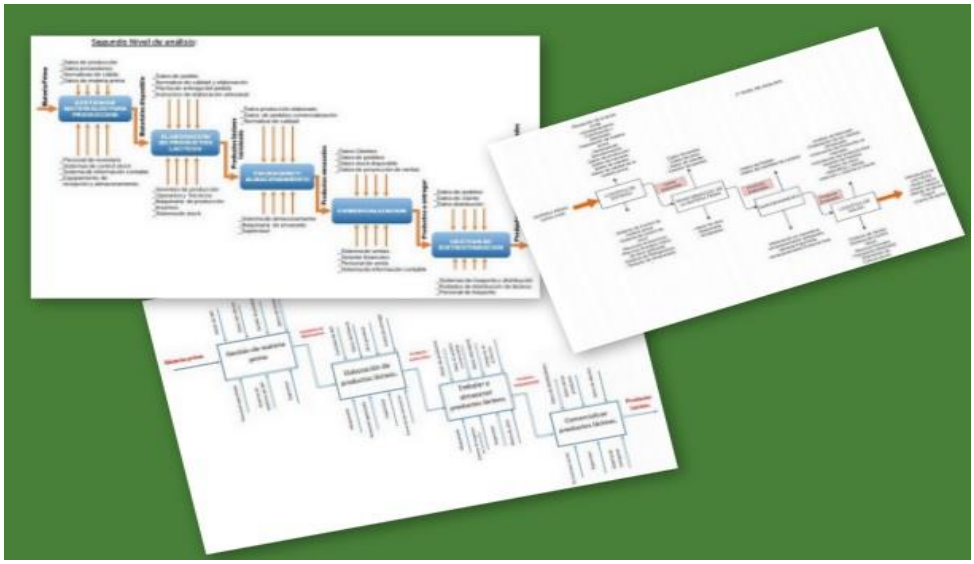


Figura 4**.** Diseño de procesos. Collage de imágenes de trabajos desarrollados por alumnos.

Teniendo en cuenta las entradas, salidas, mecanismos y controles diseñados en el SADT se especifican los procesos funcionales de la actividad primaria.

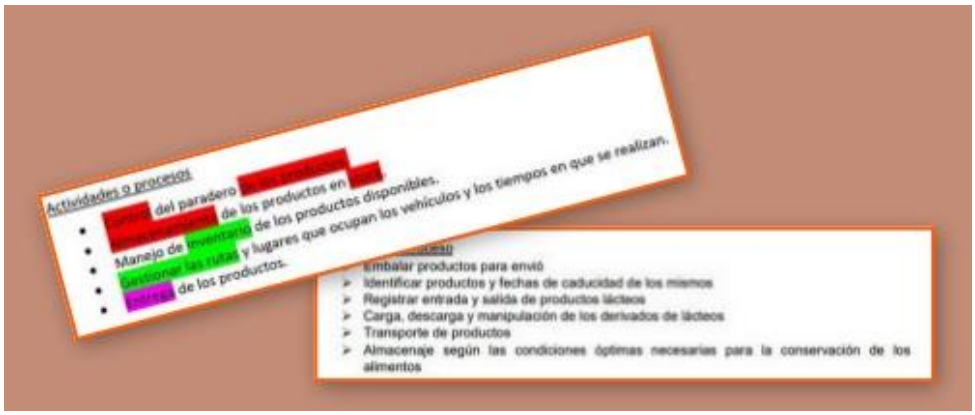


Figura 5**.** Procesos funcionales. Collage de imágenes de trabajos desarrollados por alumnos.

<u>Actividad 3:</u> Problemas y necesidades de información que se presentan en la actividad primaria.

El próximo paso, luego de tener claridad sobre el funcionamiento de los procesos que componen la actividad, es investigar e identificar los problemas y necesidades más frecuentes de información.

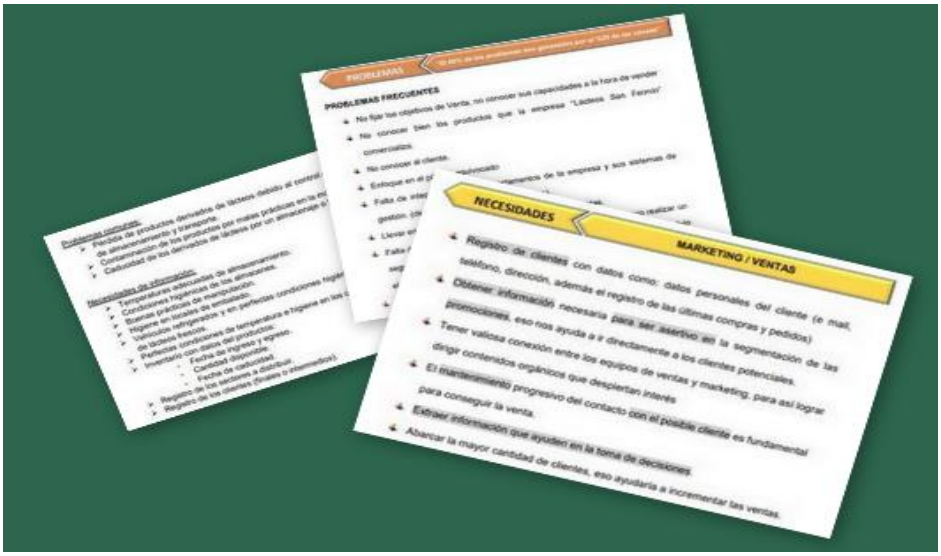


Figura 6. Problemas y necesidades de Información. Collage de imágenes de trabajos desarrollados por alumnos.

<u>Actividad 4</u>: **Alcance funcional**

Teniendo en cuenta el objetivo, los procesos, las necesidades y problemas de información se identifican las palabras claves asociadas que facilitarán realizar una descripción general de las funciones principales del sistema de información.

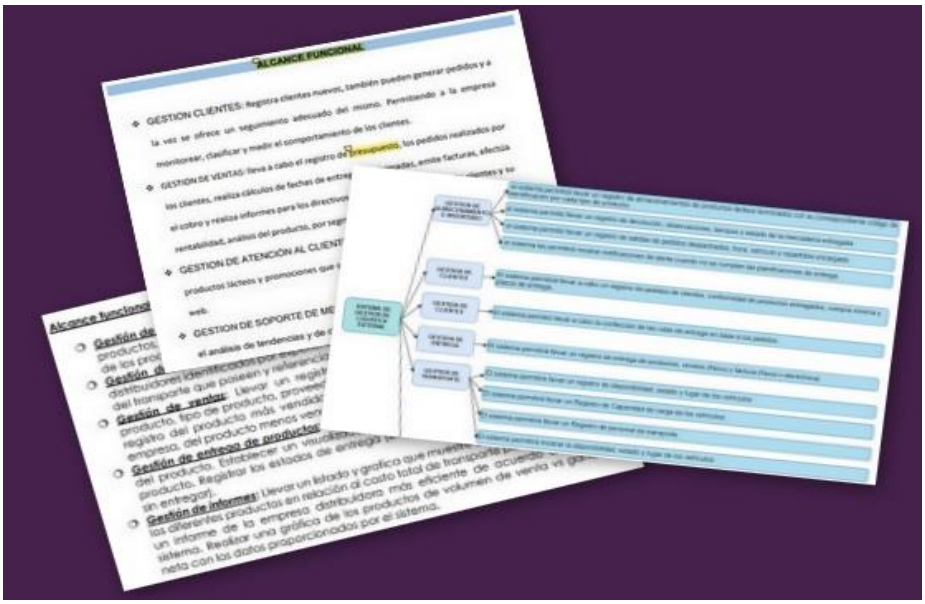


Figura 7. Alcance Funcional. Ejemplos de trabajos elaborados por los alumnos.

<u>Actividad 5</u>: **Especificar requisitos funcionales para el alcance propuesto**

Finalmente, a partir del alcance del sistema de información, se realiza la especificación de los requisitos funcionales del sistema de información.

Figura 8. Requisitos funcionales. Collage de imágenes de Catálogos de requisitos de los trabajos prácticos de los alumnos.

4 Resultados obtenidos

Los resultados que se presentan en esta sección responden a la observación, análisis y evaluación del docente sobre los trabajos desarrollados por los estudiantes.
A partir de la implementación de la propuesta didáctica, las competencias logradas por los estudiantes son:

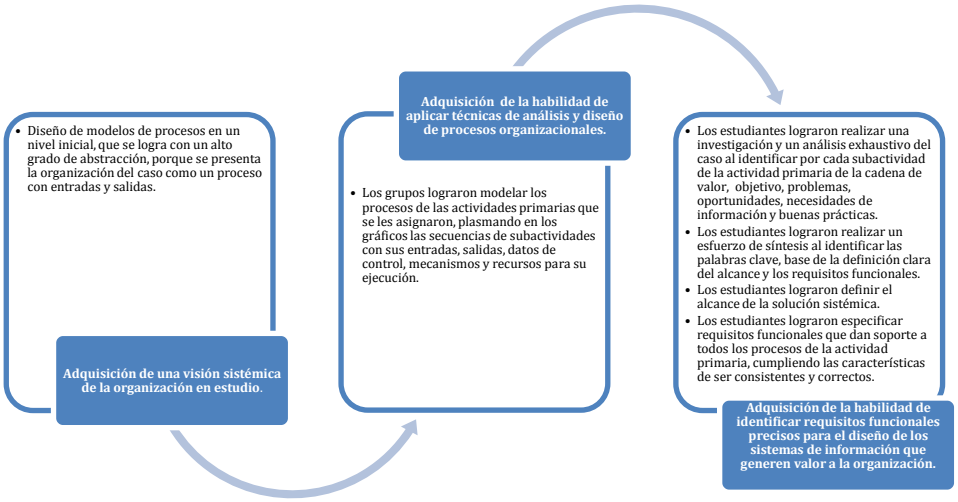


Figura 9. Competencias adquiridas por los estudiantes y sus evidencias.

- **Adquisición de una visión sistémica de la organización en estudio. La cual se evidencia en:**

  . Establecieron el objetivo y las buenas prácticas, para lograr entender la actividad primaria de la cadena de valor y poder identificar los sub-procesos de la misma.

. Lograron realizar una investigación y un análisis exhaustivo del caso, al identificar para la actividad primaria de la cadena de valor, los problemas, oportunidades y necesidades de información.

- Adquisición de la habilidad de aplicar técnicas de análisis y diseño de procesos organizacionales.

Lograron modelar los procesos de las actividades primarias que se les asignaron, plasmando en los gráficos las secuencias de subactividades con sus entradas, salidas, datos de control, mecanismos y recursos para su ejecución. Este modelado se logra con un alto grado de abstracción.

Lograron describir, en forma general, el funcionamiento de cada proceso identificado para la actividad primaria.

.

- Adquisición de la habilidad de identificar requisitos funcionales precisos para el diseño de los sistemas de información que generen valor a la organización.

. Los estudiantes lograron realizar un esfuerzo de síntesis al identificar las palabras clave sobre los objetivos, buenas prácticas, necesidades de información y problemas, que permiten mayor claridad para la definición del alcance y los requisitos funcionales.

. Los estudiantes lograron definir el alcance de la solución sistémica.

. Los estudiantes lograron especificar requisitos funcionales que dan soporte a todos los procesos de la actividad primaria, cumpliendo las características de ser precisos, completos y consistentes.

En términos generales se observa que la propuesta didáctica exige que el estudiante investigue por distintos medios toda la información referida a la actividad sobre la que se requiere conocer los servicios que debe proporcionar el sistema. Se observa el interés creado en el estudiante por cumplir con las actividades propuestas por el docente y el fortalecimiento actitudinal durante el proceso de desarrollo del trabajo.

4 Conclusiones y líneas de trabajo

El contexto influenciado por el COVID-19 nos llevó a rediseñar las prácticas docentes para lograr los objetivos de la cátedra.

La introducción de conceptos como cadena de valor, para focalizar el ámbito de estudio, y benchmarking, como herramienta de medición y comparación, posibilitaron la obtención y análisis de la información necesaria para la especificación de requisitos funcionales, reemplazando en parte la realización de técnicas y prácticas tradicionales del relevamiento de datos.

Los estudiantes lograron realizar una especificación de requisitos funcionales de forma precisa, ágil, con mayor grado de detalle y ajustada a la actividad primaria asignada.

Si bien, en la práctica se podrá ir depurando la didáctica y mejorando aún más los resultados, se puede tomar la guía de actividades presentada como una innovación en la enseñanza de los sistemas de información.

Como trabajo a futuro se continuará la búsqueda e investigación de modelos y herramientas de calidad y de gestión que colaboren en el proceso de especificación de requisitos, con el fin de lograr mejores desarrollos de sistemas de información.

Bibliografía

1. Méndez, Gonzalo (2008) Especificación de Requisitos según el estándar de IEEE 830. IEEE Std. 830-1998. Extraído de https://www.fdi.ucm.es/profesor/gmendez/docs/is0809/ieee830.pdf

2. Sommerville, Ian (1998). Ingeniería del software. Séptima Edición. Pearson Educación S.A.

3. Encinas, Gonzalo M. de las Pueblas. Trabajo de fin de grado para la obtención del Título de Ingeniería en Tecnologías Industriales. Definición de requisitos funcionales bajo especificación IEEE para un sistema de ingeniería. Junio 2019.

4. Porter, Michael (1991). Ventaja competitiva. Creación y Sostenimiento de un Desempeño Superior. Editorial Rei Argentina S.A.

5. Piattini, Velthus, Mario G., García, Rucio, Félix O., Caballero, Muñoz-Reja, Ismael (2007). Calidad de sistemas informáticos. Primera Edición. Alfaomega Grupo Editor. S.A. de C.V., México - ISBN 978-970-15-1267-8

6. Bigbluebutton. Solución de conferencia web de código abierto para el aprendizaje en línea. Extraído de: https://moodle.com/es/certified-integrations/bigbluebutton/

7. Consejo Superior de Informática (2001). Metodología MÉTRICA Versión 3. Extraído del Portal de Administración electrónica (PAe). Gobierno de España: https://administracionelectronica.gob.es/pae_Home/dam/jcr:da7d91fa-d6bd-467c-be32-a72e27c603b3/METRICA_V3_Tecnicas.pdf

8. Martínez, Sandra, Oliveros Alejandro, Zuñiga, Javier, Corbo, Sergio, Forradelas, Patricia. “Aprendizaje de la elicitación y especificación de requerimientos”. Extraído de: .http://sedici.unlp.edu.ar/bitstream/handle/10915/42142/Documento_completo.pdf?sequence=1&isAllowed=y

9. Oliveros, Alejandro, Zuñiga, Javier, Wehbe, Ricardo, Rojo, Silvana del Valle, Sardi, Florencia. Enseñanza de elicitación de requerimientos. Extraído de: http://sedici.unlp.edu.ar/bitstream/handle/10915/23852/Documento_completo.pdf?sequence=1&isAllowed=y

10. Gomez Vieites A. & C. Suarez Rey, Sistemas de Información- Herramientas prácticas para la gestión empresarial, Alfaomega, México, 2007.

ISBN 978-987-633-574-4
ISBN 978-987-633-575-1

CACIC 2021

WORKSHOP SEGURIDAD INFORMATICA

COORDINADORES

Javier Diaz (UNLP)
Hugo Ramón (UNNOBA)
Claudio Aciti (UNCPBA)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978–987–633–574–4

Un método de ensamble basado en subsecuencias a nivel de palabras para la autenticación de usuarios con cadencias de tecleo en textos libres

Nahuel González[1*], Jorge S. Ierache[1], Enrique P. Calot[1], and Waldo Hasperué[2]

[1] Laboratorio de Sistemas de Información Avanzados, Facultad de Ingeniería, Universidad de Buenos Aires, Ciudad AutBuenos Aires, Argentina
{ngonzalez,jierache,ecalot}@lsia.fi.uba.ar

[2] Instituto de Investigación en Informática (III-LIDI) Facultad de Informática Universidad Nacional de La Plata Investigador Asociado – Comisión de Investigaciones Científicas (CIC)
whasperue@lidi.info.unlp.edu.ar

Resumen Utilizando sólo los tiempos entre eventos de presión y liberación de teclas, es posible construir un segundo factor de autenticación basado en la cadencia de tecleo tanto para endurecer claves de usuario como para verificar la identidad en forma continua dentro de una muestra de texto libre. Dentro de este último caso se propone un método de ensamble para la autenticación de usuarios que utiliza subsecuencias a nivel de palabras. En contraste con otros métodos del estado del arte, que alcanzan tasas de error cercanas al 5 % con entrenamientos muy reducidos del orden de 250 caracteres, nuestro método demanda grandes cantidades de información para el entrenamiento inicial, en el orden de 50.000 caracteres, pero alcanza un EER más bajo, cercano al 3,6 %, al ser evaluado con un conjunto de datos públicamente accesible y capturado en condiciones realistas.

Keywords: seguridad informática, biometría comportamental, cadencias de tecleo, texto libre, aprendizaje automático, métodos de ensamble

1. Introducción

Las sutiles variaciones en la forma en que distintas personas teclean son suficientes para revelar su identidad. Hace cuarenta años, Gaines et al. [1], pioneros del análisis de cadencias de tecleo, reconocieron la utilidad de este fenómeno para la autenticación de usuarios. Utilizando sólo los tiempos entre eventos de presión y liberación de teclas, es posible construir un segundo factor de autenticación para endurecer claves de usuario [2] y para verificar identidad en forma

* Autor para correspondencia: Nahuel González (ngonzalez@lsia.fi.uba.ar)

ISBN 978–987–633–574–4

continua con texto libre [3]. Más recientemente, el análisis de cadencias de tecleo también ha encontrado usos fuera del dominio de la seguridad informática; por ejemplo, descubrir ciertas características fisiológicas o impedimentos clínicos del usuario [4], e incluso determinar en forma aproximada las variaciones de su estado emocional mientras escribe, basándose en autoreporte [5] o aplicando una interfaz cerebro-máquina para etiquetar las muestras [6].

El análisis de cadencias de tecleo ha dejado atrás una infancia difícil. Hace más de diez años eran muchos los problemas metodológicos que aquejaban a la disciplina, como la ausencia de conjuntos de datos públicamente accesibles de tamaño suficiente y el uso inconsistente de métricas de error incompatibles al expresar los resultados [7]. La posibilidad de plantear experimentos comparativos reproducibles y generalizables a las condiciones del mundo real se encontraba muy limitada y los estudios más rigurosos se restringían a métodos sencillos para verificar claves estáticas, reportando tasas de error del orden del 10 % [8]. Verificar textos libres demandaba muchas muestras muy extensas, con más de 800 caracteres, para acercarse a una precisión aceptable [9].

El estado de situación actual es, inversamente, satisfactorio y alentador. Hoy contamos con muchos conjuntos de datos enormes y públicamente accesibles; por ejemplo, [10] abarca casi 200.000 usuarios y contiene más de 136 millones de caracteres de texto libre capturado en condiciones realistas. A la par con muchas otras disciplinas relacionadas, el análisis de cadencias de tecleo ha integrado los métodos generales de aprendizaje automatizado en detrimento de técnicas *ad hoc*. Al éxito de este abordaje lo ilustra un estudio reciente, que empleando una sofisticada red neuronal recurrente del tipo siamesa alcanza un EER del 5 % al autenticar texto libre, aún restringiendo el entrenamiento a sólo 250 caracteres e incluso luego de escalar el sistema a más de 100.000 usuarios [11].

La cuestión que aquí nos compete es el problema dual. Mientras los autores del anterior exploran cuánto puede reducirse el tamaño del conjunto de entrenamiento sin comprometer las tasas de error y la escalabilidad del método, nosotros nos preguntamos cuánto puede reducirse la tasa de error si permitimos crecer al conjunto de entrenamiento. Proponemos un método de ensamble que utiliza subsecuencias a nivel de palabras, derivado de un estudio exploratorio previo de los mismos autores [12] sobre las correlaciones internas de los tiempos entre eventos de tecleo dentro de fronteras semánticas. Esperamos motivar procedimientos híbridos, que al combinar métodos de autenticación de convergencia rápida con aquellos asintóticamente óptimos logren, a la vez, tasas de error aceptables con mínimo entrenamiento y tasas de error óptimas en el largo plazo, cuando la plantilla biométrica del usuario cuente con suficientes muestras.

Contribuciones. El objetivo de este estudio es proponer un método de ensamble para autenticación con cadencias de tecleo en textos libres y evaluar su rendimiento cuando se cuenta con grandes cantidades de información, en la forma de muestras de escritura, para cada usuario. Las principales contribuciones ofrecidas son:

- Proponemos un método de ensamble para la autenticación de usuarios basada en cadencias de tecleo en textos libres, que fragmenta las muestras a

ISBN 978–987–633–574–4

verificar en subsecuencias a nivel de palabras y utiliza clasificadores individuales para cada una de ellas.
- Evaluamos el método propuesto sobre un conjunto de datos públicamente accesible, de gran extensión, capturado en condiciones realistas, y que ha sido utilizado en estudios anteriores [13].
- Ofrecemos en forma abierta los conjuntos de datos de entrenamiento y de resultados [14, 15] para permitir la verificación independiente y para facilitar ulteriores exploraciones y mejoras de este tipo de métodos.

Organización. El resto del artículo está organizado como se describe a continuación. La sección 2 reseña brevemente algunos estudios previos sobre el tema. La sección 3 describe el método propuesto. La sección 4 detalla la metodología del experimento, incluyendo el conjuntos de datos utilizado, el preprocesamiento y la limpieza de los datos, el proceso de clasificación, y la disponibilidad de los conjuntos de datos y resultados. La sección 5 discute los resultados. Finalmente, la sección 6 resume las conclusiones.

2. Estudios previos

Si bien existen antecedentes como el de Monrose y Rubin [2] para la autenticación de usuarios con cadencias de tecleo en textos libres, se trata más bien de estudios exploratorios. Sólo a partir de los trabajos de Bergadano, Gunetti, y Picardi [16] con la métrica R se alcanzan tasas de error equiparables a la verificación con claves estáticas o textos fijos. Sin embargo, esta métrica requiere muestras muy grandes, de más de 800 caracteres [9], para alcanzar resultados óptimos. También se ha observado que, al replicar el experimento utilizando un conjunto de datos capturado en condiciones realistas (en contraste con condiciones de laboratorio), las tasas de error se elevan sobremanera, hasta cuatro o cinco veces los valores reportados originalmente [17]. El método de modelado con contextos finitos [18] logra sortear esta última dificultad, alcanzando tasas de error óptimas en torno a los 250 caracteres y sin que estas se degraden notoriamente con la dificultad del conjunto de datos de evaluación.

El empleo de técnicas de aprendizaje automático para la autenticación de usuario en base a su cadencia de tecleo tiene una larga historia. Entre otros, Yu y Chao [19] han utilizado atributos derivados y un clasificador SVM para la tarea, Obaidat [20] ha explorado diversos tipos de redes neuronales, y Killourhy y Maxion [21] han aplicado bosques aleatorios pero en el caso de PINs.

El exponente más actual de aprendizaje automático aplicado a la autenticación de usuarios utilizando cadencias de tecleo es el de Ancien et al. [11]. Este estudio destaca no sólo por el método y su rendimiento sino también por la dificultad del protocolo de evaluación, en el que el clasificador propuesto sorprende con bajas tasas de error. Los autores proponen la utilización de una red neuronal recurrente, del tipo siamesa, con dos capas LSTM de 128 neuronas. Lo más sorprendente es la escasa cantidad de información por usuario utilizada para entrenar la red neuronal; hay sólo 15 muestras por usuario en el conjunto de datos

ISBN 978–987–633–574–4

de evaluación elegido, que entre ellas suman no mucho más de 250 caracteres. El conjunto de datos [10] cuenta con unos 200.000 usuarios y aproximadamente 136 millones de caracteres en total, lo que lo hace óptimo para evaluar la posibilidad de escalar a tamaño masivo los sistemas de autenticación por medio de cadencias de tecleo. Los autores reportan un EER de 4,8 % para mil usuarios, con una única muestra de evaluación por usuario, de aproximadamente 50 caracteres. Al incrementar la cantidad de usuarios por encima de 100.000, el rendimiento decrece un 5 % en términos relativos.

Los métodos de ensamble han sido utilizados extensivamente en tareas de aprendizaje automático y sus aplicaciones. Probablemente los bosques aleatorios, que no necesitan presentación ulterior, sean la implementación más reconocida. Hasta donde alcanza nuestro conocimiento de la literatura del tema, no se han ensayado métodos de ensamble para la autenticación de usuarios con cadencias de tecleo en textos libres como el que aquí se propone, excepto como clasificadores enlatados luego de un proceso de extracción de atributos [21].

3. Método propuesto

Supongamos que contamos con una muestra $M = \{K, R, L\}$ de texto libre, de largo m, y queremos verificar que pertenezca al usuario legítimo, utilizando su perfil biométrico que cuenta con muestras pasadas. Es este un problema de clasificación binaria, ya que las únicas respuestas posibles son sí o no. La secuencia de teclas de M es $K = k_1 \dots k_m$, sus tiempos de retención (intervalos entre el evento de presión y liberación de cada tecla) son $R = r_1 \dots r_m$ y sus tiempos de latencia (intervalos entre eventos de presión de teclas sucesivas) son $L = l_1 \dots l_m$.

Sea E un conjunto de caracteres que incluye la tecla espacio, teclas de puntuación, caracteres especiales, etc. Particionamos M en las posiciones de todos los caracteres que pertenecen a E y descartamos las subsecuencias vacías, para obtener un ensamble de palabras P_i, con sus subsecuencias de teclas, tiempos de retención, y de latencia. Por ejemplo, si $K =$ *hola, mundo, soy ng123.* y $E = \{\ ,.\}$, tenemos que $P_1 =$ *hola*, $P_2 =$ *mundo*, $P_3 =$ *soy*, y $P_4 =$ *ng123*. En particular y salvo que se indique lo contrario, presupondremos que E se compone de todos los caracteres no alfanuméricos y, por lo tanto, que las palabras resultantes de la partición son secuencias alfanuméricas ininterrumpidas.

Ahora queremos autenticar, independientemente, cada palabra P_i y luego combinar los resultados para responder si M pertenece a un usuario legítimo o a un impostor. Nuestro objetivo es entrenar un clasificador para cada P_i de la muestra M de este usuario utilizando observaciones de la misma palabra en otras muestras de este y otros usuarios, que serán tratados como impostores.

Particionamos las muestras existentes en el perfil del usuario legítimo y recolectamos todas las observaciones pasadas disponibles de cada palabra P_i, generando para cada una de ellas una instancia de entrenamiento de la forma $r_1 \dots r_{m_i} l_2 \dots l_{m_i}$, en donde m_i es el largo de P_i. Estas instancias contienen $2m_i - 1$ atributos, uno para cada tiempo de retención y uno para cada laten-

ISBN 978–987–633–574–4

cia, exceptuando la de la primera tecla. El primer tiempo de latencia se excluye pues corresponde al intervalo entre la tecla especial anterior (que puede ser cualquiera) y la primer tecla de P_i; no es representativa de la palabra en sí, y no presenta la misma consistencia entre observaciones [12] que las demás. Todas estas instancias de entrenamiento se rotulan con la clase *legítimo*. Para generar los instancias de impostores, empleamos una colección de muestras de otros usuarios, que una vez más particionamos en la misma forma a nivel de palabra para extraer tantas muestras de cada P_i como instancias de entrenamiento del usuario legítimo tengamos. De esta forma, mantenemos balanceadas las clases simplificando la tarea del clasificador. A estas otras instancias de entrenamiento las rotulamos con la clase *impostor*.

Finalmente, con el conjunto de instancias de legítimo e impostor resultantes de ambos procesos, entrenamos el ensamble de clasificadores y luego registramos los veredictos para las correspondientes P_i. Cada veredicto individual de cada P_i otorga un voto a la decisión global del ensamble para la muestra M, que se define por mayoría de votos. La evaluación de estrategias de ponderación de los votos se plantea como una futura línea de investigación en la sección 5.1.

Si no contamos con suficientes observaciones de alguna palabra en el perfil del usuario (o entre las muestras de impostores) para generar un conjunto de entrenamiento, se elimina la P_i correspondiente del ensamble. Hemos utilizado un umbral de diez observaciones requeridas como mínimo.

4. Evaluación experimental

4.1. El conjunto de datos

Para este trabajo se ha utilizado el conjunto de datos LSIA de [17, 18], actualizado para incluir muestras adicionales capturadas desde entonces. El mismo está compuesto de muestras de texto libre, ingresadas en un teclado convencional. Se han registrado las teclas presionadas y los tiempos de retención (intervalo entre evento de presión y evento de liberación de tecla) y latencia (intervalo entre eventos de presión de teclas sucesivas) con precisión de milisegundos, junto con la identidad del usuario correspondiente. Debido a ciertas restricciones de la plataforma de captura, en ocasiones los tiempos fueron redondeados a múltiplos de 8 o 16 milisegundos.

Las muestras fueron capturadas en un entorno realista durante más de cuatro años, con usuarios de ambos sexos en un rango de edad entre 28 y 60 años, y aptitud para la escritura con grandes variaciones. El texto corresponde a lenguaje natural compuesto durante el transcurso de la labor cotidiana de los usuarios. Luego del preprocesamiento y limpieza de los datos descrita en la sección siguiente, quedaron disponibles 7897 muestras de 79 usuarios para la evaluación del método aquí propuesto.

Disponibilidad pública. Tanto el conjunto de datos de entrenamiento como el de resultados se encuentran a disponibilidad del público en forma abierta y gratuita, en los repositorios de Mendeley Data [14] y IEEE DataPort [15].

ISBN 978–987–633–574–4

4.2. Preprocesamiento y limpieza de los datos

Las muestras del conjunto de datos fueron preprocesadas por medio de una herramienta propia para experimentos de cadencias de tecleo, con el objetivo de convertir a un formato abierto el esquema binario propietario en el cual se encuentra almacenada la información original. Para cada muestra de cada usuario, se utilizó el proceso de partición descrito en la sección 3, descartándose todas aquellas P_i con algún tiempo de retención o latencia que faltara, tuviera valores negativos, o superara los 3000 milisegundos; este último criterio es para eliminar las pausas que no se corresponden con el ritmo normal de escritura.

Para cada usuario, cada muestra, y cada P_i que cumpliera los criterios anteriores, se generó un archivo CSV con $2m - 1$ columnas, en donde m es el largo de la palabra, y una fila por cada instancia de entrenamiento, tanto para los rótulos legítimo como impostor. Los tiempos de retención y latencia del P_i en consideración, o de otras repeticiones de la misma palabra en la muestra, no se incluyeron en el conjunto de entrenamiento para no contaminar este con la instancia a clasificar, lo que sesgaría el sistema hacia una menor tasa de error que la alcanzable en un caso real. Así, aunque una palabra aparezca en varias muestras de un cierto usuario, para cada una de ellas el conjunto de entrenamiento difiere, pues no se incluye ninguna subsecuencia de la muestra actual; incluirlas sería hacer trampa. Todos los usuarios restantes fueron considerados como potenciales impostores, y sus muestras disponibles para extraer instancias de entrenamiento y evaluación con rótulo impostor. Como de esta forma es esperable que haya muchas más observaciones de cada palabra entre las muestras de impostores, se realizó un muestreo aleatorio de las mismas hasta recolectar tantas observaciones como del usuario legítimo haya disponibles.

Finalmente, se generó un archivo CSV para cada usuario y cada muestra, conteniendo la lista de palabras que no fueron rechazadas y sus tiempos de retención y latencia, para ser evaluadas por los clasificadores individuales.

4.3. Clasificación

El siguiente proceso se realizó para cada usuario y cada muestra, obteniendo una lista de clasificaciones para cada P_i, junto con el valor de exactitud obtenida al evaluar el modelo correspondiente. La salida de la etapa de clasificación para cada usuario y cada muestra es un archivo CSV, cuyas filas enumeran la palabra evaluada, la exactitud del modelo, y el rótulo asignado (legítimo o impostor).

Para cada P_i que haya sobrevivido a los filtros, se utilizó la implementación de bosques aleatorios `RandomForestClassifier`, versión 0.24.2, de la librería `scikit-learn` [22] para entrenar un clasificador con el conjunto de datos de entrenamiento generado en la etapa de preprocesamiento. El motivo de esta elección puede hallarse en un estudio anterior de los autores [12], donde se realiza una comparación de rendimiento para la tarea de verificación de palabras individuales con distintos clasificadores, en donde los bosques aleatorios obtienen una precisión similar a las redes neuronales con una fracción del costo computacional. No se realizó escalado y normalización de los atributos en las instancias

ISBN 978–987–633–574–4

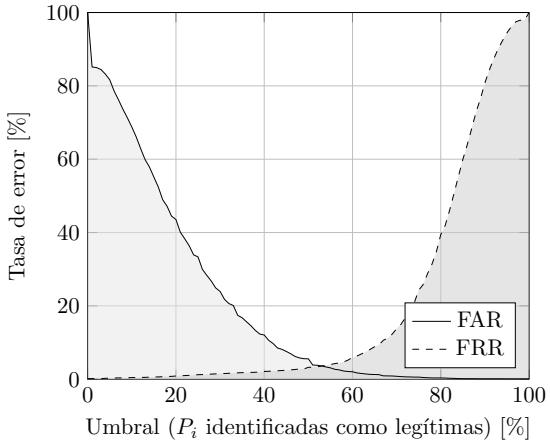


Figura 1: Distribución de falsos positivos y negativos para distintos umbrales

de entrenamiento, pues los bosques aleatorios no lo precisan [23], conservándose los valores originales en milisegundos.

La evaluación de exactitud de los modelos para palabras individuales fue realizada con el método `cross_val_score` de la librería `scikit-learn` [22], utilizando validación cruzada de cinco iteraciones. Las muestras de impostores utilizadas para evaluar la tasa de falsos positivos surgen, para cada usuario, de una selección aleatoria de las muestras de otros usuarios, del mismo tamaño que el conjunto de muestras del usuario legítimo.

5. Resultados y discusión

El proceso descrito en la sección anterior se llevó a cabo para cada muestra en el conjunto de datos de entrada, y se registró el porcentaje de P_i reconocidas como legítimas dentro de la muestra, tanto para los usuarios legítimos como para los impostores. En la figura 1 pueden observarse las tasas de falsos positivos (FAR) y falsos negativos (FRR) resultantes al fijar un umbral de aceptación, que es el porcentaje de votos positivos otorgados por las P_i de cada muestra requeridos para clasificarla como perteneciente al usuario legítimo.

Contrastemos estos resultados con aquellos de [11], citado más arriba, que puede considerarse el pináculo más reciente de la autenticación por medio de cadencias de tecleo en texto libre. Los autores reportan un EER de aproximadamente 5 % utilizando un entrenamiento de sólo 250 caracteres. Aquí alcanzamos un 3,6 % y entrenar cada clasificador por palabra requiere un mínimo de $10m$ caracteres, en donde m es el largo de la misma. Sin embargo, al tratarse de texto libre en donde muchas palabras son poco comunes, necesitamos una gran cantidad de muestras anteriores para conseguir suficientes observaciones de las P_i en consideración. Alcanzar esta tasa de error ha demandado suficientes muestras

ISBN 978–987–633–574–4

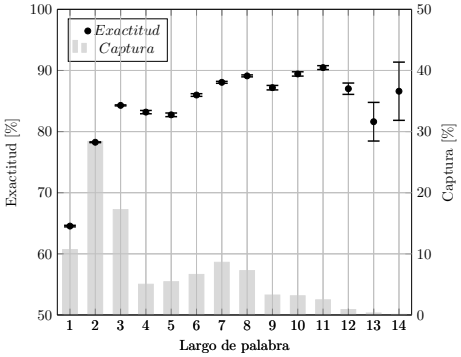


Figura 2: Captura y exactitud por largo de palabra para muestras legítimas

del usuario como para sumar aproximadamente 50.000 caracteres. La enorme mayoría del contenido de las muestras no se utiliza directamente en el entrenamiento de los clasificadores, pero esta ineficiencia es insalvable. No podemos elegir qué ha escrito el usuario y debemos utilizar el texto existente.

No es esperable que el método conserve su precisión al reducir este número. Lamentablemente, la implementación de [11] no se encuentra disponible en forma pública, y el dataset utilizado por los autores [10] no cuenta con suficientes caracteres por usuario para evaluar nuestro método. Para poner en perspectiva el tamaño del entrenamiento requerido, un usuario que hace un uso diario de la computadora intenso y prolongado teclea aproximadamente 15.000 caracteres por día, mientras que para el uso liviano el valor se reduce a aproximadamente un quinto [24]; convertido a días, los 50.000 caracteres requeridos para el entrenamiento oscilan entre cuatro y treinta. Una comparación de ambos se muestra en el cuadro 1. La contribución de palabras de distintos largos a la clasificación de

Característica	**Acien et al.**	**Método propuesto**
Clasificador	ANN siamesa recurrente	Ensamble de palabras + RF
Entrenamiento	250 caracteres	50.000 caracteres
EER	$\approx 5\,\%$	$\approx 3{,}6\,\%$

Cuadro 1: Comparación de principales características

muestras legítimas puede observarse en la figura 2. Se han incluido intervalos de confianza del 95 % para la exactitud promedio de los clasificadores individuales. Nótese que esta mejora con el largo de palabra hasta acercarse al 90 %, consistentemente con lo reportado en [12], pero que este efecto es difícil de aprovechar pues se utilizan con menor frecuencia. El intervalo de confianza crece con el largo de la palabra pues el tamaño de la población disminuye significativamente.

Es interesante notar una dificultad adicional. Al evaluar muestras de impostores, las palabras que pasan los filtros descritos en las secciones anteriores (sin

ISBN 978–987–633–574–4

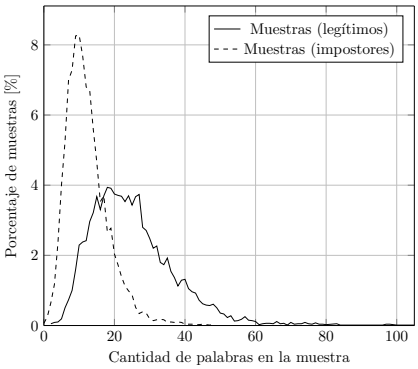


Figura 3: Porcentaje de muestras por cantidad de palabras utilizadas

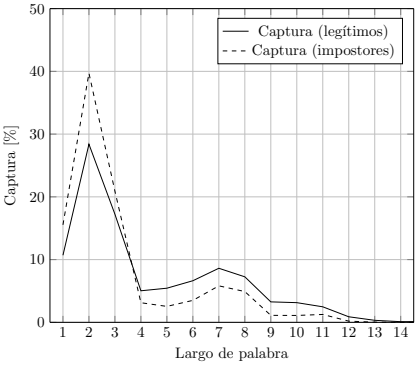


Figura 4: Captura por largo de palabra, para legítimos e impostores

atributos faltantes, ausencia de pausas, y suficientes muestras disponibles) tienden a ser menos y más cortas. Este efecto, que puede observarse en las figuras 3 y 4, es esperable ya que distintos usuarios tienden a utilizar distintas palabras con distintas frecuencias, y sólo aquellas que aparecen en común con suficiente frecuencia pueden ser utilizadas por este método.

5.1. Futuras líneas de investigación

La figura 2 muestra que con el incremento del largo de palabra la exactitud de los clasificadores individuales mejora, confirmando las conclusiones de [12]. En el método propuesto cada palabra brinda un voto al ensamble, pero el fenómeno antedicho apunta a la posibilidad de mejorar el rendimiento utilizando distintos pesos para distintas palabras, en base a su largo, la frecuencia de uso, y la exactitud de su modelo individual. La exploración de esta mejora se relega a futuras líneas de investigación.

6. Conclusión

En el presente estudio se propuso un método de ensamble para la autenticación de usuarios con cadencias de tecleo en textos libres, que utiliza subsecuencias a nivel de palabras. En contraste con otros métodos del estado del arte [11], que alcanzan tasas de error en el orden del 5 % con entrenamientos muy reducidos, nuestro método demanda grandes cantidades de información para el entrenamiento inicial pero alcanza un EER más bajo, del orden del 3,6 %. La combinación de los dos enfoques en un esquema mixto permitiría en principio lograr ambos objetivos, utilizando alguno de los primeros cuando se cuenta con poca información para alcanzar tasas aceptables rápidamente, y delegando al segundo cuando se hayan acumulado suficientes muestras. Los conjuntos de datos de entrenamiento y de resultados fueron puestos a disposición en forma pública y abierta en IEEE DataPort [15] y Mendeley Data [14].

Bibliografía

[1] R Stockton Gaines, William Lisowski, S James Press, and Norman Shapiro. Authentication by keystroke timing: Some preliminary results. Technical report, Rand Corp Santa Monica CA, 1980.

[2] Fabian Monrose and Aviel D Rubin. Keystroke dynamics as a biometric for authentication. *Future Generation computer systems*, 16(4):351–359, 2000.

[3] Patrick Bours and Hafez Barghouthi. Continuous authentication using biometric keystroke dynamics. In *The Norwegian Information Security Conference (NISK)*, volume 2009, 2009.

[4] Antony Milne, Katayoun Farrahi, and Mihalis A Nicolaou. Less is more: Univariate modelling to detect early parkinson's disease from keystroke dynamics. In *International Conference on Discovery Science*, pages 435–446. Springer, 2018.

[5] Clayton Epp, Michael Lippold, and Regan L Mandryk. Identifying emotional states using keystroke dynamics. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 715–724, 2011.

[6] Enrique P Calot, Jorge S Ierache, and Waldo Hasperué. Robustness of keystroke dynamics identification algorithms against brain-wave variations associated with emotional variations. In *Proceedings of SAI Intelligent Systems Conference*, pages 194–211. Springer, 2019.

[7] Kevin S Killourhy and Roy A Maxion. Should security researchers experiment more and draw more inferences? In *CSET*, 2011.

[8] Kevin S Killourhy and Roy A Maxion. Comparing anomaly-detection algorithms for keystroke dynamics. In *2009 IEEE/IFIP International Conference on Dependable Systems & Networks*, pages 125–134. IEEE, 2009.

[9] Daniele Gunetti and Claudia Picardi. Keystroke analysis of free text. *ACM Transactions on Information and System Security (TISSEC)*, 8(3):312–347, 2005.

[10] Vivek Dhakal, Anna Feit, Per Ola Kristensson, and Antti Oulasvirta. Observations on Typing from 136 Million Keystrokes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, 2018. doi:https://doi.org/10.1145/3173574.3174220.

[11] Alejandro Acien, Aythami Morales, Ruben Vera-Rodriguez, Julian Fierrez, and John V Monaco. Typenet: Scaling up keystroke biometrics. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2020.

[12] Nahuel González, Germán Concilio, Enrique P. Calot, Jorge S. Ierache, and Waldo Hasperué. Exploring internal correlations in timing features of keystroke dynamics at word boundaries and their usage for authentication and identification. In *Computer Science–CACIC 2020: 26th Argentine Congress, CACIC 2020, San Justo, Buenos Aires, Argentina, October 5–9, 2020, Revised Selected Papers*, volume 1, page 321. Springer Nature.

[13] Enrique P. Calot. Keystroke dynamics keypress latency dataset. Database, jan 2015. URL `http://lsia.fi.uba.ar/pub/papers/kd-dataset/`.

[14] Nahuel González. Dataset for an ensemble method for keystroke dynamics authentication in free-text using word boundaries, 2021. URL `https://data.mendeley.com/datasets/xvg5j5z29p/1`.

[15] Nahuel González. Dataset for an ensemble method for keystroke dynamics authentication in free-text using word boundaries, 2021. URL `https://ieee-dataport.org/documents/dataset-ensemble-method-keystroke-dynamics-authentication-free-text-using-word-boundaries`.

[16] Francesco Bergadano, Daniele Gunetti, and Claudia Picardi. User authentication through keystroke dynamics. *ACM Transactions on Information and System Security (TISSEC)*, 5(4): 367–397, 2002.

[17] Nahuel González, Enrique P Calot, and Jorge S Ierache. A replication of two free text keystroke dynamics experiments under harsher conditions. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6. IEEE, 2016.

[18] Nahuel González and Enrique P Calot. Finite context modeling of keystroke dynamics in free text. In *Biometrics Special Interest Group (BIOSIG), 2015 International Conference of the*, pages 1–5. IEEE, 2015.

[19] Enzhe Yu y Sungzoon Cho. Ga-svm wrapper approach for feature subset selection in keystroke dynamics identity verification. In *Neural Networks, 2003. Proceedings of the International Joint Conference on*, volume 3, pages 2253–2257. IEEE, 2003.

[20] Balqies Obaidat, Mohammad S y Sadoun. Verification of computer users using keystroke dynamics. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 27(2): 261–269, 1997.

[21] Kevin S Maxion, Roy A y Killourhy. Keystroke biometrics with number-pad input. In *Dependable Systems and Networks (DSN), 2010 IEEE/IFIP International Conference on*, pages 201–210. IEEE, 2010.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[23] Yanjun Qi. Random forest for bioinformatics. In *Ensemble machine learning*, pages 307–323. Springer, 2012.

[24] Average keyboard use statistics per user. URL `https://whatpulse.org/stats/overall/numbers/#averages-per-user`.

ISBN 978–987–633–574–4

Detección de Patrones de Comportamiento en la Red a través del Análisis de Secuencias

Carlos Catania, Jorge Guerra, Juan Manuel Romero, Franco Palau, Gabriel Caffaratti, and Martín Marchetta

Universidad Nacional de Cuyo
Facultad de Ingeniería
Laboratorio de Sistemas Inteligentes (LABSIN)
Mendoza, Argentina
{harpo, jorge.guerra}@ingenieria.uncuyo.edu.ar
juaanromeero66@gmail.com
{franco.palau, gabriel.caffaratti,
martin.marchetta}@ingenieria.uncuyo.edu.ar

Resumen Los enfoques de detección por comportamiento en el tráfico de red se basan en encontrar patrones comunes que sigue un ataque a lo largo de su ciclo de vida, tratando de generalizarlos para poder detectar una traza de ataque no vista con anterioridad. Un enfoque común consiste en la generación de secuencias basadas en caracteres para representar comportamientos maliciosos, y luego aplicar modelos como Cadenas de Markov para generalizar a otros comportamientos similares. Sin embargo, estos últimos presentan limitaciones para explorar más allá del estado anterior. En el presente trabajo se analizan las ventajas y limitaciones de tres arquitecturas de redes neuronales para detectar comportamientos maliciosos capaces de recordar patrones vistos mucho tiempo atrás. Para esto se realizó una evaluación sobre un conjunto de datos específicamente diseñado que incluye comportamientos maliciosos y normales de diversas fuentes. Los resultados preliminares indican que, a pesar de su simplicidad, la aplicación de cualquiera de las arquitecturas de red es un enfoque válido para detectar comportamientos de red maliciosos, lo cual es prometedor para su aplicación a problemas de etiquetado de tráfico de red en el contexto de un flujo de trabajo con interacción humana.

Keywords: Botnet · Redes Neuronales · Seguridad Informática

1. Motivación

El presente trabajo se basa en la aplicación de modelos de comportamiento de trazas de red a partir del estudio de las características a largo plazo presentes en el tráfico. Los enfoques de detección basados en modelos de comportamiento han demostrado ser adecuados para hacer frente a los constantes cambios de actividades que presentan los ataques [2]. Un enfoque de detección por comportamiento se basa en encontrar aquellos patrones comunes que sigue un ataque a lo largo de su ciclo de vida, tratando de generalizarlos para ser capaz de detectar una traza de ataque no vista con anterioridad. Por ejemplo, en el caso de

las llamadas Botnets, periódicamente todos los bots necesitarán conectarse al Botmaster para recibir nuevas instrucciones. Esto constituye un claro patrón de comportamiento, que sólo es observado después de un largo período de tiempo.

En el marco del proyecto Stratosphere [5], se han implementado modelos de comportamiento en el sistema gratuito SLIPS (Stratosphere Linux Intrusion Prevention System [5]). SLIPS utiliza secuencias basadas en caracteres para representar comportamientos maliciosos, y luego aplica modelos como Cadenas de Markov (MCM, Markov Chain Models) para generalizar a otros comportamientos de este tipo. Los modelos basados en MCM se eligen por ser eficientes computacionalmente, pero no generalizan bien frente a ataques con nuevos comportamientos. Además los MCM tienen la limitación de que el cálculo de las probabilidades de transición depende únicamente del estado anterior.

Por otro lado, otras técnicas como ser las Redes LSTM (Long-Short Term Memory) [4], las Redes 1DCNN (1D Convolutional Neural Networks) [1], y más recientemente los modelos ATTE (Attention Models) [10], han demostrado ser eficientes para extraer patrones en secuencias y utilizarlos para clasificación [3]. Las arquitecturas LSTM y ATTE han sido exitosas en el tratamiento de patrones de dependencia a largo plazo, mientras que las 1DCNN han mostrado algunas limitaciones en estos casos. Sin embargo, la ventaja de las 1DCNN es que pueden entrenarse hasta 9 veces más rápido que las otras dos arquitecturas.

Al igual que [9], este trabajo se centra en analizar el problema de la detección del comportamiento Botnet utilizando las tres arquitecturas de redes neuronales mencionadas, con una complejidad mínima (con un número mínimo de capas). La hipótesis es que, a pesar de la simplicidad de las arquitecturas utilizadas, estas pueden aprender las propiedades comunes de los diferentes comportamientos de los ataques y el rendimiento resultante podría constituir una base para medir el de otras arquitecturas más complejas y costosas computacionalmente.

La principal contribución del trabajo es un análisis de las ventajas y limitaciones de tres arquitecturas de redes neuronales para detectar comportamientos maliciosos en la red. Para esto se realizó una evaluación sobre un conjunto de datos específicamente diseñado que incluye comportamientos maliciosos de 14 ataques reales diferentes y 6 comportamientos normales de diversas fuentes.

El trabajo se organiza como sigue. La Sección 2 presenta un modelo de representación de comportamiento basado en secuencias de caracteres. La Sección 3 describe las diferentes arquitecturas de las redes. La Sección 4 ofrece los detalles del diseño de los experimentos y sus resultados. La Sección 5 presenta las tareas de optimización de hiper-parámetros de los modelos. La Sección 6 muestra el desempeño de los modelos en ejemplos nuevos. Finalmente la Sección 7 expone las conclusiones del trabajo.

2. Representación del Comportamiento Mediante Secuencia de Caracteres

El enfoque de SLIPS modela el comportamiento de una traza de tráfico de red agregando los flujos según una 4-tupla compuesta por: la dirección IP de

origen, la dirección IP de destino, el puerto de destino y el protocolo. Todos los flujos de red que coinciden con una tupla se juntan y se los denomina *Conexión Stratosphere* (SC). A partir de una captura de tráfico se crean varias SC, cada una conteniendo un grupo de flujos. En base a esta representación la secuencia de comportamiento de una SC se calcula como sigue:

1. Se extraen tres características de cada flujo: tamaño, duración y periodicidad.
2. Se asigna a cada flujo un símbolo de *estado* según las características extraídas y la estrategia de asignación mostrada en la Tabla 1.
3. Después de la asignación, cada *conexión* tiene su propia cadena de símbolos que representa su comportamiento en la red.

Tabla 1. Estrategia de asignación de caracteres para la representación del comportamiento

	Tamaño			**Chico**			**Medio**			**Grande**	
	Dur. Corta	**Dur. Med**	**Dur. Larga**	**Dur. Corta**	**Dur. Med**	**Dur. Larga**	**Dur. Corta**	**Dur. Med**	**Dur. Larga**		
Per. Fuerte	a	b	c	d	e	f	g	h	i		
Per. Débil	A	B	C	D	E	F	G	H	I		
No-Per. Fuerte	r	s	t	u	v	w	x	y	z		
No-Per. Débil	R	S	T	U	V	W	X	Y	Z		
Sin Datos	1	2	3	4	5	6	7	8	9		

	Diferencia de tiempo				
	0s a 5s	**5s a 1m**	**1m a 5m**	**5m a 1h**	**> 1h**
Símbolo	.	,	+	*	0

Un ejemplo de un *modelo de comportamiento basado en caracteres* se observa en la Figura 1, donde se muestran los símbolos que representan todo el flujo para una SC basada en el protocolo UDP.

2.4.R*R.R.R*a*b*a*a*b*b*a*R.R*R.R*a*a*b*a*a*a*a*

Figura 1. Un ejemplo de un SC desde la dirección 10.0.2.103, y destino a la dirección 8.8.8.8 utilizando el puerto 53 del protocolo UDP.

3. Arquitecturas de la Red Neural

Se consideraron tres arquitecturas de redes neuronales: (a) una basada en una red convolucional 1D (1DCNN) [1], (b) una red neuronal recurrente basada en LSTM [9], (c) una arquitectura que incluye un mecanismo de Atención [10]. El presente trabajo se enfoca en evaluar el rendimiento de las redes neuronales consideradas con un mínimo de capas adicionales además de las capas Convolucionales, LSTM y Atención. Una descripción visual de cada arquitectura se muestra en la Figura 2. A continuación se describen las capas utilizadas.

3.1. Capa Embedding

Esta capa permite proyectar secuencias de caracteres de entrada de longitud l en una secuencia de vectores R^{lxd}, donde l debe determinarse a partir de la

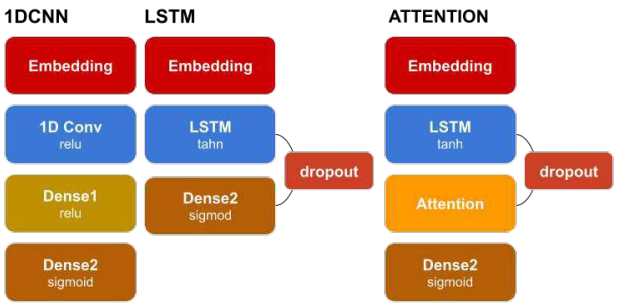


Figura 2. Esquema simplificado de las arquitecturas de las tres redes aplicadas al problema de clasificación de secuencias.

información proporcionada por las secuencias del conjunto de entrenamiento, y d es un parámetro libre que indica la dimensión de la matriz resultante [9]. Mediante esta capa la red neuronal aprende de manera eficiente el conjunto óptimo de características que representan los datos de entrada.

3.2. Capa 1D Convolucional

Para el problema de detección de comportamiento se utiliza una capa convolucional de una dimensión, compuesta por un conjunto de filtros convolucionales que se aplican a diferentes porciones de la secuencia.

Al aplicar el mismo filtro en toda la secuencia se reduce el tiempo de cálculo en comparación con las arquitecturas tradicionales como el Perceptrón Multicapa (MLP). Además, dado que un núcleo convolucional opera de forma independiente sobre cada 4-grama, es posible recorrer toda la capa de entrada de forma simultánea. Estas características aportan ventajas sobre otros enfoques de aprendizaje profundo que se suelen utilizar para el procesamiento de textos, como ser la LSTM [4, 8, 9].

3.3. Capa LSTM (Long Short-Term Memory)

La capa LSTM [4] ha sido aplicada con éxito a problemas de Procesamiento de Lenguaje Natural (NLP). La principal ventaja de las redes que usan una capa LSTM es que son capaces de memorizar información vista previamente en una secuencia. Esta capacidad se traduce en el aprendizaje de patrones de caracteres que no son necesariamente contiguos en la secuencia, sino que pueden haberse observado temporalmente muy distanciados.

3.4. Capa de Atención

A pesar de su capacidad para recordar patrones vistos muy atrás en el tiempo, las redes LSTM tienen problemas para detectar patrones cuando la secuencia supera un determinado tamaño. Para atacar este problema surge la idea de

construir un vector de contexto que preste especial “atención” a determinados estados ocultos de la capa LSTM (en lugar de sólo el último estado) [10]. Luego el modelo mismo aprende a cuáles elementos de la secuencia prestar atención y a cuáles es mejor ignorar. Los modelos de atención han probado ser una mejora significativa respecto a las redes LSTM.

3.5. Capa Densa

Esta es una capa clásica de una red de tipo MLP totalmente conectada. Esta capa se aplica sobre las características extraídas por la arquitectura 1DCNN, y como parte de las tres arquitecturas consideradas para obtener la probabilidad de que una secuencia pertenezca a la clase *Botnet* o *Normal*.

4. Diseño Experimental

4.1. Métricas

Para evaluar las arquitecturas propuestas, se utilizan las siguientes métricas estándares: **Sensibilidad**, **Especificidad** y **Exactitud Balanceada**. La Sensibilidad se calcula como el cociente entre los elementos correctamente identificados como positivos sobre el total de verdaderos positivos. La Especificidad se calcula como el cociente entre los elementos de la clase negativa identificados como negativos sobre el número total de ejemplos de la clase negativa. La Exactitud Balanceada es la media entre la Sensibilidad de ambas clases, donde la clase positiva contiene las SC que posee comportamiento malicioso.

4.2. Descripción del conjunto de datos

Las tres arquitecturas de redes neuronales fueron entrenadas y evaluadas con un conjunto de datos de seguridad de redes específicamente diseñado para el problema de la detección de botnets, conformado por veinte capturas de red publicadas por el grupo de investigación Stratosphere IPS de la Czech Technical University (CTU) [6]. El conjunto de datos referido como **CTU [A]** contiene cinco capturas de botnet y cuatro capturas normales, las cuales incluyen tráfico como DNS, HTTPS y P2P. En total, todas las capturas representan 20747 SC, teniendo 17826 etiquetadas como “Botnet” y 1449 etiquetadas como “Normal”. Todas estas capturas fueron recogidas entre 2013 y 2017.

Además, se utilizó un segundo conjunto de datos, denominado **CTU [B]** para realizar una evaluación final e independiente de los modelos resultantes. El CTU [B] se compone de cinco redes de bots y dos capturas normales. En total, las siete capturas representan 1574 SC con 1437 etiquetadas como “Botnet” y 137 etiquetadas como “Normal”.

La Tabla 2 proporciona una breve descripción de cada captura para ambos conjuntos de datos. La primera columna determina el conjunto de datos. Luego, las dos siguientes columnas muestran el ID de la captura según la denominación

utilizada por en el proyecto Stratoshpere y el tipo de tráfico incluido en la captura (es decir, Botnet o Normal). A continuación, en las columnas cuatro y cinco, el nombre del malware y el número de SC incluidas en la captura. En muchos casos el nombre del malware no se encontraba disponible, por lo tanto, se asume que todas las capturas marcadas como desconocidas no difieren significativamente en el tipo de ataque utilizado.

Tabla 2. Capturas de tráfico en los conjuntos CTU [A] y CTU [B]

Conj.	MCFP Captura ID	Clase	Malware	N. Con.
CTU [A]	2013-10-01_capture-win8	botnet	desconocido	3463
	2013-10-01_capture-win12	botnet	desconocido	2197
	2013-08-20_capture-win15	botnet	Kelihos	9844
	2013-11-06_capture-win6	botnet	Zbot	2088
	2014-03-12_capture-win3	botnet	Zbot	234
	2017-04-30_win-normal	normal	-	348
	2017-05-02_kali-normal	normal	-	562
	2017-04-18_win-normal	normal	-	387
	2017-04-19_win-normal	normal	-	152
CTU [B]	2013-08-20_capture-win2	botnet	Zbot	28
	2014-01-25_capture_win3	botnet	Zbot	52
	2014-02-10_capture-win3	botnet	Zbot	47
	2013-11-25_capture-win7-2	botnet	desconocido	627
	2014-01-31_capture-win7	botnet	desconocido	581
	2014-01-14_capture-win2	botnet	Hicrazyk.A	102
	2013-12-17_capture1	normal	-	32
	2017-04-25_win-normal	normal	-	105

5. Ajuste de Hiper-parámetros

Para ajustar los parámetros de las arquitecturas presentadas en la Sección 3, se utilizó un barrido matricial. La gama completa de valores de los hiper-parámetros se presenta en la Tabla 3. Las dos primeras columnas se refieren a los parámetros y a los valores considerados durante la búsqueda de parámetros, mientras que las columnas restantes presentan los valores óptimos obtenidos.

Tabla 3. Rango de los hiper-parámetros y los valores óptimos encontrados para las tres arquitecturas. Los dos modelos con menor variabilidad fueron seleccionados.

Hiper-parámetros	Valores	1DCNN		LSTM		ATTE	
		(1)	(2)	(1)	(2)	(1)	(2)
Tamaño LSTM	128,64,32	-	-	64	64	128	64
Filtros 1DCNN	256,128,64	256	128	-	-	-	-
Kernel 1DCNN	8,4,2	4	4	-	-	-	-
Tamaño Embedding	128,64,32	128	128	32	64	128	32
Tamaño Denso	256,128,64	128	256	-	-	-	-
Valor Dropout	0.5,0.1	-	-	0.1	0.1	0.5	0.1
Tamaño Batch	1024,256	1024	1024	1024	256	1024	256

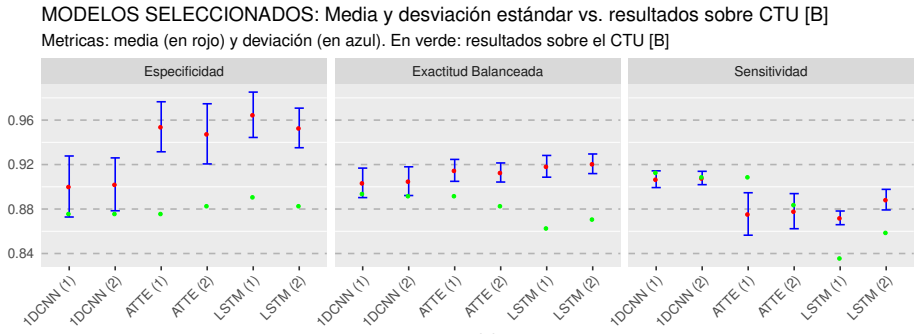


Figura 3. Media y desviación estándar para cada uno de los 6 modelos que obtuvieron mejores resultados.

A fin de evaluar cada conjunto de hiper-parámetros en un plazo de tiempo aceptable se analizó el porcentaje mínimo de muestras necesarias para el entrenamiento/prueba sin que afecte significativamente al rendimiento del modelo. Se muestrearon 30 pares de conjuntos de datos sobre un subconjunto del 25 % de CTU [A], donde el 20 % de las muestras se seleccionaron para entrenamiento y el 5 % restante se utilizó para la prueba.

No se observaron diferencias significativas en las métricas de rendimiento durante la búsqueda matricial. Por esto, se seleccionaron los modelos con un valor medio de Exactitud Balanceada superior a 0,89, y la menor varianza.

La Figura 3 complementa los resultados de la Tabla 3 con los valores de la media y la desviación estándar para las tres métricas consideradas. En general, todas las arquitecturas mostraron una baja variabilidad, en particular para las métricas de Exactitud Balanceada y Sensibilidad (número de botnets correctamente detectados). Sin embargo, en la Especificidad (número de conexiones normales correctamente detectadas), se observa una variabilidad considerablemente mayor. Esto último podría explicarse por el bajo número de ejemplos normales presentes en el conjunto de datos y su alta variabilidad.

Por otro lado, las arquitecturas basadas en 1DCNN mostraron los mejores resultados en cuanto a la detección correcta de los comportamientos de las botnets (Sensibilidad) mientras que ATTE y LSTM observaron un mejor rendimiento cuando se trataba de comportamientos normales (Especificidad). Sin embargo, al considerar la métrica de Exactitud Balanceada, las tres arquitecturas de red no parecieron mostrar diferencias significativas.

6. Evaluación sobre nuevos ejemplos

Se realizó una evaluación independiente de los modelos con los hiper-parámetros seleccionados en la Sección 5. Esta vez, las tres arquitecturas de red se entrenaron con CTU [A] completo y se evaluaron contra CTU [B] completo. Los resultados se muestran en la Tabla 4.

En general, todas las arquitecturas de red mostraron valores aceptables en todas las métricas consideradas (es decir, presentaron valores superiores al 80 %). Al igual que los resultados observados en la Sección 5, la 1DCNN mostró el mejor valor de Sensibilidad mientras que las redes LSTM funcionan mejor en términos de Especificidad. Sin embargo, cuando se consideran los valores para la Exactitud Balanceada, la red 1DCNN supera claramente a las redes LSTM. Por otro lado, las redes con una capa de atención mostraron un rendimiento similar al de 1DCNN en todas las métricas consideradas.

A pesar de los resultados similares entre las tres arquitecturas de redes neuronales (ver Tabla 4), surgen algunas diferencias cuando analizamos capturas específicas de comportamientos normales y de botnets. En la Figura 4, se observan los valores de Sensibilidad para cada una de las capturas utilizando los mejores modelos en cada arquitectura considerada. Con la excepción de la captura de `2014-01-14-win2`, todas las capturas de botnets fueron detectadas con una Sensibilidad superior al 80 % por cada una de las arquitecturas de red.

Sin embargo, las redes 1DCNN y ATTE ofrecen un mejor rendimiento en comparación con LSTM para todas las capturas de botnets. Como se muestra en el boxplot de la parte inferior de la figura, 1DCNN y ATTE mostraron valores de Sensibilidad superiores al 90 % para las capturas de botnets. En particular, la captura de botnet `2013-11-25-capture-win7-2` con 625 SC es la responsable del bajo rendimiento de Sensibilidad de LSTM. Según se puede observar, varias SC no son detectadas por las arquitecturas LSTM (79 % de Sensibilidad), mientras que 1DCNN y ATTE muestran un rendimiento de detección considerablemente mejor de 91 % y 89 %, respectivamente.

En cuanto a las capturas normales, no se observan diferencias significativas entre todas las arquitecturas de red. En particular, todos los modelos mostraron un bajo rendimiento de detección para la captura `2013-12-17-capture1`. Sin embargo, casi todas las SC de la captura normal `2017-04-25-win-normal` (105 SC en total) son detectadas correctamente por LSTM (es decir, 99 % de Sensibilidad), mientras que para los modelos 1DCNN y ATTE la Sensibilidad disminuye al 97 %. Esto último puede explicar el mejor rendimiento mostrado por LSTM en la métrica de Especificidad de la Tabla 4.

Por otro lado, la captura de botnet `2014-01-14-capture-win2` es la única que podemos corroborar que fue diferente a los comportamientos de botnet

Tabla 4. Modelos seleccionados entrenados con la totalidad del CTU19[A] y evaluados sobre CTU19[B].

Arquitectura de la red	Especificidad	Sensibilidad	Exactitud Balanceada
ATTE (1)	0.876	0.909	0.892
ATTE (2)	0.883	0.884	0.883
LSTM (1)	0.891	0.836	0.863
LSTM (2)	0.883	0.859	0.871
1DCNN (1)	0.876	0.913	0.894
1DCNN (2)	0.876	0.909	0.892

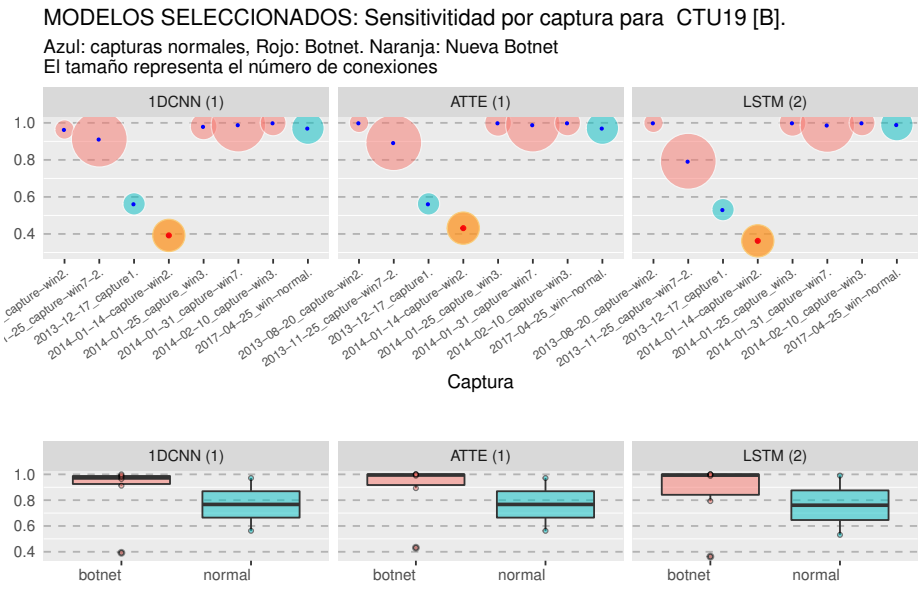


Figura 4. Arriba: Valores de Sensibilidad discriminados por captura de trafico. **Abajo:** Sensibilidad discriminada por clase.

utilizados durante el entrenamiento. En cuanto a la Sensibilidad, todas las arquitecturas de red detectaron menos del 50 % de las SC presentes. Para esta captura en particular, el mejor rendimiento de detección lo mostraron los modelos ATTE con un 43 % mientras que 1DCNN y LSTM mostraron valores de sensibilidad por debajo del 40 %. Frente a esta disminución, parece claro que el comportamiento de esta nueva Botnet difiere del resto de las Botnets. Sin embargo resulta interesante observar que todos los modelos lograron cierto nivel de generalización.

7. Conclusiones y trabajos futuros

Se han evaluado tres arquitecturas de redes neuronales sencillas y bien establecidas para la clasificación de secuencias en un conjunto de datos compuesto por 20 capturas de tráfico de diferentes comportamientos maliciosos y normales. En general parece que, a pesar de su simplicidad, la aplicación de cualquiera de las arquitecturas de red es un enfoque válido para detectar comportamientos de red malicfiosos. Todos los modelos han mostrado una precisión equilibrada por encima del 80 %. A pesar de una evidente reducción del rendimiento, todos los modelos también han mostrado un rendimiento aceptable cuando se introdujeron nuevos comportamientos de botnet en la evaluación independiente (el 40 % de las SC se detectaron correctamente).

Sorprendentemente la arquitectura 1DCNN, a pesar de no ser capaz de capturar dependencias a largo plazo, ha mostrado resultados aceptables a lo largo

de todo el proceso de evaluación (mostró un valor de Exactitud balanceada entre el 88 % y el 92 %). Este buen rendimiento en términos de detección de ataques y de trazas normales hacen de las redes basadas en 1DCNN una buena opción para detectar los comportamientos de la red. Esto cobra más importancia si se tiene en cuenta que el re-entrenamiento periódico requerido por el desvío de la distribución de datos se ha convertido en un proceso común en aplicaciones de aprendizaje automático, y que 1DCNN requiere un tiempo considerablemente menor durante el proceso de entrenamiento.

A futuro se planea la evaluación de estas arquitecturas y otras más complejas a problemas de etiquetado de trafico de red en el contexto de un flujo de trabajo con interacción humana [7].

8. Agradecimientos

Los autores agradecen el apoyo recibido por la SIIP-UNCuyo y la ANPCyT (proyectos 06/B363, 06/B374 y PICT 1435), y el de NVIDIA Corporation (donación de una GPU Titan V).

Referencias

1. Catania, C., Garcia, S., Torres, P.: An analysis of convolutional neural networks for detecting DGA. In: Proceedings of XXIV Congreso Argentino de Ciencias de la Computación (2018)
2. Garcia, S.: Identifying, Modeling and Detecting Botnet Behaviors in the Network. Ph.D. thesis, UNICEN University (2014). https://doi.org/10.13140/2.1.3488.8006
3. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. Adaptive computation and machine learning, The MIT Press, Cambridge, Massachusetts (2016)
4. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Computation **9**(8), 1735–1780 (11 1997)
5. Stratosphere Research Laboratory: Stratosphere IPS for Linux (2019), https://www.stratosphereips.org/stratosphere-ips-for-linux, [Online; accessed May-2021]
6. Stratosphere Research Laboratory: Malware Capture Facility Project (2020), https://www.stratosphereips.org/datasets-malware, [Online; accessed May-2021]
7. Torres, J.L.G., Catania, C.A., Veas, E.: Active learning approach to label network traffic datasets. Journal of Information Security and Applications **49**, 102388 (Dec 2019)
8. Torres, P., Catania, C., Garcia, S., Garino, C.G.: An analysis of Recurrent Neural Networks for Botnet detection behavior. In: 2016 IEEE Biennial Congress of Argentina (ARGENCON). pp. 1–6. IEEE, Buenos Aires, Argentina (Jun 2016)
9. Woodbridge, J., Anderson, H., Ahuja, A., Grant, D.: Predicting domain generation algorithms with long short-term memory networks. ArXiv (2016), http://arxiv.org/abs/1611.00791
10. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical Attention Networks for Document Classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 1480–1489. Association for Computational Linguistics, San Diego, California (2016)

ISBN 978–987–633–574–4

Monitoreo de Llamadas al Sistema como Método de Prevención de Malware

Fabián A. Gibellini, Sergio Quinteros, Germán N. Parisi, Milagros N. Zea Cárdenas, Leonardo Ciceri, Federico J. Bertola, Ileana M. Barrionuevo, Juliana Notreni, Analía L. Ruhl

Laboratorio de Sistemas / Dpto. de Ingeniería en Sistemas de Información/
Universidad Tecnológica Nacional / Facultad Regional Córdoba
Maestro M. Lopez esq. Cruz Roja Argentina S/N, Ciudad Universitaria (X5016ZAA) - Córdoba, Argentina
{fabiangibellini, ser.quinteros, germannparisi, milyzc, leonardorciceri, federicorbertola ilebarrionuevo, julinotreni, analialorenaruhl }@gmail.com

Resumen. De acuerdo con la categoría de un malware, se puede inferir que existen patrones de llamadas al sistema (syscalls) que permitirían descubrir qué tipo de malware se está ejecutando sobre un Sistema Operativo GNU/Linux y de esa manera reaccionar ante un ataque de estas características. Para esto es necesario monitorizar las llamadas al sistema en dicho sistema operativo. La herramienta que se presenta es un monitor de llamadas al sistema en tiempo real. Esta herramienta es parte de un proyecto homologado, cuyo objetivo es detectar malware basándose en patrones de llamadas al sistema en GNU/Linux.

Palabras Claves: Seguridad. Syscalls. Kernel. Linux. Detección.

1 Introducción

Existen diferentes tipos de malwares, además de los clásicos virus, podemos encontrar ransomwares, spywares, rootkits, troyanos y los más recientes, malwares residentes en memoria.

Según Raymond et al, el mayor desafío de crear un esquema completo de nombrado de malwares, se debe al número de muestras existentes de malware y a la frecuencia con la que nuevas muestras son descubiertas [1]. Si se considera la clasificación basada en comportamiento, propuesta por C. Elisan [2], se pueden distinguir a ransomwares, keyloggers, spywares, gusanos, troyanos, etc.

Cada uno de estos malwares intentan generar algún daño y para lograrlo es normal que utilicen al kernel para acceder a los recursos que necesitan. Un ransomware, por ejemplo, es una forma de software malicioso utilizado en ataques, en los que no se busca destruir irreversiblemente los datos, sino cifrar y cobrar por el servicio de recuperación de los datos cifrados [3] y para esto realiza operaciones de lectura y escritura sobre el disco, utilizando el kernel. Otro ejemplo es un keylogger, que es un software que se ubica entre el hardware y el sistema operativo e intercepta cada

pulsación de tecla y la almacena, para lo cual también ejecuta estas operaciones por medio del kernel.

Otro tipo de malware son los spywares, que se cargan de manera clandestina en una PC sin que su propietario se entere, y corre en segundo plano para ejecutar acciones a espaldas del propietario. Una de las formas en la que una máquina se infecta de spyware es por medio de troyanos. Existe una cantidad considerable de software gratuito que contiene spyware y el autor del software puede hacer dinero con este spyware [4].

Existe un tipo de malware, cuya variante se conoció en los últimos años, se trata de los malwares residentes en memoria o sin archivo (fileless). Son infecciones que no implican que los archivos maliciosos se descarguen o se escriban en el disco del sistema [5]. Generalmente están destinados a robar información y el atacante utiliza software existente, aplicaciones permitidas y protocolos autorizados en la víctima como portadores de actividades maliciosas. Este tipo de malware no puede ser identificado por los antivirus [6] [7].

Dentro de los casos reales de ataques de malware rememoramos algunos, como un estudio de Deloitte, el cual reporta que más de 360.000 computadoras, en más de 180 países, fueron afectadas por un ransomware conocido como WannaCry en mayo del 2017. Este ransomware generó costos económicos de 200 millones de dólares [8].

Como caso real de ataques sin archivos es la de un grupo de ciberdelincuentes, llamado Lurk, el cual fue uno de los primeros en emplear efectivamente técnicas de infección sin archivos en ataques a gran escala, técnicas que posiblemente se convirtieron en elementos básicos para otros malhechores. Se creía que Lurk había extraído más de $ 45 millones de dólares de organizaciones financieras, lo que finalmente afectó las operaciones, la reputación y los resultados de las víctimas [9].

Otro ejemplo de malware residente en memoria es Gold Dragon, que entre sus objetivos estaban los Juegos Olímpicos de Invierno en Corea del Sur. Dicho malware consistió de dos funciones primarias [10].

Una consultora de cibereconomía y ciberseguridad estima que los daños del cibercrimen tendrán un costo anual y global de seis millones de millones de dólares en 2021 [11]. Estos costos incluyen daños y destrucción de datos, dinero robado, pérdida de productividad, robo de propiedad intelectual, robo de datos personales y financieros, malversación, fraude, interrupción posterior al ataque en el curso normal de los negocios, investigación forense, restauración y eliminación de datos perjudicados y sistemas, y daño a la reputación [12].

Es debido al impacto asociado a estos malwares que conlleva al hecho de una búsqueda permanente para encontrar nuevas técnicas de prevención, o actualizar continuamente las ya existentes, de forma tal que minimice el impacto de estas amenazas. La más conocida los antivirus o antimalware, un antivirus compara los datos contra una base de datos de software malicioso (firma), si los datos mapean hacia alguna firma, entonces el antivirus muestra que el archivo está infectado [13].

Para los casos en que el malware supere las líneas de defensa planteadas y logre ejecutarse, es que se invierten costos y tiempos en investigar métodos reactivos de detección, de forma tal que una vez detectados puedan ser detenidos. Con relación a esto, predominan los trabajos dirigidos a la detección de un malware en particular, algunos de los mismos se mencionan a continuación.

Luckett et al, presenta una investigación para la detección de rootkits, donde examina una variedad de algoritmos de aprendizaje automático (como el de vecinos más cercanos, árboles de decisión, redes neuronales y máquinas de vectores de soporte) y propone un método de detección de comportamiento con un bajo consumo de energía de la CPU. Evaluando este método en los sistemas operativos Windows 10, Ubuntu Desktop y Ubuntu Server, junto con cuatro rootkits diferentes e identificando los algoritmos de mejor desempeño [14]. Muchos de los esfuerzos actuales para detectar los rootkits se basan en fuentes conocidas y son principalmente específicos de cada sistema operativo, por lo tanto, son ineficaces para detectar rootkits recién mutados, ocultos y desconocidos. Partiendo de esto, Ramani et al proponen un sistema para la detección de rootkits mediante la identificación de archivos ocultos. Este proceso de detección define un marco de monitoreo de procesos que mantiene continuamente una lista de archivos activos y puede detectar rootkits conocidos y desconocidos con una sobrecarga de rendimiento mínima.

También distinguimos un estudio que se centra en Cloud Computing, ésta es una instalación compartida y distribuida que son accedidas de forma remota por cualquier usuario final. Debido a esto, es que estos ambientes son vulnerables a varios ataques y requieren atención inmediata. Gupta y Kumar se concentran en ataques como rootkits, gusanos y troyanos en sistemas en un entorno Cloud Computing. Ellos describen críticamente y analizan técnicas para la detección de llamadas de sistema maliciosas y proponen una nueva técnica basada en la estructura de firmas de llamadas al sistema inmediatas, para determinar las ejecuciones de programas maliciosos en la nube [15].

Respecto a la detección de ransomware, malware que ha tenido un gran impacto en los últimos años, Popli y Girdhar realizaron un estudio al comportamiento de los ransomwares WannaCry y Petya, incluyendo análisis de procesos, análisis del sistema de archivos, análisis de persistencia y análisis de red, además de simulaciones en la herramienta Cuckoo con el objetivo de identificar técnicas que permitan distinguir cuándo un ransomware se convierte en un malware polimórfico y metamórfico [16].

Otros enfoques utilizan machine learning para identificar ransomwares y malwares en general, ya que implica un aprendizaje de los patrones en los datos para crear un modelo. Alrawashdeh y Purdy proponen un sistema una precisión limitada para la detección de ransomware dinámicamente utilizando machine learning [17]. Honeypots, es otra de las técnicas utilizadas, que a través de una configuración de archivos de señuelo permite “engañar” a un ransomware. Una vez que estos archivos son accedidos, el ransomware puede ser identificado. Incluso se ha aplicado también el análisis estadístico de llamadas a APIs entre una operación normal y un ransomware para generar modelos de detección de estos [18].

La siguiente y última técnica presentada se basa en identificar patrones en las llamadas al sistema (system calls) para así inferir si un proceso, en un sistema operativo GNU/Linux, puede ser considerado como un software potencialmente maligno. El proyecto de I+D (código SIUTNCO0007850) dentro el cual está inserto el presente trabajo propone ampliar esta técnica, en conjunto con identificación de patrones de datos para la detección de distintos malwares, como fileless, ransomware, entre otros de forma tal de lograr un monitoreo y detección de malwares en tiempo real y minimizar los daños económicos o de cualquier otra índole.

Una llamada al sistema o system call es un método o función que puede invocar un proceso para solicitar un cierto servicio al kernel o núcleo del sistema operativo [19]. Es válido aclarar que la necesidad de acceder a los servicios que brinda kernel a través de llamadas al sistema no es algo exclusivo de los malwares, sino que cualquier proceso lo realiza. La diferencia se encuentra en la manera de acceder, tanto hacia qué llamada, como la frecuencia y los parámetros con las que cada una es solicitada.

La detección de malwares en general y herramientas que permitan detectar cualquier tipo de software malicioso es un campo que sigue en expansión, dado a que el mapa de malwares tiende a seguir creciendo día a día. Actualmente se aplican diversas técnicas o combinaciones de una o más, como por ejemplo el análisis de comportamiento de procesos, el machine learning, las redes neuronales, el data mining y la clasificación de datos basada en comprensión [20]. De los cuales mencionamos algunos trabajos destacados, que junto con los descriptos anteriormente son considerados puntos de partida para el presente proyecto, y demostrando la importancia de generar nuevas herramientas que puedan detectar malware o inferir si un software (en ejecución) es malicioso antes que genere un impacto mayor del que ya ha generado.

Entre los métodos de detección de malware de análisis de comportamiento a través de llamadas al sistema, destacamos a Canzanese et al, que sostienen que un conjunto relativamente pequeño de tipos de llamadas al sistema provee una precisión comparable a la de modelos más complejos al momento de detectar procesos maliciosos. Además, afirma que su gran contribución son técnicas de extracción de características de procesos malicioso, con una tasa de falsos positivos muy baja [21]. En un artículo posterior, amplía este trabajo y describe un sistema que usa características extraídas de un seguimiento a las llamadas al sistema, lo que proporciona una lista de las llamadas que se ejecutan y el orden en el que lo hacen.

Combinando la minería de procesos y llamadas al sistema se encuentra Acampora et al, con un modelo para la detección de malware obtenido mediante un enfoque declarativo de minería de procesos a partir del análisis de algunos malwares en ejecución. La idea principal es que el conjunto de relaciones y patrones de ejecución recurrentes entre las llamadas al sistema de un malware en ejecución se pueden modelar para obtener una huella digital del mismo. Estas huellas se comparan y clasifican mediante un algoritmo de agrupación difusa para recuperar el mapa de relaciones de malware de todos los tipos de de malware considerados. La evaluación de este enfoque se realizó sobre un conjunto de datos de más de 4,000 software infectados en 39 tipos de malware [22].

Machine learning es aplicado en muchos trabajos, dentro de los cuales resaltamos los de Asmithad con Vinod y Saxe junto a Berlin. Los primeros proponen un sistema que utiliza machine learning para identificar procesos maliciosos en Linux. Extraen llamadas al sistema dinámicamente utilizando la herramienta Strace e identifican el mejor conjunto de características de procesos malignos y benignos para construir un modelo de clasificación de malware eficiente [23]. Saxe y Berlin incluyen, además de machine learning, redes neuronales en su sistema clasificador de malware de redes neuronales profundo que logra una tasa de detección utilizable del 95%, a una tasa de falsos positivos extremadamente baja y, según los autores, se adapta a volúmenes de ejemplos de capacitación en el mundo real sobre productos de hardware comunes [24].

Para finalizar mencionamos un trabajo que tiene en cuenta la complejidad de la información, y es expuesto por Alshahwan et al, en el que estudia la distancia de compresión normalizada (NCD) aplicada directamente a los binarios. La NCD es una medida teórica de la información y le permite obtener un 97.1% de precisión y una tasa de falsos positivos del 3% al momento de decidir si un programa sospechoso presenta mayor similitud a un malware o a un software benigno. Además, demostraron que esa precisión se puede optimizar combinando NCD con las tasas de compresibilidad de los ejecutables. Alshahwan remarca que el tiempo y el costo de cálculo de este método no es trivial [25].

Si bien hoy existen más tipos de malware que los descriptos, nos centramos en los que más impacto han generado en los últimos años. Lo mismo sucede con las técnicas de detección, para cada método o técnica (por ejemplo, machine learning) existen innumerables trabajos de los cuales solo hemos descripto los que consideramos más representativos y puntos de partida para el presente proyecto.

2 Desarrollo

Para determinar si un software es maligno es necesario monitorear sus procesos, por lo que para saber si en una computadora se está ejecutando un software malicioso es obligatorio monitorizar todos los procesos que se ejecutan en la misma. Para las computadoras que tienen un Sistema Operativo Linux/GNU es necesario monitorear las llamadas al sistema que realiza cada uno de estos procesos. Una llamada al sistema es una interfaz fundamental entre una aplicación y el kernel de Linux [26].

Para lograr este monitoreo se desarrolló un módulo del kernel [27] que intercepta ciertas llamadas al sistema y monitorea los procesos actuales a partir de dichas syscalls. De las 313 llamadas al sistema (syscalls) [19] que existen se seleccionó un conjunto, que permitirá identificar si un proceso es maligno. Las llamadas al sistema seleccionadas para esta versión del detector de malware están descriptas en la Tabla 1.

Table 1. Llamadas al sistema seleccionadas para esta primera versión del detector de malware.

ID	Nombre	Descripción
0	read	Lee bytes de un archivo referenciado por un file descriptor a un buffer.
1	write	Escribe bytes desde un buffer al archivo referenciado por el file descriptor
4	stat	Obtiene información sobre un archivo, como por ejemplo tiempo
2	openat	Abre un archivo
5	fstat	Obtiene información sobre un archivo

3	close	Cierra un file descriptor, por lo tanto, el archivo referenciado ya no puede ser accedido
101	ptrace	Permite observar y controlar la ejecución de un proceso

Estas llamadas al sistema fueron seleccionadas en base a la experiencia de los integrantes con ransomwares.

Para lograr este componente de monitoreo de llamadas al sistema fue necesario definir las siguientes interrogantes: ¿Cómo interceptar las llamadas al sistema? ¿La máxima cantidad de procesos se va a poder interceptar? ¿La máxima cantidad de llamadas al sistema por proceso se van a interceptar? ¿Se puede lograr que este monitoreo sirva para cualquier versión del kernel linux? Cuando se habla de proceso, se hace referencia al proceso asociado al programa que se está ejecutando en el sistema operativo GNU/Linux.

Lo primero que hace el módulo de monitoreo es inicializar la estructura de datos que va a usar para almacenar, a la cual se la puede conceptualizar como una matriz de 500x400 (Fig .1.).

La elección de 500 como límite de procesos a monitorizar es empírica, puede variar eventualmente. Por otro lado, el límite de 400 esta definido por la cantidad de llamadas al sistema definida en la Tabla de llamadas al sistema para sistemas x86 y x86_64 [19].

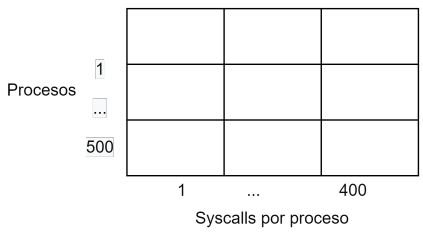


Fig. 1. Conceptualización estructura de datos para almacenar.

Una vez inicializada esta estructura, el siguiente paso es encontrar la tabla de llamadas al sistema, para poder "engancharse" e interceptar todas las llamadas que realizan las aplicaciones al kernel. Una vez que se encuentra esta tabla, es necesario quitar la protección contra escritura, osea el bit 16 del CR0 (Control Registry 0).

Los registros de control (CR0, CR1, CR2, CR3 y CR4) determinan el modo de operación del procesador y las características de ejecución de las tareas concurrentes.

El setear el bit 16 del CR0 permite a los procedimientos de nivel superior escribir en páginas de solo lectura, como por ejemplo en la página del Sistema Operativo donde se encuentra la tabla de llamadas al sistema [28]. Con esta protección contra escritura quitada queda "engachar" la función que cuenta las llamadas a cada syscall y por último volver a setear dicha protección contra escritura. A continuación, se muestra parcialmente el código que intercepta las syscalls originales y las redirige a una función

que acumula estas llamadas en base al PID y el ID de la syscall interceptada, para luego llamar a la función original de la syscall.

```
if (syscall_table != NULL) {
    write_cr0 (read_cr0 () & (~ 0x10000));
    original_open = (void *)syscall_table[__NR_open];
    syscall_table[__NR_open] = (long) &new_open;
    original_write = (void *)syscall_table[__NR_write];
    syscall_table[__NR_write] = (long) &new_write;
    original_read = (void *)syscall_table[__NR_read];
    syscall_table[__NR_read] = (long) &new_read;
    write_cr0 (read_cr0 () | 0x10000);
    printk(KERN_INFO "Syscall top iniciado\n");
 }
```

Una vez que este módulo es cargado al sistema, inicia con la lógica explicada recientemente, acumulando la cantidad de llamadas al sistema solicitadas por cada proceso que corre en el sistema operativo.

Cuando el módulo es descargado del kernel, se quita la protección contra escritura (bit 16 CR0 del procesador), se restablecen las funciones originales de cada syscall y se vuelve a establecer dicha protección contra escritura.

Una vez que este módulo es cargado al sistema, inicia con la lógica explicada recientemente, acumulando la cantidad de syscalls ejecutadas por cada proceso que corre en el sistema operativo.

Por último, se imprime un resumen de por cada PID “vivo” que el módulo encontró:

```
"PID: %1 - SYSCALL(%2) = %3"

Donde:

%1 Representa el Process ID del proceso.

%2 Representa el ID de la llamada al sistema.

%3 Representa la cantidad de veces que la llamada al
sistema con ID %2 fue llamada por el proceso con ID %1.
```

Las pruebas de esta herramienta de monitoreo por el momento han sido con ransomwares, logrando detectar un posible umbral de llamadas al sistema de escritura (write) de 1000, lo mismo para las llamadas al sistema de lectura (read) frente a programas “normales” o de uso diario.

3 Conclusiones

Este monitoreo de llamadas al sistema permitirá en base a ciertas reglas, por ejemplo, de cantidad de lecturas (read) o escrituras (write) o una combinación de ambas detener el proceso asociado a dichas llamadas al sistema.

Es importante mencionar que cuando se diseñó este módulo todavía no había salido la versión 5 del kernel Linux. Después de analizar esta nueva versión concluimos que, si bien este módulo de monitoreo no se ejecuta sobre la misma, el esfuerzo necesario para que esto ocurra no es mucho. Es decir, con algunas adaptaciones es posible llevar este monitoreo a la versión 5 del kernel Linux.

Como se mencionó previamente esta herramienta es un proyecto cuyo objetivo es detectar malware basado en patrones de llamadas al sistema en sistemas GNU/Linux cuyo resultado previsto es el desarrollo de un sistema de monitoreo y detección de malware para sistemas GNU/Linux, compuesto de dos herramientas.

La primera es una herramienta de monitoreo sobre las llamadas al sistema que cada proceso ejecuta y la cual generará diferentes vistas para que la información pueda ser comprensible. El actual trabajo forma parte de la primera herramienta y cuyos próximos pasos es generar una forma de visualización de estos datos. Esta forma de visualización debe ser en tiempo real y de lectura compresible, para posteriormente proceder con la segunda herramienta, la cual se integrará con la primera y permitirá detectar patrones posiblemente maliciosos en los procesos para informar al usuario de esta situación, pudiendo tomar una decisión.

Por otro lado, si bien las pruebas realizadas hasta ahora han sido con ransomwares también se ha identificado que esta técnica podría detectar minado de criptomonedas. Para esto sería necesario realizar pruebas para detectar llamadas al sistema usadas y posibles umbrales de cantidad de llamadas realizadas para inducir que se podría estar dando un minado de criptomonedas no autorizado.

Referencias

1. Canzanese R. (2015). Detection and Classification of Malicious Processes Using System Call Analysis. Recuperado el 28 de Mayo de 2019 https://pdfs.semanticscholar.org/8060/eae74c98a66cfcc736f4fca61d46f4dbc1d4.pdf.
2. Elisan C. (2015) Advanced Malware Analysis. McGraw-Hill. Capítulo 2. ISBN: 9780071819756.
3. K. Savage, P. Coogan, and H. Lau, (2018). The Evolution of Ransomware. Secur. Response, p. 57, 2015.
4. Tannenbaum A. (2009). Sistemas operativos modernos. Tercera edición. Pearson Educación. ISBN: 978-607-442-046-3.
5. Cruz M. (Junio 2017). Security 101: The Rise of Fileless Threats that Abuse PowerShell. Recuperado el 28 de Mayo del 2019 de https://www.trendmicro.com/vinfo/us/security/news/security-technology/security-101-the-rise-of-fileless-threats-that-abuse-powershel.

6. Viscuso M. (Febrero 2017). What Is a Non-Malware (or Fileless) Attack?. Recuperado el 28 de Mayo de 2019 de https://www.carbonblack.com/2017/02/10/non-malware-fileless-attack/.
7. Fileless Malware Attacks Are on the Rise, SentinelOne Finds, http://eds.b.ebscohost.com/eds/detail/detail?vid=2&sid=ee6fa74a-009d-4d30-9b94-fe6f826e0804%40sessionmgr103&bdata=Jmxhbmc9ZXMmc2l0ZT1lZHMtbGl2ZQ%3d%3d#AN=131651919&db=bsx.
8. Barros R., San-José P., Villanueva X. (2017) ¿Qué impacto ha tenido el ciberincidente de WannaCry en nuestra economía?. Deloitte. Recuperado de http://perspectivas.deloitte.com/hubfs/Campanas/WannaCry/Deloitte-ES-informe-WannaCry.pdf.
9. Yarochkin F., Kropotov V. (Febrero 2017). Lurk: Retracing the Group's Five-Year Campaign. Trend Micro. Recuperado el 28 de Mayo de 2019 https://blog.trendmicro.com/trendlabs-security-intelligence/lurk-retracing-five-year-campaign/?_ga=2.257191439.544304570.1558717075-418567566.1558717075.
10. Beek C., Dunton T., Grobman S., Karlton M., Minihane N., Palm C., Peterson E., Samani R., Schmugar C., Sims R. A., Sommer D., Sun B. (Junio 2018). McAfee Labs Threats Report. McAfee. Recuperado el 28 de Mayo de 2019 https://www.mcafee.com/enterprise/en-us/assets/reports/rp-quarterly-threats-jun-2018.pdf.
11. Morgan S. (Mayo 2017). 2018 Cybersecurity Market Report. Recuperado el 28 de Mayo de 2019 de https://cybersecurityventures.com/cybersecurity-market-report/.
12. Morgan S. (Diciembre 2018). Cybercrime Damages $6 Trillion By 2021. Recuperado el 28 de Mayo de 2019 de https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021.
13. Choudhary, S., Saroha, R., & Beniwal, S. (2016). How Anti-virus Software Works?. International Journal of Advanced Research in Computer Science and Software Engineering, (April 2013), 5–7.
14. Luckett P., McDonald J., Glisson W., Benton R., Dawson B., Doyle A. (2018). Identifying stealth malware using CPU power consumption and learning algorithms". Journal of Computer Security 26(2018) 589-613. DOI 10.3233/JCS-171060.
15. Gupta, S. & Kumar, P. (2015). An Immediate System Call Sequence Based Approach for Detecting Malicious Program Executions in Cloud Environment. Wireless Pers Commun, 81: 405.
16. Popli N.K., Girdhar A. (2019) Behavioural Analysis of Recent Ransomwares and Prediction of Future Attacks by Polymorphic and Metamorphic Ransomware. In: Verma N., Ghosh A. (eds) Computational Intelligence: Theories, Applications and Future Directions - Volume II. Advances in Intelligent Systems and Computing, vol 799. Springer, Singapore.
17. Ransomware Detection Using Limited Precision Deep Learning Structure in FPGA.
18. Kok S.H., Abdullah A., Jhanjhi N. Z., Supramaniam M. (2019). Ransomware, Threat and Detection Techniques: A Review. IJCSNS International Journal of Computer Science and Network Security, VOL.19 No.2. Recuperado el 28 de Mayo de 2019 de http://paper.ijcsns.org/07_book/201902/20190217.pdf.
19. Searchable Linux Syscall Table for x86 and x86_64. https://filippo.io/linux-syscall-table/. Última visita 08/08/2021.
20. Roman Gonzalez A. (2012). Clasificación de Datos Basado en Compresión. Revista ECIPeru, pp.69-74. ffhal-00697873. Recuperado el 28 de Mayo de 2019 de https://hal.archives-ouvertes.fr/hal-00697873/document.
21. Canzanese, R., Mancoridis, S., & Kam, M. (2015). System Call-Based Detection of Malicious Processes. In Proceedings - 2015 IEEE International Conference on Software Quality, Reliability and Security, QRS 2015 (pp. 119-124). [7272922] Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/QRS.2015.26.

22. Acampora, G., Bernardi, M. L., Cimitile, M., Tortora, G., Vitiello, A. (2018). A fuzzy clustering-based approach to study malware phylogeny. IEEE International Conference on Fuzzy Systems, , 2018-July doi:10.1109/FUZZ-IEEE.2018.8491625.
23.Asmithad K. A., Vinop P. (2014). A machine learning approach for linux malware detection. In: 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), ISBN 978-1-4799-2900-9.
24. Saxe J., Berlin k. (2015). Deep Neural Network Based Malware Detection Using Two Dimensional Binary Program Features. Recuperado de https://ia802808.us.archive.org/14/items/arxiv-1508.03096/1508.03096.pdf.
25. Alshahwan N., Barr E.T., Clark D., Danezis G. (2015). Detecting Malware with Information Complexity. Recuperado de https://archive.org/details/arxiv-1502.07661/page/n5
26. Manual de Programación de Linux. Página Oficial: http://man7.org/linux/man-pages/man2/syscalls.2.html.
27.Modulos kernel Linux. https://www.kernel.org/doc/html/v4.16/admin-guide/README.html
28. Mayo 2020. The Intel 64 and IA-32 Architectures Software Developer's Manual. Intel. Link a https://software.intel.com/content/dam/develop/public/us/en/documents/325384-sdm-vol-3abcd.pdf. Última visita 08/08/2021.

ISBN 978–987–633–574–4

Métricas para blockchain

Javier Díaz[1], Mónica D. Tugnarelli[2], Mauro F. Fornaroli[2],
Facundo N. Miño[2], Lucas Barboza[2]

[1]Facultad de Informática – Universidad Nacional de La Plata
jdiaz@unlp.edu.ar
[2] Facultad de Ciencias de la Administración – Universidad Nacional de Entre Ríos
{monica.tugnarelli, mauro.fornaroli, lucas.barboza}@uner.edu.ar

Abstract. En este artículo se presentan un conjunto de métricas iniciales para abordar el análisis de rendimiento de tecnologías blockchain, principalmente con herramientas aplicadas a la Blockchain Federal Argentina y una primera aproximación a mediciones sobre Hyperledger Fabric. Considerando que si bien hay aspectos conocidos para medir el rendimiento, aun no existe un marco común que facilite la tarea de lograr una medición comparativa entre las distintas soluciones de blockchain, lo cual, y considerando el sostenido uso de esta tecnología en un amplio campo de aplicación, se muestra como un área de vacancia sobre la cual consideramos que es necesario avanzar con el objetivo de evaluar el rendimiento en diferentes casos de uso y escenarios.

Keywords: blockchain, métricas, BFA, Ethereum, Hyperledger

1 Introducción

En este artículo se presentan los avances del PID-UNER 7059 denominado "*Tecnología Blockchain para aseguramiento de evidencia digital en entornos Forensic Readiness*" que tiene como objetivo principal analizar el impacto de la utilización de esta tecnología aplicada a la preservación, integridad y trazabilidad de evidencia digital, la cual es obtenida a priori de los activos señalados como esenciales en una organización, en un entorno preventivo como lo es Forensic Readiness, también llamado Preparación Forense [1].

El PID cuenta con varias etapas, en las cuales se trabaja con blockchain sin criptomoneda asociada para lograr implementar un prototipo donde realizar pruebas que permitan analizar cómo reacciona esta tecnología frente a los requerimientos de Forensic Readiness tanto para el proceso de asegurar la evidencia como para el mantenimiento de la cadena de custodia.

En trabajos anteriores [2] se ha analizado las características de diferentes tipos de blockchain disponibles en el mercado y, de acuerdo a los objetivos planteados en esta investigación, se focaliza el análisis en dos soluciones representativas, una plataforma pública, distribuida y descentralizada como Ethereum [3] y la solución privada Hyperledger Fabric [4] de administración centralizada, considerando para este análisis

aspectos tales como: privacidad, seguridad, velocidad de validación de transacciones, casos de uso, estándar abierto, entre otros [5] .

2 Métricas

A medida que se avanzó con las etapas del PID, se hizo necesario contar con algunos indicadores que ayuden a medir la perfomance y el rendimiento de cada tipo de blockchain.
Si bien se plantean aspectos conocidos para la medición del rendimiento no hay un marco común que facilite la tarea de lograr una medición comparativa en las diferentes implementaciones de las soluciones de blockchain, lo cual considerando el sostenido uso de esta tecnología en distintos ámbitos se muestra como un área de vacancia sobre la cual consideramos que es conveniente avanzar.
Al respecto, se delinearon algunas métricas iniciales sobre la Blockchain Federal Argentina como ejemplo de Ethereum y una primera revisión del tema sobre Hyperledger Fabric instalada como base de pruebas en laboratorio. Para el primer caso se utilizaron dos herramientas disponibles en el sitio de BFA, *bfascan*[1] desarrollada por Última Milla y un monitor implementado con *Grafana*[2] , las cuales se presentan en el siguiente punto.

2.1 Métricas sobre BFA

La Blockchain Federal Argentina [6] es una plataforma multiservicios abierta y participativa basada en tecnología Ethereum y pensada para integrar servicios y aplicaciones sobre blockchain. Está conformada por sectores públicos, privados, académicos y de la sociedad civil que participan desde la ingeniería organizacional hasta el despliegue de la infraestructura donde ningún sector tiene mayoría y eso evita que pueda ser manipulada. Cuenta con una variedad de casos de uso interesantes y con la ventaja de que el servicio de Sello de Tiempo 2.0 de BFA provee una hora oficial segura para usar en distintos procesos, el cual permite demostrar que el contenido de cualquier documento digital existió en un momento y que desde entonces no ha cambiado.
En cuanto a su operatoria, cada entidad que administra un nodo de BFA es responsable de su mantenimiento y monitoreo y no existe en la red un sistema central de administración. Como apoyo, BFA implementa un esquema de monitoreo a través del NOC (Network Operation Center), "*que estará atento al funcionamiento de los nodos selladores y gateway*". El mismo no tiene una ubicación centralizada, sino que está distribuido geográficamente y entre varias partes de la organización. Cabe destacar que la Facultad de Ciencias de las Administración cuenta con un nodo sellador administrado por dos integrantes de este equipo de investigación.

[1] BFA SCAN http://www.bfascan.com.ar/
[2] Monitor https://bfa.ar/monitor

Sobre esa distribución de nodos se realizó la captura de datos y se aplicaron las herramientas disponibles de análisis.
La siguiente tabla y gráficos muestran datos obtenidos sobre la Blockchain Federal Argentina en cuanto a cantidad de nodos operativos, volumen de transacciones y cuentas creadas en la BFA:

Cantidad de nodos que componen la red	-Cantidad total de nodos transaccionales.	90 nodos transaccionales
	-Cantidad total de nodos selladores	21 nodos selladores en línea y operativos
Transacciones del día	Cantidad total de transacciones realizadas desde las 0:00hs	Información variable: promedio de 10.000 transacciones en días laborables.
Total de transacciones	Cantidad total de transacciones realizadas.	Información variable: acumulado 5350650
Total de Direcciones	Total de cuentas creadas en la BFA.	Información variable: 805

Tabla 1. Recopilación de datos BFA con BFA Scan 16/8/21. Fuente: Elaboración propia

Fig.1 Captura de información estadística con BFAScan 16/8/21. Fuente: Web BFAScan

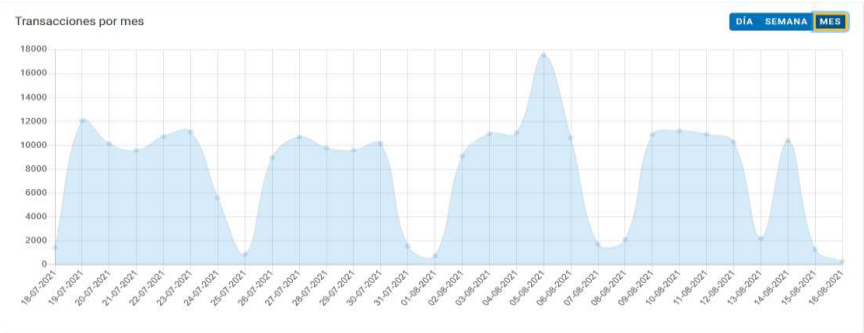


Fig.2 Captura de información estadística con BFAScan 16/8/21. Fuente: Web BFAScan

En base a la información anterior se proponen algunas métricas iniciales:

- **TPT_BFA**: Tiempo promedio de una transacción = Transacciones del Día / tiempo transcurrido desde las 0:00hs (al momento de calcular).
- **CTPN_BFA**: Cantidad de transacciones promedio por nodo = Total de Transacciones / Total de Nodos Selladores
- **CTND_BFA**: Cantidad de transacciones promedio por nodo del día = Transacciones del Día / Total de Nodos Selladores.

Además de la herramienta anterior, el sitio *BFASCAN* ofrece una API llamada *API REST BFA Scan* que presenta distintos métodos para obtener información contenida en la BFA.

a) **Método getBlocks** *url: http://201.190.184.52/bfascan/Blocks/getBlocks*

- HTTP GET. Devuelve información de los últimos 10 bloques.
- HTTP POST. Devuelve información de los últimos 100 bloques, o de uno en particular (a partir del hash).
- HTTP POST para usuarios registrados: Devuelve información de los últimos *n* bloques (como máximo 1000), o de uno en particular (a partir del hash). La consulta requiere un token que es enviado en un correo electrónico luego del registro.

De la información retornada por el método *getBlocks*, pueden utilizarse los *timestamps* de cada registro para calcular el tiempo promedio de creación de un bloque. Por ejemplo:

- **TPCB_BFA: Tiempo promedio de creación de bloque** = tiempo transcurrido desde el primer al último bloque devueltos por la consulta / el total de bloques devueltos por la consulta. (cuando más bloques puedan consultarse más precisa podrá resultar la aproximación).

Otro campo que devuelve la consulta en cada registro es *transactions_associated*. Este valor podría utilizarse para calcular la cantidad de transacciones promedio por bloque sumando la cantidad de transacciones de cada bloque por el total de bloques.

- **CPTB_BFA: Cantidad promedio de transacciones por bloque** = suma de las transacciones asociadas a cada uno de los bloques devueltos por la consulta / el total de bloques devueltos por la consulta.

b) Método getTxs *url: http://201.190.184.52/bfascan/transactions/getTxs*

- HTTP GET. Devuelve información de las últimas 10 transacciones creadas en la BFA.
- HTTP POST. Devuelve información de las últimas 100 transacciones, o de una en particular (a partir de su hash)
- HTTP POST para usuarios registrados: Devuelve información de las últimas transacciones (como máximo 1000), o de una en particular (a partir del hash). La

consulta requiere un token que es enviado en un correo electrónico luego del registro.

Al igual que con el método *getBlocks* podrían utilizarse los *timestamps* de los registros devueltos por la consulta para calcular el tiempo promedio de creación de una transacción (o tiempo entre transacciones). Por ejemplo:

- **TPCT_BFA: Tiempo promedio de creación de una transacción** = tiempo transcurrido entre la primera y la última transacción retornada / cantidad de transacciones retornadas (lo mismo que en el caso anterior, a mayor cantidad de registros consultados más aproximada podría ser la estimación)

Estas consultas podrían repetirse a intervalos de tiempo regulares, y luego realizar una estimación a partir de todos los tiempos promedios obtenidos.

El monitor BFA de Grafana, y a partir de la información que se muestra, permite obtener información para otra métrica inicial:

- **CPBS_BFA**: Cantidad promedio de bloques sellados por nodos (a partir de la información de último bloque sellado en *x* período de tiempo (permite consultar por últimos *n* minutos, horas, días) Estos valores se podrían calcular en distintos momentos de tiempo y luego calcular un promedio sobre estos para una mejor estimación.

Fig.3 Captura de información de monitoreo 16/8/21. Fuente: Web Monitor

Y otro indicador:

- **TF_BFA**: Tiempo de finalización: tiempo necesario para alcanzar la inmutabilidad de transacciones y bloques.

2.1 Métricas sobre Hyperledger

En cuanto a Hyperledger Fabric que es la blockchain instalada en laboratorio, provee un archivo llamado *configtx.yaml* para la configuración de ciertas características de la red. En él se encuentran dos parámetros importantes *BatchSize y BatchTimeout* que permiten configurar el rendimiento y la latencia de las transacciones.

- **Batch size:** Define cuántas transacciones recopilará el nodo ordenador antes de cerrar un bloque. Ningún bloque superará el tamaño de *AbsoluteMaxBytes* ni tendrá más de *MaxMessageCount* transacciones dentro. El tamaño ideal de construcción del bloque es de *PreferredMaxBytes*. Las transacciones que sean mayores a este tamaño aparecerán en un bloque propio.

- **Batch timeout:** es un mecanismo de reserva si el bloque no se llena en un tiempo específico. Este valor proporciona un límite superior para el tiempo que se tarda en cerrar un bloque de transacciones. Al disminuir este valor se mejorará la latencia, pero al hacerlo demasiado pequeño puede que se reduzca el rendimiento al no permitir que el bloque se llene a su capacidad máxima.

Cuanto menores sean los valores de *Batch timeout* y *Batch size*, mayor va a ser el número total de bloques generados por segundo. En cambio, mientras mayor sean sus valores, menor será el número de bloques generados por segundo.
Reduciendo el valor de *Batch timeout* disminuirá la latencia, pero a expensas del rendimiento total. Por el contrario, aumentar el *MaxMessageCount* hará que aumente el rendimiento total pero a expensas de la latencia de la transacción. Esta latencia, que se obtiene de restar tiempo de confirmación – tiempo de envío, es una vista de toda la red relacionada a la cantidad de tiempo que tarda una transacción en hacerse efectiva y propagarse por toda la red.
En términos generales, no existen valores ideales a definir, por lo que se deberán hallar de acuerdo a los requerimientos del prototipo instalado en el marco de este proyecto de investigación y luego aplicar las métricas definidas en el punto anterior para ver su correlación entre esquemas de blockchain.
Y, como agregado, las diferentes versiones de Hyperledger Fabric (HLF), por ejemplo, HLF v0.6 y HLF v1.0, deben compararse en el mismo marco de evaluación para demostrar las ventajas / desventajas de rendimiento de las nuevas versiones [7].

3. Conclusiones y trabajos futuros

En este trabajo se han presentado algunas primeras métricas que servirán de base para medir la calidad, performance y escalabilidad de las aplicaciones de blockchain. El beneficio de contar con métricas que ayuden a identificar y encontrar posibles problemas en el funcionamiento de la blockchain, que permitan analizar el trabajo de los nodos actuando a la vez y de manera descentralizada, a identificar cuellos de botella, a determinar el uso de recursos, a optimizar los protocolos de consenso y a detectar la ocurrencia de ataques sobre la seguridad de la red ayudará a crear un entorno más controlado y seguro de operar.

Como trabajos futuros se aplicarán las métricas propuestas por un periodo suficiente para establecer un banco de pruebas como base para el análisis, tanto sobre la blockchain BFA como en la Hyperledger desplegada en laboratorio y realizar ajustes en caso de ser necesario. Además se avanzará en lograr indicadores relacionados con aspectos de seguridad en la cadena de bloques, principalmente tratar de medir como es la relación del esquema de seguridad implementado versus las funcionalidades que debe brindar la blockchain.

Referencias

[1] TAN, John. (2001). Forensic Readiness. http://isis.poly.edu/kulesh/forensics/forensic_readiness.pdf

[2] Díaz, Francisco Javier; Tugnarelli, Mónica Diana; Fornaroli, Mauro F.; Barboza, Lucas. Blockchain para aseguramiento de evidencia digital en entornos Forensic Readiness. XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020). ISBN: 978-987-3714-82-5 http://sedici.unlp.edu.ar/handle/10915/103377

[3] Ethereum. https://ethereum.org/en/

[4] Hyperledger. https://www.hyperledger.org/use/fabric

[5] Michael Crosby, et. al. BlockChain Technology: Beyond Bitcoin. Applied Innovation Review (AIR). Issue No. 2 June 2016. Berkeley. http://scet.berkeley.edu/wp-content/uploads/AIR-2016-Final-version-Int.pdf

[6] Blockchain Federal Argentina https://bfa.ar/

[7] C. Fan, S. Ghaemi, H. Khazaei and P. Musilek, "Performance Evaluation of Blockchain Systems: A Systematic Survey," in IEEE Access, vol. 8, pp. 126927-126950, 2020, doi: 10.1109/ACCESS.2020.3006078.

ISBN 978-987-633-574-4
ISBN 978-987-633-575-1

CACIC 2021

TRACK “GOBIERNO DIGITAL Y CIUDADES INTELIGENTES”

COORDINADORES

Elsa Estevez (UNS)
Ariel Pasini(UNLP)

RedUNCI

CACIC 2021

Universidad Nacional de Salta

ISBN 978–987–633–574–4

Sensado móvil como estrategia de participación ciudadana en Ciudades Inteligentes

Juan Fernández Sosa [1], Verónica Aguirre [1], Leonardo Corbalán [1]
Lisandro Delía [1], Pablo Thomas [1], Patricia Pesado [1]

[1] Instituto de Investigación en Informática LIDI (III-LIDI). Facultad de Informática – Universidad Nacional de La Plata, La Plata, Argentina.
Centro Asociado a la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)
{jfernandez, vaguirre, corbalan, ldelia, pthomas, ppesado}@lidi.info.unlp.edu.ar

Resumen. Las ciudades inteligentes utilizan las Tecnologías de la Información y la Comunicación para mejorar la calidad de los servicios públicos y el bienestar general de sus habitantes. En este modelo de urbe, la participación ciudadana permite a los vecinos de la comuna involucrarse en la vida social, política y económica de la ciudad, exponiendo sus reclamos y propuestas. En este trabajo se presenta un nuevo sistema de sensado participativo para informar anomalías en la red de distribución de agua potable domiciliaria. Las imágenes capturadas y los reportes generados por los ciudadanos con sus propios dispositivos móviles se publican y comparten. Como resultado se genera un mapa colaborativo de la calidad del agua que expone las zonas donde el suministro es inadecuado o deficiente.

Keywords: Dispositivos móviles; Sensado móvil; Sensado participativo; Ciudades inteligentes; Participación ciudadana.

1. Introducción

El desarrollo de las tecnologías digitales, la expansión de Internet y el crecimiento de las redes de telefonía móvil han generado profundos cambios políticos, económicos y sociales a nivel global. Entre las ventajas más destacadas que ofrece la tecnología digital actual están las facilidades que brinda para mejorar la participación ciudadana en la gobernanza de las ciudades.

Las políticas de gobierno que hacen uso de las Tecnologías de Información y la Comunicación (TIC) ofrecen excelentes oportunidades para transformar a las administraciones públicas en instrumentos del desarrollo sustentable. Al conjunto de estas prácticas mediadas por la tecnología que llevan a cabo los organismos del Estado se las conoce con el nombre genérico de “Gobierno Electrónico” o *e-Government*.

Desde su aparición, el teléfono móvil ha sido generador de grandes y veloces cambios sobre aspectos fundamentales de la sociedad. Como consecuencia, la calidad de vida de los ciudadanos ha cambiado, mejorando la forma en que se relacionan y comunican con los demás, la manera de trabajar y el estilo de vida que llevan en general.

Los teléfonos inteligentes actuales, *smartphones* de aquí en adelante, tienen en promedio mayor capacidad tecnológica que el Apolo 11 cuando llegó a la Luna en 1969. Ninguna tecnología conocida, ni siquiera Internet, se ha impuesto con tanta rapidez, ni ha evolucionado tanto en tan poco tiempo como la tecnología móvil [1].

La penetración y utilización masiva de la telefonía celular ha permitido a los estados establecer canales de comunicación efectivos y bidireccionales con la población. Por medio de estos canales, las administraciones ofrecen servicios de manera cómoda, efectiva y económica para los ciudadanos. A su vez, en flujo

retroalimentativo, los ciudadanos pueden aportar información e ideas valiosas a los organismos estatales, las 24 horas del día, los 7 días de la semana, sin importar el lugar físico donde se encuentren. Al conjunto de estas prácticas de gobierno mediadas por la tecnología y los dispositivos móviles se las conoce con el nombre genérico de "Gobierno Móvil" o *m-Government*.

Las ciudades inteligentes y la innovación que supone el *m-Government* implican mejores servicios urbanos y un aumento en el bienestar ciudadano. La comunicación C2G (*citizen to government*) que ocurre en *m-Government*, es una de las formas más usuales de participación ciudadana. Las notificaciones que las personas envían a las administraciones estatales respecto de problemas en la vía pública, o sobre cualquier incidencia en el ámbito de los servicios públicos, son ejemplos de este tipo de comunicación.

Sin lugar a duda, la tecnología móvil, al servicio de las ciudades inteligentes, ofrece grandes oportunidades para la participación directa de los habitantes en cuestiones de gobernanza, lo que empodera a la comunidad.

El resto de este trabajo se organiza de la siguiente manera: en el capítulo 2 se aborda el concepto de ciudades inteligentes y diferentes mecanismos que existen para la participación ciudadana, luego en el capítulo 3 se profundiza sobre los sistemas de sensado móvil y su aplicación en el ámbito de las ciudades inteligentes. En el capítulo 4 se presenta el diseño y descripción de un nuevo sistema de sensado móvil, propuesto para el monitoreo de la calidad de agua de la red de distribución domiciliaria. Finalmente se presentan conclusiones y trabajo futuro.

2. Ciudades inteligentes y participación ciudadana

Las ciudades inteligentes son aquellas que hacen uso de las TICs para mejorar el estilo de vida y la calidad de los servicios públicos ofrecidos a una comunidad. Este paradigma se construye en función de los cinco pilares que se grafican en la figura 1: *social*, *económico*, *medioambiental*, *gobernanza* e *infraestructura urbana* [2].

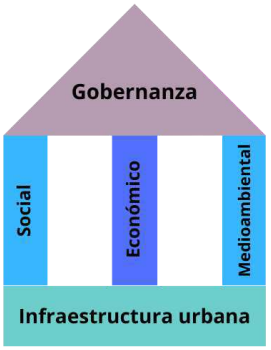


Fig 1. Pilares de una ciudad inteligente y sustentable.

La *infraestructura urbana* aporta las herramientas esenciales para satisfacer las necesidades de la comunidad, por ejemplo, el soporte para las comunicaciones informáticas. Sobre la infraestructura se apoyan los pilares *social*, *económico* y *medioambiental*. El primero está relacionado con la calidad de vida de las personas y tiene en cuenta factores tales como la salud, educación, seguridad y cultura, entre otros. El segundo pone el foco en el crecimiento económico responsable y la generación de oportunidades laborales. El tercer pilar se centra en la protección, monitoreo y restauración del medioambiente, además de la aplicación de prácticas eco-sustentables. Estos pilares se construyen a partir de políticas administradas por

la gobernanza estatal que formula e implementa marcos regulatorios, políticas públicas y sociales.

No basta con incorporar tecnología en las ciudades para que sean consideradas “inteligentes”. Estas tecnologías deben ser aprovechadas y explotadas por sus habitantes. La *Participación Ciudadana* es un nuevo modelo que ha surgido en los últimos años en el contexto de las ciudades inteligentes y que pone en valor las ideas y aportes de los ciudadanos [3]. Este modelo puede instrumentarse de tres formas diferentes: (1) ciudadanos como participantes democráticos, (2) ciudadanos como co-creadores y (3) ciudadanos como usuarios de las TICs.

En el primer caso, los ciudadanos forman parte de la toma de decisiones sobre las políticas a desarrollar en su ciudad. Este mecanismo de participación puede llevarse a cabo por ejemplo mediante sistemas de voto electrónico.

En el segundo caso de participación, los ciudadanos se convierten en generadores de ideas, informando a las administraciones locales sobre las necesidades reales de la sociedad. Un ejemplo de participación ciudadana para recolectar ideas y necesidades son los *focus groups*.

En el último caso de participación, se ponen a las TICs al servicio de los ciudadanos mediante:

- Aplicaciones orientadas al ciudadano y conectadas con la infraestructura tecnológica de la ciudad, que permiten reportar eventos en el tráfico, estado de las rutas, tratamiento de residuos, entre otros fenómenos de la ciudad.
- Infraestructura, integrando por ejemplo sensores e internet de las cosas en objetos de la vida cotidiana que son capaces de registrar y modificar el entorno.
- *Open data*, que consiste en difundir libre y abiertamente todos los datos que se producen en la ciudad: información del tráfico, el clima, el presupuesto del sector público, la información turística, entre muchos otros. Los desarrolladores pueden utilizar estos datos de libre acceso para crear aplicaciones de código abierto, que faciliten la colaboración entre los ciudadanos con el fin de resolver problemas a diferentes escalas (barrio, ciudad o incluso país).

3. Sensado con dispositivos móviles

Los *smartphones* son pequeños dispositivos móviles que además de las funciones básicas de un teléfono ordinario, ofrecen un conjunto de prestaciones tales como procesamiento de datos, cómputo, conectividad y acceso a Internet. Un conjunto de periféricos y sensores de *hardware* incrementan aún más estas capacidades. Actualmente, resultan cotidianos y se han convertido en herramientas fundamentales para la vida en sociedad. Es destacable la rapidez con la que se ha desarrollado la tecnología móvil, que llegó para instalarse definitivamente y promover al mismo tiempo, cambios sustanciales en los hábitos sociales y prácticas gubernamentales. Según estimaciones de la Universidad de las Naciones Unidas (UNU), en 2017 ya existían más cantidad de teléfonos móviles que habitantes en el mundo (7.700 millones de suscripciones a teléfonos móviles para una población mundial de 7.400 millones de personas) [4].

La cantidad de servicios y prestaciones ofrecidas por los *smartphones* se han incrementado de manera exponencial en los últimos años, apoyándose en el desarrollo y evolución de tres componentes fundamentales: el *hardware*, el *software* y las *tecnologías de comunicación inalámbrica*.

Respecto de los avances del *hardware*, se puede mencionar una mayor capacidad de procesamiento y de memoria, junto con la incorporación, en las últimas dos décadas, de diversos sensores que permiten a los *smartphones* "percibir" datos de diferente naturaleza. Asimismo, se ha ampliado la capacidad para anexar otros sensores externos de propósitos específicos.

Desde la perspectiva del *software*, se destaca la evolución de los actuales sistemas operativos que han permitido y promovido el desarrollo y ejecución de aplicaciones para dispositivos móviles. A través de estas aplicaciones se ofrecen infinidad de servicios a los usuarios.

Por último, los avances en las *tecnologías de comunicación inalámbrica* han permitido una mejora sustancial para el acceso e intercambio de información a través de Internet, utilizando las redes 3G, 4G y 5G o por medio de WiFi. Además se han incorporado diferentes tecnologías de comunicación de bajo consumo energético como Bluetooth y NFC, entre otras.

Resulta especialmente interesante notar la cantidad de sensores que actualmente incorporan los *smartphones* (ver figura 2). Esto ha permitido el desarrollo de una nueva área de estudio conocida como *Sensado con dispositivos móviles* [5]. Aquí se explora la capacidad de percepción del entorno a partir de la creación de redes de sensado en diferentes niveles: *personal*, *grupal* y *comunitario*. Esta práctica genera oportunidades en diversas áreas como el comercio, el monitoreo del ambiente, de la salud, del comportamiento humano, del tránsito y en otras actividades dentro del contexto de las ciudades inteligentes [6].

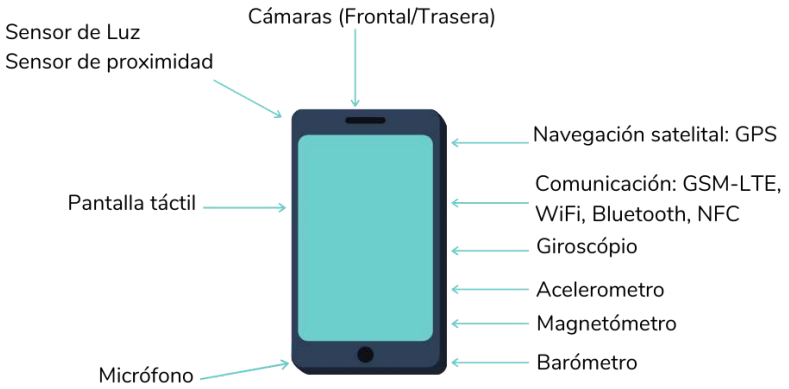


Fig 2. Lista no exhaustiva de sensores incorporados en los *smartphones*.

La figura 3 presenta una clasificación para las redes de sensado con dispositivos móviles. En función del objeto de estudio, estas redes pueden estar centradas en las *personas* o en el *ambiente* [7, 8]. Son ejemplos del primer caso las redes orientadas a documentar actividades, controlar ejercicios físicos o monitorear la salud y bienestar de las personas. Por su parte, las redes centradas en el ambiente buscan obtener información sobre el espacio que rodea al usuario, por ejemplo la calidad del aire, condiciones de caminos, tráfico, etc.

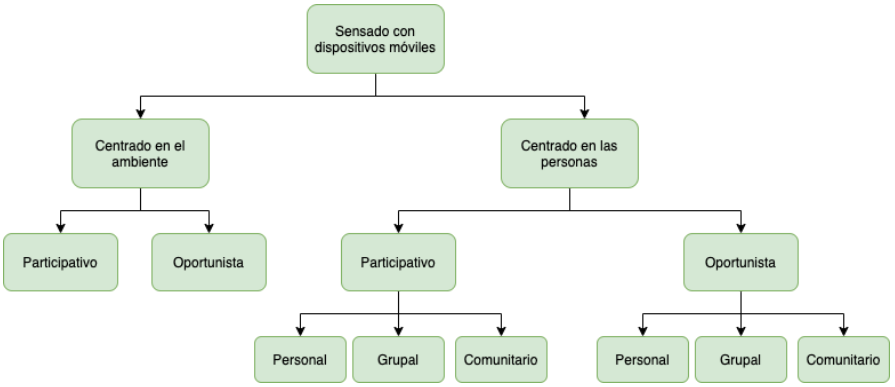


Fig 3. Taxonomía de los sistemas de sensado con dispositivos móviles

Dependiendo del grado de compromiso de los usuarios que participan en estas redes, surgen dos nuevas categorías: las redes *participativas* y las *oportunistas* [5, 6]. Las primeras requieren del rol activo del usuario quien debe elegir conscientemente qué datos recopilar, cómo, dónde, etc. Tal es el caso de un sistema para la recolección de muestras de ruido ambiental donde el usuario debe tener la iniciativa de tomar su *smartphone* y realizar la captura de audio correspondiente. Este tipo de paradigma de sensado permite detectar eventos y realizar operaciones complejas, apelando a la inteligencia y criterio de las personas. En contrapartida, la calidad de la información recopilada dependerá del entusiasmo de la persona en formar parte de la red de sensado. Por ello es necesario estudiar procesos de reclutamiento e incentivos adecuados.

En redes oportunistas, el usuario asume un rol pasivo como portador del dispositivo que de manera autónoma recolecta datos en *background*. Tal es el caso de las aplicaciones que buscan continuamente redes WiFi en el entorno por donde se mueve el usuario o que monitorean su actividad física (cantidad de pasos dados, escalones subidos/bajados, etc). Los problemas que aquí se presentan están relacionados con las dificultades para conocer el contexto de la recolección de los datos. Por ejemplo, una aplicación que registra en *background* el nivel de ruido presente en el entorno del usuario, debería descartar las muestras recolectadas cuando el dispositivo se encuentre guardado en el bolsillo.

Finalmente, las redes se clasifican según la escala o nivel de sensado en las siguientes categorías [5, 9]:

- *Personal*: Recolectan datos del portador del dispositivo. Ejemplos: Monitoreo del sueño, patrones de movimiento, medios de transporte utilizados, etc.
- *Grupal*: Recolectan y comparten información dentro de un grupo de individuos que persiguen un objetivo o interés común. Ejemplos: redes de trabajo, vecinales, universitarias, etc.
- *Comunitario*: Interviene un gran número de personas. La recolección, análisis e intercambio de datos se realiza para beneficiar a una comunidad entera. Ejemplos: aplicaciones para conocer el avance de una enfermedad, congestionamientos de tránsito, mapas de ruido, etc.

3.1 Sensado con dispositivos móviles en Smart Cities

En el contexto de las ciudades inteligentes, el sensado con dispositivos móviles ofrece a sus habitantes nuevas formas de participación en cuestiones relevantes de la vida social, política y económica de su comunidad. Algunos dominios de aplicación

son el transporte, monitoreo ambiental, salud pública, seguridad urbana, economía y educación [10].

Para aplicaciones que requieren la recolección de grandes volúmenes de datos, tales como el monitoreo ambiental, la industria o el gobierno suelen desplegar redes de sensores inalámbricas (WSN por sus siglas en inglés). Estas redes se construyen a partir de la interconexión de dispositivos electrónicos con capacidad de sensado, cómputo y conectividad inalámbrica [11]. Dichos dispositivos colaboran entre sí monitoreando algún fenómeno específico [12].

Como se mencionó anteriormente, los *smartphones* fueron adquiriendo mayores capacidades de cómputo y de sensado con el paso de los años. Gracias a ello, son posibles las redes de sensado móvil, donde los *smartphones* se erigen como nodos brindando ventajas y nuevas oportunidades en comparación con las WSN tradicionales [13]:

- El mantenimiento de cada nodo está a cargo del dueño del terminal. Esto incluye la recarga de batería, reparación y recambio del equipo por uno de mayores prestaciones.
- Se pueden formar redes de mayor tamaño y cobertura, aprovechando que estos dispositivos ya se encuentran desplegados en la sociedad.
- Para escalar el sistema sólo se debe reclutar nuevos usuarios.
- Un mismo dispositivo es capaz de producir datos de diferente naturaleza, y puede ejecutar múltiples apps para cubrir diferentes necesidades.

El sensado que se lleva a cabo utilizando los dispositivos móviles de los ciudadanos permite mejorar la visualización de los eventos urbanos. Ello posibilita informar sobre situaciones sociales o ambientales que antes requerían del despliegue de redes de sensores estáticas, costosas de mantener y escalar.

Existen varias aplicaciones que emplean el sensado móvil participativo y comunitario para el monitoreo ambiental. Un ejemplo de ello se presentó en [14] por medio de un *framework* para monitorear la contaminación ambiental. Este fue puesto a prueba en la ciudad de Dhaka, Bangladesh. El sistema está compuesto por una aplicación de *software* que se ejecuta en los dispositivos móviles de los participantes, un servidor central encargado del procesamiento de los datos recibidos y una aplicación web para visualizar la información, dirigida especialmente a quienes tienen poder de decisión al respecto.

La aplicación móvil permite a los usuarios tomar fotografías, grabar audios y video de accidentes de contaminación junto con la información de geolocalización provista por el dispositivo. Estos datos, y los de otros usuarios, pueden visualizarse en un mapa embebido en la app. Los reportes incluyen distintas categorías como desechos, contaminación sonora, lumínica, del agua, del aire etc.

En [15] y [16] se presentan dos iniciativas aplicadas al monitoreo del ruido ambiental: “Ear-Phone” y “NoiseTube” respectivamente. Dichas aplicaciones procesan las mediciones del micrófono de los dispositivos móviles para obtener el valor del nivel de ruido existente. Ambos proyectos de sensado participativo y comunitarios poseen dos componentes: una aplicación de *software* que se ejecuta en los *smartphones* de los participantes y un servidor central. Los datos recolectados por los dispositivos móviles son enviados junto con información georeferenciada para la construcción de un mapa de ruido.

4. Diseño propuesto para sistema de monitoreo de la calidad del agua en la red de distribución domiciliaria

En este capítulo se presenta el diseño de un nuevo sistema para el reporte de incidencias en la red de distribución de agua potable de una ciudad. Más adelante se exponen los detalles de implementación y aspectos técnicos abordados durante el desarrollo de su componente más distintivo: *la aplicación móvil*. Esta aplicación, instalada en los *smartphones* de los participantes, conforma la red de sensado móvil comunitaria que da soporte a la propuesta. Así, cualquier ciudadano puede reportar una variación en la calidad del servicio de distribución de agua domiciliaria utilizando la cámara y GPS de su dispositivo móvil.

4.1 Motivación y presentación del sistema

La red pública de distribución de agua potable se encarga de suministrar a los domicilios este elemento vital, asegurando condiciones de cantidad y calidad. Existen diferentes factores que pueden degradar esta calidad, como por ejemplo la presencia de metales pesados, microbios y sedimentos. Ello impacta directamente en la salud pública, exponiendo a la comunidad a brotes de enfermedades intestinales y otras infecciones [17]. Por lo tanto, es importante alertar a las autoridades y a las empresas responsables del servicio, ante cualquier cambio observable en las características del agua suministrada.

Sobre este escenario, se implementa un sistema de sensado participativo para reportar anomalías en el suministro domiciliario de agua. El conjunto de reportes contribuye a la construcción de un mapa de la calidad del agua, exponiendo zonas en donde el servicio no es el adecuado.

Los reportes, enviados a un servidor central a través de Internet, incluyen la siguiente información:

1. Una imagen obtenida con la cámara del dispositivo móvil del participante que expone la incidencia detectada.
2. La puntuación asignada por el usuario a la calidad del agua observada en un rango de 0 a 10. Este dato ayudará en la sistematización de la información y en la visualización de los reportes en el mapa.
3. Las respuestas del usuario a preguntas sobre características específicas del agua suministrada (color, olor, nivel de transparencia, etc.)
4. Información georeferenciada del dispositivo para ubicar el reporte en un mapa y también para alertar a otros participantes de la zona, quienes a su vez pueden realimentar al sistema generando sus propios reportes.

4.2 Aspectos técnicos

La solución desarrollada utiliza la arquitectura típica propia de los sistemas para el sensado móvil, donde se destacan tres componentes principales: los usuarios participantes, la aplicación de *software* para dispositivos móviles y un servidor central [18] (ver figura 4). Los usuarios interactúan con el sistema enviando reportes a través del aplicativo, instalado previamente en sus dispositivos móviles. La plataforma de sensado se encarga, en principio, de procesar, almacenar e informar dichos reportes.

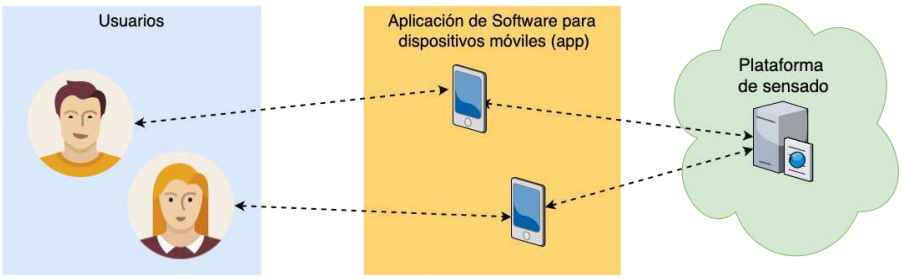


Fig 4. Componentes del sistema de sensado presentado.

El código fuente que implementa las interfaces de usuario y los casos de uso que dan soporte a los requerimientos funcionales de acceso a los sensores y GPS, se encuentran accesibles de manera pública en un repositorio de GitLab en [19].

La construcción de aplicaciones de *software* para dispositivos móviles incluye básicamente dos enfoques distintos: el *desarrollo nativo* y el *desarrollo multiplataforma* [20, 21]. El enfoque nativo consiste en el desarrollo de una aplicación específica para cada uno de los sistemas operativos soportados. En contraposición, el enfoque multiplataforma permite desarrollar un proyecto único generando desde el mismo código fuente las aplicaciones que se ejecutarán en las distintas plataformas o sistemas operativos. Mientras que en el enfoque nativo se utilizan las herramientas de desarrollo -SDK- provistos por cada una de las plataformas, en el enfoque multiplataforma los aplicativos pueden ser desarrollados empleando una gran diversidad de tecnologías.

Para el desarrollo de la aplicación de este sistema de sensado móvil, se adoptó un enfoque multiplataforma, de modo que permitiera, a partir de un único código fuente, generar las aplicaciones para los sistemas operativos destino, en este caso Android e iOS. Se utilizó Ionic [22], un *framework* que permite la construcción de Apps utilizando tecnologías web (HTML, CSS y JS). Dicho *framework* brinda acceso a diferentes características de los dispositivos por medio de *plugins* de código abierto creados por la comunidad [23].

El desarrollo de la aplicación requirió instalar los siguientes plugins y librerías:

- Camera [24]: *plugin* especializado para la captura de imágenes o videos. Se utiliza para tomar fotografías del estado del agua domiciliaria.
- Geolocation [25]: este *plugin* permite obtener la ubicación del dispositivo, en términos de latitud y longitud. Así es posible anexar al reporte la información necesaria para poder ubicar la incidencia en el mapa y alertar a usuarios en las cercanías.
- API de Javascript de Google Maps [26] para la visualización y personalización del mapa, con los diferentes reportes enviados por los participantes del sistema de sensado.

4.3 Interfaz de usuario

La aplicación móvil presenta una interfaz simple y amigable de modo que cualquier usuario pueda utilizarla, independientemente de su experticia. En la figura 5 se visualiza la pantalla principal. En ella se puede observar el mapa con diferentes reportes creados por otros participantes de la red. El mapa toma en consideración los datos del GPS del dispositivo para centrarse en esa ubicación. Presionando sobre los reportes en el mapa se accede a una descripción detallada de éstos conformada por

una imagen, etiquetas que describen el fenómeno reportado e información de la fecha y hora en que se recolectó esta información.

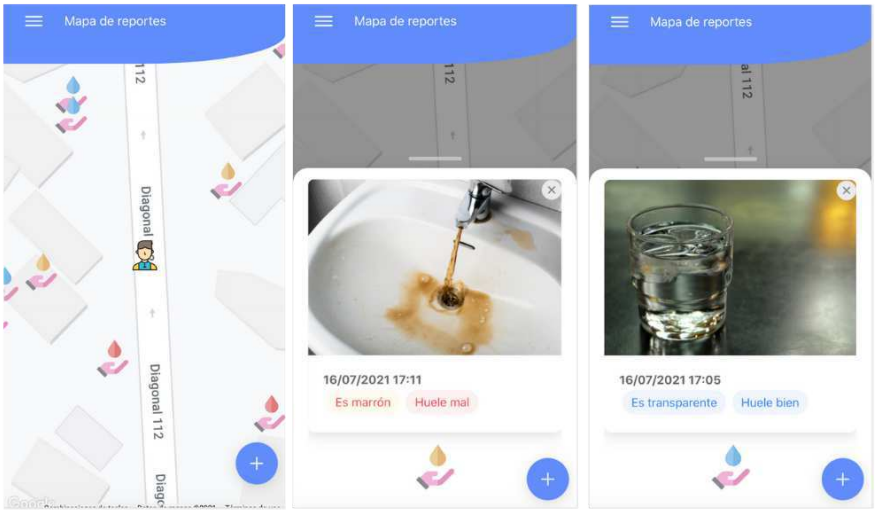


Fig 5. Pantalla principal de la app y detalles de algunos reportes de incidencias

El botón que se encuentra en el borde inferior derecho de la pantalla principal permite la generación y posterior envío de un nuevo reporte por medio del cuestionario que se visualiza en la figura 6. Aquí se debe suministrar la siguiente información:

1. Una fotografía que refleje el fenómeno reportado. La aplicación solicita acceso a la cámara del dispositivo para poder realizar la captura.
2. Un puntaje para evaluar la calidad del agua, entre 0 (mala) y 10 (buena). Este dato surge de la observación y percepción del usuario.
3. Una evaluación sobre el aspecto y olor del agua por medio de una selección entre múltiples opciones preconfiguradas.

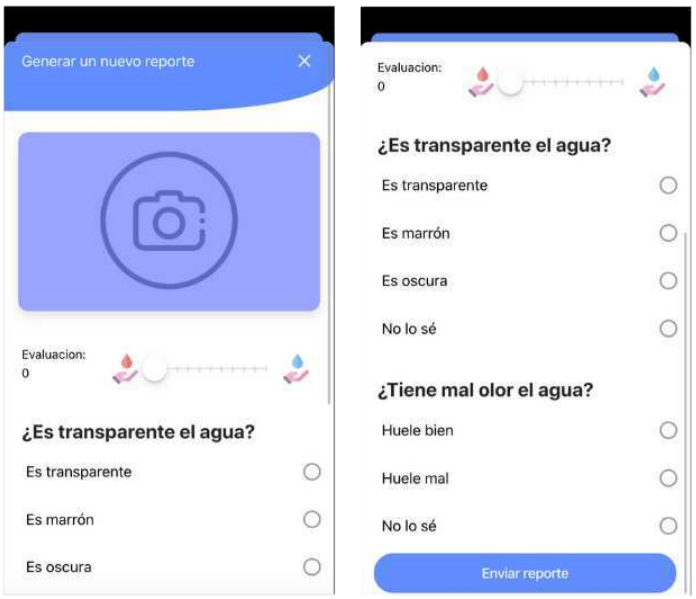


Fig 6. Pantalla para envío de reporte.

5. Conclusiones y Trabajo Futuro

El nuevo modelo para el diseño y desarrollo de la urbe moderna, la *ciudad inteligente*, hace uso de las *Tecnologías de Información y la Comunicación* para transformar a las administraciones públicas en instrumentos del desarrollo sustentable, al mismo tiempo que mejora la prestación de servicios y el bienestar general de sus habitantes. La *participación ciudadana*, promovida por este nuevo modelo, empodera a la comunidad otorgándole facultades para intervenir en cuestiones de gobernanza.

Una forma de participación ciudadana, la *comunicación C2G* (desde el ciudadano hacia el gobierno), permite a los vecinos de una comunidad informar a las autoridades sobre cualquier incidente que pudiese afectarles directa o indirectamente. El uso de la tecnología móvil juega aquí un papel destacado debido a que facilita considerablemente este tipo de comunicación.

Existen millones de dispositivos móviles desplegados actualmente en la sociedad. En los últimos años, han experimentado mejoras sustanciales en sus capacidades de cómputo, almacenamiento y prestaciones de servicios. La incorporación de una gran cantidad de sensores otorga a estos dispositivos la habilidad para "percibir" el contexto donde se encuentran. Las posibilidades que de aquí se desprenden son enormes y se estudian bajo el nombre de *sensado móvil*. Los resultados competen directamente a las ciudades inteligentes y son aplicables en áreas tales cómo la salud, educación, transporte, comercio y monitoreo ambiental entre muchas otras.

Los sistemas de sensado móvil típicamente utilizan tres componentes principales: los usuarios participantes, una aplicación de *software* instalada en sus dispositivos y una plataforma de sensado desplegada en la nube. De esta forma los ciudadanos participan en actividades comunales con sus propios dispositivos, generalmente un *smartphone*, monitoreando fenómenos ambientales y urbanos e informando a las autoridades estatales sobre ellos y otros tipos de incidencias de interés para la comunidad.

En este artículo se presentó un sistema de sensado móvil participativo. Su implementación permite al ciudadano monitorear la calidad del agua de red que llega a su domicilio. El resultado de su participación junto con la de otros vecinos, se obtiene a partir del sistema en forma de un mapa colaborativo que expone aquellas zonas donde el suministro presenta anomalías. De esta manera es posible alertar a toda la comunidad, a las empresas responsables del servicio y a las autoridades gubernamentales que tienen capacidad de acción para corregir el problema.

La aplicación móvil que se desarrolló permite generar el reporte de la anomalía observada y publicarlo para ser compartido en la nube. La información enviada al servidor central consiste en una fotografía tomada con el *smartphone* del usuario, la puntuación asignada a la calidad del agua, y la descripción de las características observadas. Los datos de geolocalización, aportados por el GPS del dispositivo, son adjuntados automáticamente en cada envío. La aplicación también permite la consulta por área geográfica de aquellos reportes generados por el propio usuario u otros participantes de la comunidad.

Se plantea como trabajo futuro la realización de pruebas de usabilidad para evaluar y eventualmente mejorar la interfaz de la aplicación desarrollada. Se incorporará la opción para generar un reporte extendido, con mayor cantidad de información, luego de un segundo examen más minucioso de la anomalía observada en el suministro de agua. También se extenderá la aplicación para dar soporte a dispositivos que no cuenten con GPS, permitiendo que el usuario elija manualmente

la ubicación del incidente en el mapa. Asimismo se dará soporte a la pérdida momentánea de conexión, almacenando el reporte localmente hasta que se pueda realizar el envío al servidor central.

Referencias

[1] Carrasco, C. A. P., & Ipanaqué, C. I. V. (2014). Adopción del m-government en el sector público. Quipukamayoc, 22(41), 155-164.

[2] Muñoz, R., Pasini, A. C., & Pesado, P. M. (2020). Innovación en el sector público para ciudades inteligentes sostenibles. In XXVI Congreso Argentino de Ciencias de la Computación (CACIC)(Modalidad virtual, 5 al 9 de octubre de 2020).

[3] Simonofski, A., Asensio, E. S., & Wautelet, Y. (2019). Citizen participation in the design of smart cities: Methods and management framework. Smart Cities: Issues and Challenges, 47-62.

[4] Baldé, C. P., Forti, V., Gray, V., Kuehr, R., & Stegmann, P. (2017). The global e-waste monitor 2017: Quantities, flows and resources. United Nations University, International Telecommunication Union, and International Solid Waste Association

[5] Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A. T. (2010). A survey of mobile phone sensing. IEEE Communications magazine, 48(9), 140-150.

[6] Khan, W. Z., Xiang, Y., Aalsalem, M. Y., & Arshad, Q. (2012). Mobile phone sensing systems: A survey. IEEE Communications Surveys & Tutorials, 15(1), 402-427

[7] Laport-López, F., Serrano, E., Bajo, J., & Campbell, A. T. (2020). A review of mobile sensing systems, applications, and opportunities. Knowledge and Information Systems, 62(1), 145-174.

[8] Kanhere, S. S. (2013, February). Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces. In International Conference on Distributed Computing and Internet Technology (pp. 19-26). Springer, Berlin, Heidelberg.

[9] Ganti, R. K., Ye, F., & Lei, H. (2011). Mobile crowdsensing: current state and future challenges. IEEE communications Magazine, 49(11), 32-39.

[10] Xiao, Z., Lim, H. B., & Ponnambalam, L. (2017). Participatory sensing for smart cities: A case study on transport trip quality measurement. IEEE Transactions on Industrial Informatics, 13(2), 759-770.

[11] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. Computer networks, 38(4), 393-422.

[12] Ali, S., Khusro, S., Rauf, A., & Mahfooz, S. (2014). Sensors and mobile phones: evolution and state-of-the-art. Pakistan journal of science, 66(4), 385

[13] Ma, H., Zhao, D., & Yuan, P. (2014). Opportunities in mobile crowd sensing. IEEE Communications Magazine, 52(8), 29-35.

[14] Ahmed, A. A. N., Haque, H. F., Rahman, A., Ashraf, M. S., Saha, S., & Shatabda, S. (2017). A participatory sensing framework for environment pollution monitoring and management. arXiv preprint arXiv:1701.06429.

[15] Rana, R. K., Chou, C. T., Kanhere, S. S., Bulusu, N., & Hu, W. (2010, Abril). Ear-phone: an end-to-end participatory urban noise mapping system. In Proceedings of the 9th ACM/IEEE international conference on information processing in sensor networks (pp. 105-116).

[16] Maisonneuve, N., Stevens, M., Niessen, M. E., & Steels, L. (2009). NoiseTube: Measuring and mapping noise pollution with mobile phones. In Information technologies in environmental engineering (pp. 215-228). Springer, Berlin, Heidelberg.

[17] Edition, F. (2011). Guidelines for drinking-water quality. WHO chronicle, 38(4), 104-108.
[18] Restuccia, F., Das, S. K., & Payton, J. (2016). Incentive mechanisms for participatory sensing: Survey and research challenges. ACM Transactions on Sensor Networks (TOSN), 12(2), 1-40.
[19] https://gitlab.com/jffs/trabajoespecialista-sensadomovil.git
[20] Delia, L., Galdamez, N., Thomas, P., Corbalan, L., & Pesado, P. (2015, Mayo). Multi-platform mobile application development analysis. In 2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS) (pp. 181-186). IEEE.
[21] Delia, L., Thomas, P., Corbalan, L., Sosa, J. F., Cuitiño, A., Cáseres, G., & Pesado, P. (2018, Julio). Development approaches for mobile applications: comparative analysis of features. In Science and Information Conference (pp. 470-484). Springer, Cham.
[22] https://ionicframework.com/
[23] https://ionicframework.com/docs/native/community
[24] https://ionicframework.com/docs/native/camera
[25] https://ionicframework.com/docs/native/geolocation
[26]https://developers.google.com/maps/documentation/javascript/overview?hl=es

Calidad de datos aplicada a la base de datos abierta de casos registrados de COVID-19

Ariel Pasini, Juan Ignacio Torres, Silvia Esponda, Patricia Pesado

Instituto de Investigación en Informática LIDI (III-LIDI)*
Facultad de Informática – Universidad Nacional de La Plata 50 y 120 La Plata Buenos Aires
*Centro Asociado Comisión de Investigaciones Científicas de la Pcia. de Bs. As. (CIC)
{apasini,jitorres,sesponda,ppesado}@lidi.info.unlp.edu.ar

Abstract. Para lograr una mayor transparencia y fomentar la participación de los ciudadanos en la toma de decisiones, y responder de manera eficiente, los gobiernos de diferentes países ponen una gran cantidad de información a disposición de su comunidad. Esta información, tomó más relevancia durante la pandemia del COVID-19. Sin embargo, si estos datos no poseen un buen nivel de calidad, la información pierde confiabilidad. Se presenta una evaluación de la calidad de datos del archivo público "COVID-19. Casos registrados en la República Argentina" utilizando el modelo que provee la norma ISO/IEC 25012 y el proceso de evaluación definido por la ISO/IEC 25040.

Keywords: Calidad de datos, ISO/IEC 25000, Gobierno abierto, Datos abiertos

1 Introducción

Desde el primer caso detectado en Wuhan (Hubei, China), el COVID-19 se ha expandido rápidamente a través de aproximadamente 200 países/regiones amenazando con la vida de las personas e irrumpiendo significativamente en la economía y la sociedad mundialmente.

La publicación de datos abiertos es crucial durante un período de pandemia, ya que, a través del análisis y procesamiento de estos, los ciudadanos ayudan en el proceso de toma de decisiones para responder de manera inteligente [1]. Es fundamental que los datos públicos en formatos abiertos (ya sea a través de datasets o reportes diarios) sean de alta calidad. Una baja calidad de datos afecta de manera negativa tanto al proceso de creación de valor como a la transparencia y la confianza en el gobierno [2].

El modelo general definido en la norma ISO/IEC 25012 – "Data Quality Model" [3] permite medir la calidad de los datos almacenados en un sistema informático. Es utilizado junto con otras normas para planificar y llevar a cabo evaluaciones de

calidad de datos. Este modelo, clasifica los atributos de calidad en quince características de acuerdo con dos puntos de vista que no son excluyentes: inherente y dependiente del sistema. Las características alcanzan diferente importancia y prioridad según las necesidades de las partes interesadas.

Por otro lado, la norma ISO/IEC 25040 – "Evaluation Process" [4] proporciona un modelo de referencia general para evaluar, que tiene en consideración las entradas al proceso de evaluación, restricciones y aquellos recursos que sean necesarios para obtener las salidas que correspondan. El proceso consta de un total de cinco actividades.

El presente artículo propone realizar la medición de la calidad de los datos obtenidos desde el archivo público *"COVID-19. Casos registrados en la República Argentina"* [5], mantenido por la Dirección de Epidemiología y Análisis de Situación de Salud, tomando ciertas características de calidad definidas por la norma ISO/IEC 25012 y siguiendo las actividades definidas en el modelo de evaluación establecido por la norma ISO/IEC 25040.

En la segunda sección se introducen las normas y las métricas a utilizar para la evaluación de la calidad de los datos. En la sección tres se introduce el concepto de Gobierno Abierto, se mencionan ventajas y desventajas de la publicación de datos abiertos y se resalta la importancia de la participación de los ciudadanos en el proceso de generación de valor a través del uso de estos datos. Luego, en el capítulo 4 se introduce detalladamente el recurso a evaluar y se realiza la evaluación de la calidad de la base de datos COVID-19. Finalmente, se presentan las conclusiones del artículo.

2 Calidad de Datos

La familia de normas ISO/IEC 25000, conocida como SQuaRE (System and Software Requirements and Evaluation) propone evaluar la calidad de un producto de software bajo un marco de trabajo común. En el contexto del artículo, se utilizarán dos normas de la familia, la ISO/IEC 25012 y la ISO/IEC 25040.

2.1 Modelo de calidad de datos ISO/IEC 25012

Innegablemente, la cantidad de datos manejada por sistemas informáticos se encuentra en aumento en todas partes del mundo. Es necesario en todo proyecto de tecnología de la información maximizar la calidad de los datos que se intercambian, procesan y utilizan entre sistemas. Una baja calidad de datos puede generar información insatisfactoria y resultados que no son de utilidad.

La norma define un modelo de calidad de datos en un formato estructurado dentro de un sistema informático. Este modelo está compuesto por quince características divididas en dos puntos de vista: inherente y dependiente del sistema. Cabe destacar que algunas características son pertinentes desde los dos puntos de vista.

Calidad de los datos inherente:

Este punto de vista se refiere al grado en que las características tienen el potencial de satisfacer las necesidades establecidas e implícitas cuando se utilizan los datos bajo condiciones especificadas. Las características pertinentes desde el punto de vista inherente Son: ***Exactitud***: hace referencia a que los datos poseen atributos representados de manera correcta (puede ser sintáctica o semántica). ***Completitud***: los datos poseen valores para todos los atributos esperados. ***Consistencia***: los datos no son contradictorios y son consistentes unos con otros. ***Credibilidad***: los datos poseen atributos ciertos y creíbles (incluye concepto de autenticidad). ***Actualidad***: los datos poseen atributos que son actualizados adecuadamente.

Calidad de los datos dependientes del sistema:

Hace referencia al grado en que se alcanza y preserva la calidad de los datos en un sistema informático cuando los datos se utilizan bajo condiciones específicas. Existe una dependencia entre la calidad de los datos y el dominio tecnológico en el cual los mismos se utilizan. Las características pertinentes desde el punto de vista dependiente del sistema son: ***Disponibilidad***: define el grado en que los atributos pueden ser recuperados por los usuarios y/o aplicaciones que tengan acceso. ***Portabilidad***: grado en que los datos pueden ser instalados, reemplazados o copiados de un sistema a otro manteniendo la calidad. ***Recuperabilidad***: los datos deben mantener y preservar un nivel de operaciones y calidad ante fallos.

Calidad de los datos inherente y dependientes del sistema:

Las características que son pertinentes desde ambos puntos de vista (inherente y dependiente del sistema) son: ***Accesibilidad***: grado en que se puede acceder a los datos. ***Cumplimiento***: los datos cumplen con normas, convenciones o regulaciones en relación a la calidad de datos en un contexto específico. ***Confidencialidad***: los datos tienen atributos que deben ser accedidos e interpretados solamente por usuarios autorizados. Es parte de la seguridad de la información, junto con la integridad y la disponibilidad. ***Eficiencia***: los atributos se pueden procesar y proporcionar niveles esperados de rendimiento utilizando tipos y cantidades de recursos correspondientes. ***Precisión***: los atributos pertenecientes a los datos presentan valores exactos o que permiten discriminación. ***Trazabilidad***: se analiza si se proporciona un registro para la auditoría del acceso a los datos y de cualquier modificación respecto a los mismos. ***Comprensibilidad***: grado en que los usuarios pueden leer e interpretar los datos, que deben estar expresados de manera apropiada.

2.2 Definición de métricas según ISO/IEC 25012

La ISO/IEC 25012 provee métricas de ejemplo para cada característica. Para cada característica, se utilizarán las funciones de medición contenidas en la norma. Se presenta una a modo de ejemplo:

- Tipo (inherente o dependiente del sistema)
- Nombre de la medida de calidad

- Función de medición
- Variables

En la tabla 1 se presenta como ejemplo la métrica para la característica Completitud.

Tabla 1. Definición de métrica para la Completitud

Tipo	**Inherente**
Nombre de la medida de calidad	Completitud de los datos en un archivo
Función de medición	Valor = A/B
Variables	A = número de datos requeridos para el contexto particular en el archivo B= número de datos en el contexto particular de uso previsto

2.3 Modelo de evaluación

La norma ISO/IEC 25040 define un modelo de referencia teniendo en cuenta entradas, restricciones y recursos necesarios para obtener las salidas deseadas. Para llevar a cabo la evaluación de un producto de software, se define un proceso de cinco actividades:

1. **Establecer los requisitos de la evaluación**: establecer el propósito de la evaluación y los requisitos de calidad del software a evaluar, identificando las partes interesadas en el producto, los riesgos (si existen) y el modelo de calidad a utilizar.
2. **Especificar la evaluación:** dentro de esta actividad se seleccionan los módulos de evaluación y se definen los criterios de decisión tanto para las métricas como para la evaluación.
3. **Diseñar la evaluación**: se define el plan con las actividades de evaluación que se deben realizar, teniendo en cuenta la disponibilidad de recursos.
4. **Ejecutar la evaluación**: en esta cuarta actividad se realizan las mediciones y se aplican los criterios de decisión tanto para las métricas como para la evaluación.
5. **Concluir la evaluación**: se cierra la evaluación de la calidad, realizando un informe con los resultados finales y las conclusiones de la evaluación en base a lo obtenido.

3 Gobierno y datos abiertos

El término de gobierno abierto se fue construyendo a través del tiempo. Sin embargo, fue popularizado recién en el 2009 por la Administración de Barack Obama en el “Memorandum on Transparency and Open Government” [6] en donde se proponen tres principios:

1. Un gobierno debe ser transparente. La transparencia en un gobierno puede ser alcanzada proveyendo a los ciudadanos con información acerca de lo que está haciendo, lo cual promueve la rendición de cuentas.

2. Un gobierno debe ser participativo. La participación pública en la formación de políticas y aportando conocimiento, ideas y experiencia mejora tanto la efectividad del gobierno como la calidad de sus decisiones.
3. Un gobierno debe ser colaborativo. La colaboración involucra activamente a los ciudadanos en el trabajo de su gobierno. Este principio requiere de cooperación entre individuos, empresas, asociaciones y agentes en todos los niveles de gobierno.

Se puede entender al gobierno abierto como una plataforma tecnológica ya que utiliza la tecnología de información y comunicación a su alcance para lograr su propósito. Esto incluye tanto sitios web como las redes sociales. [7]

3.1 Participación de los ciudadanos en el análisis de los datos abiertos

La participación de la ciudadanía es crucial en el concepto de gobierno abierto. Los gobiernos alrededor del mundo buscan aumentar la participación de los ciudadanos debido a su importancia para la toma de decisiones tanto en procesos políticos como administrativos. Para ello, es fundamental poner al alcance de la sociedad datos públicos en formatos abiertos. La revolución en las TIC (tecnologías de información y comunicación) cambió radicalmente la interacción entre gobiernos y ciudadanos.

La Web social o Web 2.0, que fomenta la participación de los individuos, permite alcanzar fácilmente el cumplimiento de los tres propósitos mencionados en el apartado anterior. Este paradigma se refiere al contenido que es creado, compartido y procesado por usuarios a través de medios de comunicación social además de la formación de una red social en la cual los mismos se conectan entre sí. [8]

Los gobiernos pueden expandir el alcance de los datos abiertos que publican a través del uso de estas redes sociales electrónicas. Para impulsar esto, es necesario crear comunidades en torno a los datos, que los consulten y analicen además de compartir nuevos datos. Además, un punto clave a tener en cuenta, es la redistribución de los mismos, a mayor velocidad de redistribución, mayor será el valor adquirido por los datos.

3.2 Datos abiertos: ventajas y desventajas

La publicación de datos abiertos en sus diferentes formas, como se mencionó previamente, permite a los ciudadanos usar y crear información a través de una red colaborativa. Con la misma surge una inteligencia colectiva en la que el público ayuda a mejorar la toma de decisiones participando activamente.

Además, al fomentar la publicación de datos abiertos, se aumenta la satisfacción del pueblo, ya que promueve igualdad de acceso a los datos, y se obtiene mayor confianza y transparencia. [9]

Sin embargo, hay ciertos factores que hacen que el uso de los datos y la participación ciudadana se vea perjudicada, debido a que no promueven la generación

de valor agregado. Otro punto para tener en cuenta es la falta de un análisis sistemático sobre qué debería ser publicado y qué esperan los usuarios de los datos abiertos, en muchos casos esa falta de análisis conlleva a publicar información incompleta, o incluso excesiva, lo que impide concentrarse en lo relevante. Por lo tanto, es importante asegurar la calidad de los datos para garantizar un análisis acertado.

4 Base de datos de COVID-19

La publicación de datos epidemiológicos abiertos junto con su análisis es un elemento fundamental ante el COVID-19. Es por lo que, para profundizar la transparencia durante esta pandemia, el ministerio de Salud de la Nación publicó distintos datasets relacionados a las tareas de lucha contra el coronavirus. En particular, en el siguiente trabajo se hizo foco en el recurso *"COVID-19. Casos registrados en la República Argentina"*, de actualización diaria, que se comenzó a publicar el 15 de mayo de 2020 y es mantenido por la Dirección Nacional de Epidemiología y Análisis de Situación de Salud.

El mismo cuenta con veinticinco campos, que se mencionan a continuación junto con sus tipos: número de caso (número entero), sexo (texto), edad (número entero), edad indicada en meses o años (texto), país de residencia (texto), provincia de residencia (texto), departamento de residencia (texto), provincia de establecimiento de carga (texto), fecha de inicio de síntomas (fecha según ISO-8601 - "Data elements and interchange formats — Information interchange — Representation of dates and times"), fecha de apertura del caso (fecha ISO-8601), semana epidemiológica de fecha de apertura (número entero), fecha de internación (fecha ISO-8601), indicación si estuvo en cuidado intensivo (texto), fecha de ingreso a cuidado intensivo en caso de corresponder (fecha ISO-8601), indicación de fallecido (texto), fecha de fallecimiento en caso de corresponder (tiempo ISO-8601), indicación si requirió asistencia respiratoria mecánica (texto), código de provincia de carga (número entero), origen de financiamiento (texto), clasificación manual del registro (texto), clasificación del caso (texto), código de provincia de residencia (número entero), fecha de diagnóstico (tiempo ISO-8601), código de departamento de residencia (número entero) y última actualización (fecha ISO-8601).

4.1 Evaluación de la base de datos COVID-19

Se realizará la evaluación teniendo en cuenta las actividades propuestas por la norma ISO/IEC 25040.

4.1.1 Propósito de la evaluación.

El propósito de la evaluación es: *analizar si los datos presentes en el recurso considerados como obligatorios están presentes, qué tan actualizada está la clasificación de los casos, si los campos cumplen con los formatos esperados, si están expresados correctamente y si los datos son consistentes entre sí.*

Para lograrlo, se evaluarán las características: ***Exactitud***, ***Completitud***, ***Consistencia***, ***Actualidad*** y ***Comprensibilidad***.

4.1.2 Especificación de la evaluación.

En esta actividad, además de definir los módulos de evaluación (presentado en la sección 2.2), se definen los criterios de decisión tanto para las características como para la evaluación final.

En la tabla 2 se presentan los rangos de valores para determinar el nivel de aceptación para las características: *exactitud, consistencia, completitud y comprensibilidad.*

En el caso de la característica *actualidad*, para determinar el nivel obtenido, se deben verificar tres campos en la base de datos: c*onfirmados, fallecidos, descartados* en la tabla 3 se pude ver los valores para determinar el nivel de cada uno de ellos y en la tabla 4 la combinación de estos tres campos para establecer el nivel alcanzado por la característica.

Tabla 2. Criterios de decisión para las características Exactitud, Consistencia, Completitud y Comprensibilidad.

Características	Nivel	Rango de valores
Exactitud	Inaceptable	Si Valor >= 0 y Valor < 0,3
Consistencia	Mínimamente aceptable	Si Valor >= 0,3 y Valor < 0,7
Completitud	Rango objetivo	Si Valor >= 0,7 y Valor < 0,9
Comprensibilidad	Excede los requerimientos	Si Valor >= 0,9

Tabla 3. Criterios de decisión para los campos del criterio actualidad.

Campos	Nivel	Rango de valores
Confirmados (C)	Inaceptable	Si Valor >= 0 y Valor < 0,3
Fallecidos (F)	Mínimamente aceptable	Si Valor >= 0,3 y Valor < 0,7
Descartados (D)	Rango objetivo	Si Valor >= 0,7 y Valor < 0,9
	Excede los requerimientos	Si Valor >= 0,9

Tabla 4. Criterios de decisión para la característica actualidad

Características	Nivel	Rango de valores
Actualidad	Inaceptable	C, D y F: Inaceptable
	Mínimamente aceptable	C: D y F: Mínimamente aceptable
	Rango objetivo	C, D y F: Rango Objetivo
	Excede los requerimientos	C, D y F: Excede los requerimientos

En la tabla 5 se muestran los criterios de decisión para la evaluación final, teniendo en cuenta los criterios de decisión para cada característica.

Tabla 5. Criterios de decisión para la Evaluación final.

Nivel Final	Valor de las características
Inaceptable	E, C, A y M: Inaceptable O: Mínimamente Aceptable
Mínimamente aceptable	E: Mínimamente aceptable C y M: Rango objetivo O y A: Mínimamente Aceptable
Rango Objetivo	E, O, A y M: Rango Objetivo C: Excede los requerimientos
Excede los requerimientos	E, C, O, A y M: Excede los requerimientos

Exactitud (E) Completitud (C) Consistencia (O) Actualidad (A) Comprensibilidad (M)

4.1.3 Diseño de la evaluación.

Se propone realizar una evaluación de la calidad de los datos desde la publicación del recurso (15 de mayo de 2020) hasta el 17 de febrero de 2021. Se evalúan las características definidas en la ISO/IEC 25012 acordes al propósito de la evaluación basada en la guía propuesta por la norma ISO/IEC 25040 y luego se calcula el promedio para determinar el valor final. Para realizar la evaluación se tendrán en cuenta las especificaciones que figuran en la tabla 6.

Tabla 6. Especificaciones de las características a evaluar

Característica	Especificaciones
Exactitud (Sintáctica + Semántica)	Para cada campo se tendrá en cuenta que coincida con el tipo de dato definido por el mantenedor de la base de datos, en el contexto acorde a la descripción. Número de Caso -> Número entero Sexo -> Texto (M o F) Edad -> Número Entero entre 0 y 116 Edad indicada en meses o años -> Texto (“años” o “meses”) País de Residencia -> Texto Provincia de Residencia -> Texto Departamento de Residencia -> Texto Provincia de establecimiento de carga -> Texto Fecha de inicio de síntomas -> Fecha ISO-8601 (date) Fecha de apertura del caso -> Fecha ISO-8601 (date) Semana Epidemiológica de fecha de apertura -> Número Entero Fecha de internación -> Fecha ISO-8601 (date) – si corresponde Indicación si estuvo en cuidado intensivo -> Texto (SI/NO) Fecha de ingreso a cuidado intensivo -> Fecha ISO-8601 (date) – si corresponde Indicación de fallecido -> Texto (SI/NO) Fecha de fallecimiento -> Tiempo ISO-8601 (time) – si corresponde Indicación si requirió asistencia respiratoria mecánica -> Texto

	(SI/NO) Código de Provincia de carga -> Número Entero Origen de financiamiento -> Texto (Público/Privado) Clasificación manual del registro -> Texto Clasificación del caso -> Texto (Descartado, confirmado o sospechoso). Código de Provincia de residencia -> Número Entero Fecha de diagnóstico -> Tiempo ISO-8601 (time) Código de Departamento de residencia -> Número Entero Última actualización -> Fecha ISO-8601 (date)
Completitud	En cada registro se evalúa que cada campo obligatorio esté presente
Consistencia	Los datos deben ser consistentes entre sí. Un ejemplo de inconsistencia a tener en cuenta es que una persona que no ha fallecido tenga fecha de fallecimiento.
Actualidad	La actualidad se mide en base a la diferencia entre el día de publicación del dataset y la fecha de diagnóstico (descartado/confirmado) o fecha de fallecimiento (fallecido). Para realizar la evaluación se considera actual aquella fecha de diagnóstico o de fallecimiento que tenga una demora de máximo 2 días con respecto a la fecha de emisión del recurso.
Comprensibilidad	Se evalúa que los campos estén expresados con una unidad o forma apropiada y deben ser legibles por el usuario.

4.1.4 Ejecución de la evaluación.

Al realizarse las mediciones se obtuvieron los valores para cada característica, y al aplicarse los criterios de decisión se obtuvieron los resultados reflejados en la tabla 7.

Tabla 7. Valores y resultados para cada característica

Característica	Valor	Resultado
Exactitud	1	Excede los requerimientos
Completitud	0.82	Rango objetivo
Consistencia	1	Excede los requerimientos
Actualidad	Confirmados = 0.75	Rango objetivo
	Descartados = 0.77	Rango objetivo
	Fallecidos = 0.33	Mínimamente aceptable
Comprensibilidad	0.92	Excede los requerimientos

El resultado más bajo, se relaciona con el campo *fallecidos* de la característica de *actualidad*, el promedio de cantidad de días de demora fue de 2.84 días para los *confirmados*, 9.65 días para los *descartados* y de 13.23 días para los *fallecidos*.

4.1.5 Finalización de la evaluación.

En la tabla 8 se presentan los resultados obtenidos para cada característica, aplicando el criterio definido para la evaluación final (tabla 5), si bien tres de las cinco características analizadas exceden los requerimientos, la actualidad de los fallecidos determinó que la característica *actualidad* sea mínimamente aceptable, por lo tanto, el mayor nivel que se puede otorgar a la evaluación final del propósito analizado es: **Mínimamente Aceptable**. Siendo este, un resultado satisfactorio por lo definido en la norma ISO/IEC 25040.

Tabla 8. Resultado final de la evaluación

Característica	Nivel Obtenido
Exactitud	Excede los requerimientos
Completitud	Rango Objetivo
Consistencia	Excede los requerimientos
Actualidad	Mínimamente aceptable
Comprensibilidad	Excede los requerimientos

En cuanto a la *comprensibilidad*, si bien excede los requerimientos, se notó que no existió diferencia en el uso de los tipos Fecha ISO-8601 (date) y Tiempo ISO-8601 (time), los campos con estos tipos mostraban datos con el mismo formato.

5 Conclusiones

Se presentaron dos normas de la familia ISO/IEC 25000 (SQuaRE). Se mostraron las quince características presentes en la ISO/IEC 25012, de las cuales se seleccionaron cinco para un contexto específico de uso y para las cuales se definieron las métricas basadas en las funciones de medición definidas en la norma.

Se definió el concepto de gobierno abierto, haciendo énfasis en los beneficios y las problemáticas de publicar datos abiertos además de resaltar la importancia de la participación de los habitantes en el uso de estos para la generación de valor y ayudando a tomar decisiones de manera eficaz.

Se realizó la evaluación de la calidad de datos a la base de datos “COVID-19. Casos registrados en la República Argentina”, siguiendo el modelo de evaluación propuesto por la norma ISO/IEC 25040. El resultado final fue “mínimamente aceptable” debido a que el valor obtenido para la actualidad de los fallecidos se encontraba dentro de este rango. Esto no quita que el resultado haya sido satisfactorio.

Cabe destacar, que, en una evaluación realizada diariamente bajo las mismas condiciones durante el mes de junio de 2021, los valores se mantuvieron en los mismos rangos, lo cual marca una estabilidad en la calidad de los datos.

6 Agradecimientos

Esta publicación fue realizada en el contexto del Proyecto CAP4CITY – "Strengthening Governance Capacity for Smart Sustainable Cities" (www.cap4city.eu) co-financiado por el Programa Erasmus+ de la Unión Europea. Acuerdo Número 598273-EPP-1- 2018-1-AT-EPPKA2-CBHE-JP. Número de proyecto: 598273

7 Referencias

[1] O. Morgan, "How decision makers can use quantitative approaches to guide outbreak responses," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 374, no. 1776, 2019, doi: 10.1098/rstb.2018.0365.

[2] H. Chen, D. Hailey, N. Wang, and P. Yu, "A review of data quality assessment methods for public health information systems," *Int. J. Environ. Res. Public Health*, vol. 11, no. 5, pp. 5170–5207, 2014, doi: 10.3390/ijerph110505170.

[3] ISO/IEC 25012:2008, "Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model."

[4] ISO/IEC 25040:2011, "Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Evaluation process."

[5] "COVID-19. Casos registrados en la República Argentina" [Online]. Available: https://datos.gob.ar/. - http://datos.salud.gob.ar/

[6] O. Government, "Memorandum on Transparency and Open Government," *Fed. Regist.*, pp. 21–22, 2009, [Online]. Available: https://www.archives.gov/files/cui/documents/2009-WH-memo-on-transparency-and-open-government.pdf.

[7] R. Sandoval-Almazán, "Gobierno abierto y transparencia: Construyendo un marco conceptual," *Convergencia*, vol. 22, no. 68, pp. 203–227, 2015, doi: 10.29101/crcs.v0i68.2958.

[8] S. A. Chun, S. Shulman, R. Sandoval, and E. Hovy, "Government 2.0: Making connections between citizens, data and government," *Inf. Polity*, vol. 15, no. 1–2, pp. 1–9, 2010, doi: 10.3233/IP-2010-0205.

[9] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, Adoption Barriers and Myths of Open Data and Open Government," *Inf. Syst. Manag.*, vol. 29, no. 4, pp. 258–268, 2012.

ISBN 978–987–633–574–4

Propuesta para la construcción de un Corpus Jurídico utilizando Expresiones Regulares

Osvaldo Spositto[1], Ryckeboer Hugo[1], Viviana Ledesma[1], Gastón Procopio[1], Lorena Matteo[1], Cecilia Gargano[1], Julio Bossero[1], Edgardo Moreno[1], Victoria Saizar[1], Patricio Macias[1], Juan Ojeda[1], Fabio Quintana[1], Laura Conti[2], Sergio García[3] y Gustavo Pérez Villar[4]

[1] Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigación Tecnológicas. Florencio Varela 1903. San Justo. La Matanza. {spositto, hugor, vledesma, gprocopio, lmatteo, cgargano, jbossero, ej_moreno, vsaizar, pmacias, jmojeda}@unlam.edu.ar
[2] Universidad Nacional de La Matanza. Departamento Derecho y Ciencia Política. lconti@unlam.edu.ar
[3] Palacio de Tribunales. Departamento Judicial de Morón. Alte. Brown. Piso 4. Morón. sergiogabriel.garcia@pjba.gov.ar
[4] Subsecretaría de Tecnología Informática del Poder Judicial de la Provincia de Buenos Aires. Palacio de Justicia, avenida 13 entre 47 y 48, primer piso (La Plata). Argentina. gperez@scba.gov.ar

Abstract. En la última década, la construcción de corpus de distintas especialidades ha tenido un amplio desarrollo, debido en gran parte, por su incorporación en el proceso de recuperación de la información. Si bien, en el sistema jurídico argentino, existen varios buscadores de expedientes digitales, en este artículo se presenta una propuesta para incorporar, en un corpus jurídico, las fechas y las referencias de la norma jurídica, mediante el Reconocimiento de Entidades Nombradas (tales como Acordadas, Artículos, Leyes, entre otros), que componen los distintos documentos judiciales, utilizando Expresiones Regulares (ER). Estas cadenas de caracteres se utilizan para describir o encontrar patrones dentro de otros textos, empleando delimitadores y reglas de sintaxis. Se propone una metodología que permita identificar, clasificar y reemplazar estas entradas automáticamente, con el objetivo de ser normalizas. Por último, se presenta una propuesta para incorporar en un algoritmo de Lematización, la codificacion del proceso mencionado usando ER.

Keywords: Corpus, Expresiones Regulares, Sistema de Recuperación de Documento, Lematización, Reconocimiento de Entidades Nombradas

1 Introdución

Este trabajo, continúa con la línea de investigación y trabajo interdisciplinario entre especialistas del área jurídica provincial, técnicos de la Corte Suprema de la Provincia

de Buenos Aires e Investigadores de la Universidad Nacional de La Matanza (UNLaM). En el año 2020, el grupo abordó el análisis, diseño y construcción de una herramienta informática que ayuda a la sistematización y optimización de varios de los procesos judiciales que actualmente se realizan en forma manual o semiautomática en los juzgados de la provincia. La herramienta desarrollada, que se denomina *Experticia*[1] [1-2], pretende dar soporte a los operadores de la justicia en su decisión para la resolución de una causa. De esta manera se busca estandarizar el proceso de despacho de trámites, y a la vez agilizar y reducir los tiempos de carga, minimizando posibles errores como en el ingreso de datos. La información generada con Experticia, se almacena en el Sistema Informático de Gestión Asistida Multifuero (GAM), más conocido en el poder judicial como Augusta[2]. Este aplicativo fue creado con la finalidad de dotar al Poder Judicial de la Provincia de Buenos Aires, de una plataforma informática única e integral, que permita homogeneizar la gestión administrativa diaria de las causas. En el campo del derecho, la jurisprudencia tiene un papel importante como fuente de derecho; porque sus conclusiones apoyan la aplicación de la ley en un caso específico. El poder judicial argentino produce una gran cantidad de dictámenes, expedientes, etc. cada año, estas decisiones se guardan en documentos, haciendo que esta fuente de derecho sea cada vez mayor, lo que impulsa a los profesionales del derecho a dedicar más tiempo a la búsqueda de documentos relevantes. Por lo tanto, se necesitan técnicas sofisticadas de cómputo para minimizar el tiempo de búsqueda y mejorar la pertinencia de los documentos recuperados. Por este motivo, en el año 2021 se presentó el proyecto de investigación "*Implementación de un Sistema Web de Recuperación de la Información Orientado a Documentación Jurídica con el Proceso de Indexación Semántica Latente Paralelizado*", por el Programa de Incentivos para Docentes Investigadores de la Secretaría de Políticas Universitarias (PROINCE). Dentro de las etapas para llevar adelante este trabajo, se encuentra la construcción de un corpus jurídico. Varias investigaciones se centraron en remarcar la importancia que tiene la lingüística de corpus como herramienta de ayuda para analizar terminología y fraseología especializada en su contexto original de producción. Hoy, gran parte de los corpus, se compilan a partir de textos electrónicos y la web se ha convertido en una gran fuente de contenidos textuales de todo tipo [3,4].

Un Sistema de Recuperación de Información (SRI) [5-7] es una herramienta que interactúa entre un corpus y sus usuarios. Su efectividad depende del adecuado control del lenguaje de representación de los elementos de información y las búsquedas de sus usuarios. Para cumplir con sus objetivos, según Gabriel H. Tolosa y otros [6], un SRI debe realizar las siguientes tareas básicas:

- Representación lógica de los documentos y – opcionalmente – almacenamiento del original.
- Representación de la necesidad de información del usuario en forma de consulta.
- Evaluación de los documentos respecto de una consulta para establecer la relevancia de cada uno.

[1] https://noficcionweb.com.ar/la-suprema-corte-bonaerense-y-la-unlam-avanzan-en-la-automatizacion-de-la-justicia/

[2] https://www.scba.gov.ar/paginas.asp?id=39889

- Ranking de los documentos considerados relevantes para formar el "conjunto solución" o respuesta.
- Presentación de la respuesta al usuario.
- Retroalimentación de las consultas (para aumentar la calidad de la respuesta).

La arquitectura de un SRI que permite realizar las tareas básicas enumeradas en el párrafo anterior se puede observar en la Figura 1:

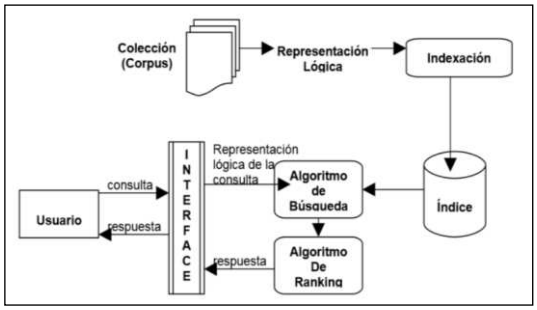


Fig. 1 – Arquitectura de un SRI. Fuente [6]

Como se puede apreciar en la Figura 1, el conjunto de todos los documentos sobre los que se deben realizar operaciones de RI se denomina corpus, colección de documentos o base de datos documental. El proceso de indexación genera la representación lógica de los documentos y las estructuras de datos denominadas índices, estas estructuras son las que permiten que se realicen búsquedas eficientes. El algoritmo de búsqueda se encarga de procesar la consulta de un usuario y de buscar en el índice cuáles documentos se asemejan a la consulta. A continuación, el algoritmo de ranking determina la relevancia de cada documento de acuerdo al nivel de semejanza y retorna un subconjunto con los documentos más relevantes. La interface de usuario permite que éste especifique la consulta, visualice la respuesta y realimente el sistema para mejorar la calidad de las respuestas.

Uno de los principales procesos de un SRI es la indexación y según Tolosa en [5] se puede dividir en las siguientes etapas:

- Análisis lexicográfico: Se extraen las palabras y se normalizan.
- Reducción (Tokenización) de palabras vacías o de alta frecuencia.
- Lematización: Se reducen palabras morfológicamente parecidas a una forma base o raíz, con la finalidad de aumentar la eficiencia de un SRI.
- Selección de los términos a indexar: Se extraen aquellas palabras simples o compuestas que mejor representan el contenido de los documentos.
- Asignación de pesos o ponderación de los términos que componen los índices de cada documento.

Si bien, estos corpus contienen información del mismo dominio, esta es habitualmente del tipo textual. En un expediente judicial, se pueden encontrar, además, referencias de fechas, en distintos formatos, como asi también, referencias a diferentes fuentes judiciales[3], como se puede ver en el párrafo siguiente: *"...de la Ley Nº 25.188, o el Decreto 41/99, o la Ley Nº 25.164 –que rige únicamente para el*

[3] https://www.conicet.gov.ar/wp-content/uploads/Ley-25164-De-Marco-de-Regulación-de-Empleo-Público-Nacional.pdf

personal……su función (artículo 3º de la Ley Nº 25.188; art. 47 del Decreto 41/99 y art. 30 de la Ley Nº 25.164…"

Los usuarios de distintos ecosistemas, que utilizan corpus "*ad hoc*", demandan cada vez más servicios, que les permitan extraer información recuperada, usando reconocimiento y categorización de Entidades Nombradas (EN o NE del inglés Named Entity) de fácil integración en aplicaciones del Procesamiento del Lenguaje Natural (PLN) [8]. En este escrito, se presenta una propuesta, que se centra, en la detección, clasificación y normalización de fechas y entidades nombradas (como Acordadas, Artículos, Leyes, Resoluciones o Decretos, etc.) que componen la normativa jurídica, mediante el uso de Expresiones Regulares (ER). La idea es poder incorporar esta información al corpus, en el proceso de la Lematización de los documentos. Esta es una de las etapas de Preprocesamiento en un SRI [6]. Por su parte la técnica de Reconocimiento de EN (REN) se divide generalmente en dos pasos [8]: la delimitación de entidades nombradas y su posterior clasificación. En este trabajo solo nos enfocamos en la primera. Esta propuesta podría incrementar la eficacia en la equiparación entre los términos del documento y los términos de la pregunta del usuario.

2 Trabajos relacionados

Se han desarrollado muchos trabajos relacionados a la temática en cuestión. Diversas propuestas han sido consideradas para la construcción de Corpus jurídicos [9-10], en este último artículo, "*El uso de corpus electrónicos para la investigación de terminología jurídica*", se encuentra una extensa lista de los corpus disponibles en Argentina y una descripción detallada de mas de 10 corpus multilingües internacionales. Respecto a los trabajos sobre construcción de corpora utilizando Expresiones Regulares para resolver las Entidades Nombradas, tenemos el trabajo desarrollado por Karen Haag, en su tesis: "*Reconocimiento de entidades nombradas en texto de dominio legal"* [8], el escrito se centra en la detección, clasificación y anotación de entidades nombradas (como Leyes, Resoluciones o Decretos, entre otros) para el corpus de *InfoLEG*, una base de datos que contiene los documentos de todas las leyes de la República Argentina. Además, se pueden mencionar, entre otros, el trabajo de Cristian Cardellino [11] "*A Low-cost, High-coverage Legal Named Entity"*. En este documento, se intenta mejorar la extracción de información en textos legales mediante la creación de un reconocedor, clasificador y vinculador de entidad con nombre legal. Otro trabajo que merece ser nombrado se encuentra en el capitulo segundo: "*Regular Expressions, Text Normalization, Edit Distance*" del libro de D. Jurafsky y J. H. Martin: "*Speech and Language Processing*" [12], donde se presenta una herramienta para realizar tareas básicas de normalización de texto que incluyen segmentación y normalización de palabras, segmentación de oraciones y derivación. Por último, se puede nombrar, además, el trabajo de Robaldo, Livio y otros: "*Compiling Regular Expressions to Extract Legal Modifications*", que presenta un prototipo para identificar y clasificar automáticamente tipos de modificaciones en el texto legal italiano [13].

3 Expresiones Regulares

Uno de los éxitos no reconocidos en la estandarización de la informática ha sido la utilización de ER, un lenguaje para especificar cadenas de búsqueda de texto [12]. Este lenguaje práctico se usa en todos los lenguajes de computadora, procesadores de texto y herramientas de procesamiento de texto como las herramientas Unix grep[4] o Emacs[5]. Formalmente, una expresión regular es una notación algebraica para caracterizar un conjunto de cadenas.

Son particularmente útiles para la búsqueda en textos, cuando tenemos un patrón y un corpus de textos donde buscar. Una función de búsqueda de expresiones regulares buscará en el corpus y devolverá todos los textos que coincidan con el patrón. El corpus puede ser un solo documento o una colección. Por ejemplo, la herramienta de línea de comandos de Unix grep toma una expresión y devuelve cada línea del documento de entrada que coincide con la expresión. En otras palabras, son notaciones simbólicas que se utilizan para identificar caracteres mediante una secuencia en el texto. En cierto modo, se parecen al método comodín del comando de Linux "Shell" para hacer coincidir los nombres de archivo y ruta, pero a una escala mucho mayor. Una expresión regular es un patrón capaz de reconocer o filtrar cadenas de caracteres según ciertos criterios. El uso de comodines "*" para indicar cadenas de caracteres cualesquiera o "?" para indicar un carácter único son ejemplos de uso de expresiones regulares. Así, el patrón "aba*" reconoce cadenas como "abaco", "abajo", "abatimiento", "abalorio", "aba-23"; el patrón "do?" reconoce cadenas como "doy", "dos", "dot", "don", "do$"; el patrón "aba*-txt" describe el conjunto de cadenas de caracteres que comienzan con "aba", contienen cualquier otro grupo de caracteres y luego la cadena "-txt". Los patrones construidos como ER que permiten reconocer cadenas de caracteres de estructura compleja. Las ER son utilizadas para realizar búsquedas o sustituciones en textos [14]. Estas son reconocidas por muchos lenguajes de programación, editores y otras herramientas. Su nombre proviene de la teoría matemática en la que se basan.

3.1 Expresiones Regulares básicas

Una ER determina un conjunto de cadenas de caracteres. Un miembro de este conjunto de cadenas se dice que aparea, equipara o satisface la expresión regular.

Con la idea de mostrar unos ejemplos, en la tabla 1, se pueden ver las ER que componen el conjunto de ER Elementales que aparean con un único carácter [14], en este mismo documento, se encuentra un tutorial del tema.

Tabla 1. Resumen de las ER Elementales que aparean con un único caracter [14].

Expresión	Aparea con
c	ER que aparea con el caracter ordinario c
.	(punto) aparea con un caracter cualquiera excepto nueva línea
[abc]	ER de un caracter que aparea con uno de a, b, c

[4] https://www.gnu.org/software/grep/
[5] http://www.gnu.org/software/emacs/

[^abc]	ER de un caracter que no sea uno de a, b, c
[0-9][a-z] [A-Z]	ERs de un caracter que aparean con cualquier caracter en el intervalo indicado El signo "-" indica un intervalo de caracteres consecutivos
\e	ER que aparea con alguno de estos caracteres (en lugar de la e): . * [\ cuando no están dentro de [ ] ^ al principio de la ER, o al principio dentro de [ ] $ al final de una ER / usado para delimitar una ER

Por lo general, se encontrará el nombre abreviado como "Regex" o "Regexp". En un editor de texto como EditPad Pro[6] o una herramienta de procesamiento de texto especializada como PowerGREP[7], puede usar la expresión regular como la siguiente:

«\b[A-Z0-9 ._%+-] + @ [A-Z0-9 .-] + \ . [AZ] {2,4} \ b » (1)

para buscar una dirección de correo electrónico. Cualquier dirección de correo electrónico, para ser exactos.

4 Reconocimiento de Entidades Nombradas

Encontramos en [15] una definición sobre el término Entidad Nombrada "*...es una palabra o secuencias de palabras que se identifican como nombre de persona, organización, lugar, fecha, tiempo, porcentaje o cantidad.*". El REN, tiene como objetivo el reconocer y clasificar nombres de personas, lugares, organizaciones o cantidades, en distintas aplicaciones del Procesamiento del Lenguaje Natural. A partir de la bibliografía consultada [8-11]. En estos trabajos se muestran distintos usos de ER para detectar patrones dentro del texto de un documento.

En el área del REN, un problema común es obtener información relevante relacionada con nombres de personas, lugares u organizaciones, por lo cual se vuelve importante el poder extraer y distinguir este tipo de elementos de todo el conjunto de palabras que componen a un documento. Aún cuando algunos elementos son relativamente fáciles de identificar, mediante el uso de patrones (por ejemplo: fechas o datos numéricos) existen otros elementos, como personas, lugares u organizaciones, que presentan otras dificultades para ser identificados como pertenecientes a un tipo específico. En un SRI, una técnica como el REN, es muy importante, ya que permite buscar información muy concreta en colecciones de documentos, extrayendo y organizando la información relevante [15]. En el trabajo de Sánchez Pérez, se menciona que en los últimos años se ha trabajado ampliamente en el desarrollo de sistemas de REN para mejorar el desempeño de clasificadores utilizando técnicas de aprendizaje automático.

[6] https://www.editpadpro.com/
[7] https://www.powergrep.com/

5 Bases metodológicas

Con el propósito de elaborar una propuesta orientada a la incorporación de datos en formato fecha y de entidades nombradas (Leyes, Acordadas, Decretos, Artículos, etc.), usando ER, en un corpus, utilizamos un texto público de Pedido de Libertad Condicional (PLC), disponible en [16]. En este documento se pueden observar cómo aparecen las referencias mencionadas en el texto (Fig.2).

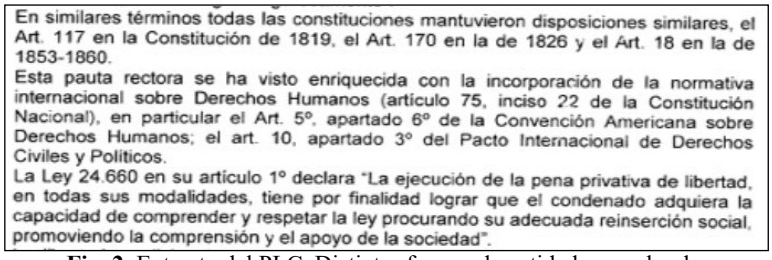

En similares términos todas las constituciones mantuvieron disposiciones similares, el Art. 117 en la Constitución de 1819, el Art. 170 en la de 1826 y el Art. 18 en la de 1853-1860.
Esta pauta rectora se ha visto enriquecida con la incorporación de la normativa internacional sobre Derechos Humanos (artículo 75, inciso 22 de la Constitución Nacional), en particular el Art. 5º, apartado 6º de la Convención Americana sobre Derechos Humanos; el art. 10, apartado 3º del Pacto Internacional de Derechos Civiles y Políticos.
La Ley 24.660 en su artículo 1º declara "La ejecución de la pena privativa de libertad, en todas sus modalidades, tiene por finalidad lograr que el condenado adquiera la capacidad de comprender y respetar la ley procurando su adecuada reinserción social, promoviendo la comprensión y el apoyo de la sociedad".

Fig. 2. Extracto del PLC. Distintas formas de entidades nombradas.

En el documento, usado de ejemplo, de casi nueve carillas, se puede contabilizar la cantidad de veces y distintos formatos, en que se encuentran estas referencias.

Tabla 2. Resumen de las EN y los formatos de fechas que figuran en el PLC.

Refencia	Ejemplo del texto que aparece	Cantidad de veces
Art. XX	Art 14	11
arts. XX y XY	arts 14 y 17	14
artículo XX	artículo 5	4
artículo XX Inciso	artículo 75 inciso 22	4
Ley XXXX	Ley 26.660	6
Fecha formato (dd/mm/aaaa)	24/11/1993	4
Fecha formato (dd de mes de aaaa)	27 de octubre de 2006	2
Fecha formato AAAA	*1993*	3

En coincidencia con [8] y también, en base a un análisis exploratorio del PLC, el patrón de REN más común, se encuentra en la siguiente forma:

< Tipo Entidad > [Nro] < Número > [/ < Año >] **(2)**

Dónde el "Tipo Entidad" es una parte de las categorías nombradas.

En la construcción de un corpus como el propuesto, para este trabajo, un problema común es obtener información relevante relacionada con todos los nombres de la normativa a normalizar, por lo cual se vuelve importante el poder extraer y distinguir este tipo de elementos de todo el conjunto de palabras que componen a un documento.

En el trabajo de Karen Haag [8], se desarrollan todas las entidades nombradas que se utilizan en el poder judicial. Algunos elementos son relativamente fáciles de identificar, mediante el uso de patrones (por ejemplo: fechas o datos numéricos). Existen muchas aplicaciones [17-18] que ayudan a convertir distintos formatos de fecha en ER. A continuación, se muestra una lista de algunas de las variantes que se pueden encontrar en este conjunto de datos:

- 20/04/2009; 20/04/09; 20/4/09; 3/04/09

- 20 de marzo de 2009; 20 de mar de 2009
- Febrero de 2009; septiembre del 09; octubre 2010
- 6/2008; 12/09 o 2009

A continuación, se muestra una colección de ER útiles para encontrar fechas:

- **Formato (dd/mm/aa o aaaa o dd-mm-aa o aaaa)**
  RegEx1: [0-9]{1,2}[\\/-][0-9]{1,2}[\\/-][0-9]{2,4} o **(2)**
  RegEx2: \d{1,2}[\\/-]\d{1,2}[\\/-]\d{2,4} **(3)**
- **Formato 'Mes, dd, aaaa', Por ejemplo, '4 de julio de 2021'.**
  (Ene(?:ro)?|Feb(?:rero)?|Mar(?:zo)?|Abr(?:il)?|May|Jun(?:io)?|Jul(?:io)?|Agost(?:o)?|Sep(?:tiembre)?|Oct(?:ubre)?|Nov(?:iembre)?|Dic(?:ciembre)?)\s+(\d{1,2})\,\s+(\d{4}) **(4)**

6 Lematización

En los SRI, la lematización (Stemming en Inglés) es una técnica empleada en la recuperación de datos en los SRI, que sirve para reducir variantes morfológicas de la forma de una palabra a raíces comunes o lexemas; con el fin de mejorar la habilidad de los motores de búsqueda para mejorar las consultas en documentos. Básicamente, este consiste en remover el plural, el tiempo, o los atributos finales de las palabras [19,20]. Cuando se realiza la extracción de palabras de un texto se obtiene una gran cantidad de entradas con formas verbales conjugadas y variantes de concordancia. Logrando la reducción morfológica de todas estas variantes se busca que el usuario recupere tanto los textos que contienen sus términos de búsqueda, como aquellos que contienen las formas derivadas de esos términos. Los algoritmos de lematización más conocidos son: Lovins[8](1968), Porter[9] (1980) y Paice[10] (1990). La descripción y comparación de estos y otros algoritmos menos conocidos, se encuentran desarrollados en el trabajo "*Comparative Study of Truncating and Statistical Stemming Algorithms"* en [21]. Todos eliminan "los finales" de las palabras en forma iterativa, y requieren de una serie de pasos para llegar a la raíz, pero no requieren "a priori" conocer todas las posibles terminaciones. Originalmente todos fueron hechos para el inglés, y se diferencian en la eficiencia del código y la elección de sufijos que identifican y eliminan. Esto es solo un ejemplo de la forma en que operan estos algoritmos. El trabajo de Porter[11], fue tomado como base por muchos investigadores [22]. El algoritmo[12] sirve para reducir variantes morfológicas de las formas de una palabra a raíces comunes o lexemas; mediante una sucesión de reglas que aplica sobre cada palabra. En esta memoria se presenta una codificación utilizando la librería Regex de .Net[13]. El ejemplo utiliza el método de *Regex.Replace* para reemplazar fechas con el formato mm/dd/aa por fechas con el formato dd-mm-aa.

[8] http://snowball.tartarus.org/algorithms/lovins/stemmer.html
[9] https://tartarus.org/martin/PorterStemmer/
[10] https://www.scientificpsychic.com/paice/paice.html
[11] https://tartarus.org/martin/index.html
[12] https://tartarus.org/martin/PorterStemmer/def.txt
[13]https://docs.microsoft.com/es-es/dotnet/standard/base-types/regular-expressions?redirectedfrom=MSDN

```
using System;
using System.Globalization;
using System.Text.RegularExpressions;
public class Class1
{   public static void Main()
   { string dateString =
DateTime.Today.ToString("d",
DateTimeFormatInfo.InvariantInfo);
      string resultString = MDYToDMY(dateString);
      Console.WriteLine("Converted {0} to {1}.",
dateString, resultString);    }
   static string MDYToDMY(string input)
   { try { return Regex.Replace(input,
     "\\b(?\\d{1,2})/(?\\d{1,2})/(?\\d{2,4})\\b",
           "${day}-${month}-${year}",
RegexOptions.None,
             TimeSpan.FromMilliseconds(150));     }
      catch (RegexMatchTimeoutException) {  return
input;}} }
```

7 Conclusiones y Trabajo Futuro

En este trabajo se propuso implementar, en un algoritmo de lematización, el uso de Expresiones Regulares para incorporar fechas y Entidades Nombradas a un corpus jurídico, para luego ser empleado en un Sistema de Recuperación de Información. Se estudiaron las expresiones regulares, que proporcionan un método eficaz y flexible para procesar texto.

Dentro de las tareas a desarrollar se puede mencionar:

- Incorporar la codificación propuesta al SRI implementado por el proyecto PROINCE mencionado en la introducción.
- Utilizar el algoritmo de Porter y analizar otros Lematizadores.
- Estudiar otras librerías existentes de ER.
- Realizar una clasificación de todas las EN dentro de la norma jurídica Argentina.

Referencias

1. Sposito O. y otros. Sistema Experto para Apoyo del Proceso de Despacho de Trámites de un Organismo Judicial. Jornadas Argentinas de Informática (JAIIO 2020).
2. Sposito O. y otros. Metodológica para evaluar un modelo de Justicia Predictiva". Trabajo presentado en CONAIISI 2020.
3. Capello, A. Sistema de recomendación para textos legales. (2018) En Línea: http://hdl.handle.net/11086/11342 Fecha de consulta: 25/6/21

4. Moreno A. Internet como fuente para la compilación de corpus jurídicos (2013) CES Felipe II (UCM) En línea: http://www.cesfelipesegundo.com/revista/Articulos2013/Art%C3%ADculoArsenioAndrade.pdf Fecha de consulta: 25/6/21
5. Kuna, H., Rey, M., Martini, E., Solonezen, L. & Podkowa, L. Desarrollo de un Sistema de Recuperación de Información para Publicaciones Científicas del Área de Ciencias de la Computación, Revista Latinoamericana de Ingeniería de Software, (2014). 2(2): 107-114.
6. Tolosa G. & Bordignon, F. Introducción a la Recuperación de Información: Conceptos, modelos y algoritmos básicos. Universidad Nacional de Luján, Argentina, (2008). En línea: http://eprints.rclis.org/12243/1/Introduccion-RI-v9f.pdf. Fecha de consulta: 25/6/21
7. González, C. M. La recuperación de información en el siglo XX. Revisión y aplicación de aspectos de la lingüística cuantitativa y la modeliza-ción matemática de la información U. N. de La Plata. (2008) Disponible en: http:// www.fuentesmemoria.fahce.unlp.edu.ar/tesis/te.350/te.350.pdf. Fecha de consulta: 25/6/21
8. Karen Haag. Reconocimiento de entidades nombradas en texto de dominio legal. Cordoba, Argentina (2019). Recuperado el 01/08/2019 de: https://rdu.unc.edu.ar/handle/11086/15323
9. Cucatto M, El lenguaje jurídico y su desconexión con el lector especialista: El caso de a mayor abundamiento. Letras de Hoje, 48 (1), 127-138. (2013). En Memoria Académica. Disponible en: http://www.memoria.fahce.unlp.edu.ar/art_revistas/pr.9102/pr.9102.pdf Fecha de consulta: 25/6/21
10. El uso de corpus electrónicos para la investigación de terminología jurídica. Disponible en: http://www.bibliotecact.com.ar/PDF/08118.pdf. Fecha de consulta: 25/6/21
11. Cardellino C. y otros. A Low-cost, High-coverage Legal Named Entity. (2017) En: https://hal.archives-ouvertes.fr/hal-01541446/document. Fecha de consulta: 25/6/21
12. Jurafsky, D. & Martin, J. Speech and Language Processing. (2020) En línea:. https://web.stanford.edu/~jurafsky/slp3/2.pdf. Fecha de consulta: 25/6/21
13. Robaldo, L. y otros. Compiling regular expressions to extract legal modifications. 250. 133-141. 10.3233/978-1-61499-167-0-133. (2012).
14. William Shotts. The Linux Command Line. (Third Internet Edition). A LinuxCommand.org Book. (2016). En línea: https://filedn.com/liGIo7rEUfzfmU4MQdhIKrh/Cursos/CursoBasicoLinux/ExpresionesRegulares.pdf. Fecha de consulta: 25/6/21
15. Sánchez Pérez C. Clasificación de Entidades Nombradas utilizando Información Global. (2008). En línea: https://inaoe.repositorioinstitucional.mx/jspui/bitstream/1009/564/1/SanchezPCR.pdf. Fecha de consulta: 25/6/21
16. Revista Pensamiento Penal. http://www.pensamientopenal.com.ar/system/files/2016/06/miscelaneas43506.pdf#viewer.action=download. Fecha de consulta: 25/6/21
17. https://regex101.com/
18. https://www.regextester.com/
19 Martínez Méndez, F. Recuperación de información: modelos, sistemas y evaluación. Disponible en: https://digitum.um.es/digitum/bitstream/10201/4316/1/libro-ri.pdf. (2004) Último acceso: 20/07/2021.
20 Herrero Pascual, Cristina. (2010). Manual de indización: teoría y práctica. Investigación bibliotecológica, 24(52), 239-240. http://www.scielo.org.mx/scielo.php ?script=sci_arttext&pid=S0187-358X2010000300010&lng=es&tlng=es. Último acceso: 20/07/2021.
21 Figuerola C. y otros (2000) Diseño de un motor de recuperación de la información para uso experimental y educativo. Univ. de Salamanca. Disponible en: https://dialnet.unirioja.es/servlet/articulo?codigo=5555288. Último acceso: 20/07/2021.
22 Bordignon F., W. Panessi. Procesamiento de variantes morfológicas en búsquedas de textos en castellano. Revista Interamericana de Bibliotecología, ISSN 0120-0976, Vol. 24, Nº. 1 (ENE-JUN), 2001, págs. 69-88. https://dialnet.unirioja.es/servlet/articulo?codigo=4291340. Último acceso: 20/07/2021.

ISBN 978–987–633–574–4

Control de tránsito en una Smart City

Juan Pablo Murdolo[1], Marcelo Taruschio[2], Rodolfo Bertone[3],

[1] Alumno de Grado FACEI UCALP
[2] Profesor Titular, FACEI UCALP
[3] Director de Carrera FACEI UCALP
jpmurdolo@gmail.com, mtarus2012@gmail.com, rodolfo.bertone@ucalp.edu.ar

Abstract. Este trabajo surge a partir de la siguiente premisa ¿cómo resolver los problemas de tránsito de una ciudad como La Plata aplicando tecnología? Teniendo varios factores en cuenta, se plantean soluciones novedosas a problemas conocidos. Se aplican conceptos ligados a GPS, sensores y placas Arduino e Internet, así como la posibilidad de contar con información en línea de un número de usuarios considerables, lo que permitiría organizar al tránsito en forma automática, a partir de la utilización de la tecnología.
El proyecto plantea generar una aplicación de control automático de semáforos de manera de hacer el tránsito más eficiente. Así, los stakeholders relacionados optimizarán el tiempo requerido para el traslado, se reducirán la emisiones y, en general, se mejorará las condiciones de vida. Además, se propone una solución donde se utiliza hardware (Arduino y sensores) cuyo valor económico es acotado, haciendo el proyecto viable para cualquier municipio.

Keywords: SmartCity, Sensores, Arduino, Aplicaciones, Automatización

1 Introducción

En control del tránsito en las ciudades de una forma eficiente, rápida, segura y principalmente en tiempo real es uno de los objetivos que se presentan en las políticas de servicios públicos municipales. El tiempo perdido, la contaminación generada (auditiva y polución) y las perdidas económicas producidas a partir de una situación caótica en el tránsito, pueden ser evitadas o al menos morigeradas a partir del uso de aplicaciones que adquieran información en tiempo real y tomen decisiones que permita agilizar el movimiento automotor en una zona determinada.

Por ejemplo, la empresa Here Mobility, con sede en la ciudad holandesa de Amsterdam, propone la idea de Smart Traffic Management. Este sistema consiste en la utilización de pequeñas luces, sensores y detectores integrados a la infraestructura. La empresa propone un sistema utilizado solo para regular el tránsito. Para ello, utiliza los sensores y las señales de tránsito para monitorear, controlar y responder a las condiciones de tráfico. Los objetivos que se proponen son: reducir la congestión día a día, priorizar el tráfico en base a condiciones reales del mismo, reducir la

polución, dar prioridad a los micros que ingresan a las intersecciones y usar semáforos para garantizar el flujo de buses a través de la ciudad, y, además, mejorar el tiempo de respuesta en caso de accidentes de tránsito. Su método de implementación es mediante, un sistema de control centralizado, la colocación de semáforos smart y cámaras. La aplicación de big data e IoT hace que el proyecto a gran escala sea una excelente opción para ciudades de gran envergadura como New York [1].

Uno de los proyectos obtenidos de la sección smartcity de la página de Viena del Gobierno de Austria, implementa cuatro ideas sobre semáforos que "piensan y se comunican". Estas ideas son:

- algoritmos para reconocer el deseo de cruzar de los peatones
- luces de tráfico interconectadas,
- integración con sistemas de navegación
- y sensores de clima y ambiente.

En septiembre de 2019 las "luces de tráfico inteligente" reemplazaron aproximadamente 200 semáforos con botones de cruce. El algoritmo planteado fue desarrollado por TU Graz. consta de una cámara que detecta no solo el paso de personas, sino que, además, registra si esa persona tiene o no el deseo de cruzar. La conexión con sistemas de navegación indica que los semáforos no deben comunicarse entre sí, sino que deben pasar información respecto al tráfico a dispositivos de navegación y smartphones.

En ese proyecto y en una etapa posterior, se pretende colocar en los semáforos sensores de clima y ambiente. Estos sensores permitirán obtener datos estadísticos que permitan lograr una mejora en la calidad del aire de la ciudad.[2]

La revista Times en una presentación de enero de 2019 plantea la evolución, o quizá la "involución", de los semáforos a lo largo del tiempo. En el trabajo se remarca que sólo el 3% de los semáforos son considerados smart en Estados Unidos. Entre otros datos presentados, se indica que algunos sistemas son anticuados, porque no tienen en cuenta a los peatones y ciclistas. Además, muchos de los sensores se encuentran debajo de las calles y se ven afectados por el clima.[3]

La revista presenta un proyecto surgido en el Carnegie Mellon Institute donde se propone utilizar una red descentralizada de semáforos inteligentes equipados con radares, cámaras y otros sensores para gestionar los flujos de tráfico. El proyecto, denominado Surtrac, despliega sensores para identificar la aproximación de vehículos, calcular trayectoria y velocidad, y así ajustar la duración de la luz verde (o roja) de los semáforos de la zona evaluada. Esto permite manejar en tiempo real el volumen de tránsito que se tiene en cada momento y no utilizar predicciones matemáticas para el mismo.[4]

Un último estudio analizado está relacionado con el trabajo académico, presentado por estudiantes y docentes de la universidad del Líbano en 2016. Aquí se plantea un circuito de semáforos inteligentes, con dos configuraciones, basándose en el ritmo de transito, y cambiando los tiempos de los 4 semáforos pertenecientes a una misma intersección. Esto se logra por medio de un microcontrolador programable 16F877A, una pantalla de LCD, transceiver XBee, sensores IR, botones para el cruce y luces LED de tres colores que reemplazarían a los semáforos actuales. [5]

2 Estudio de Caso general

Para el desarrollo de la solución propuesta se analizaron en una primera etapa algunos aspectos considerados fundamentales para el proyecto. En primer lugar, la solución promueve no realizar cambios en equipamiento o infraestructura de los semáforos existentes. Esta situación resulta muy importante por una cuestión de costos asociadas. Un municipio que pueda implementar una solución efectiva, no debería encontrarse con un costo excesivo para implantarla.

Por tal motivo, y como segunda consideración, se propone agregar sobre la base existente una nueva funcionalidad para que sea más atractivo para el usuario final.

Un aspecto adicional sería contar con una buena comunicación. Para ello se prevé que en un futuro no demasiado lejano los municipios tengan o contraten fibra óptica. La utilización de la misma generaría ventajas importantes para el proyecto de los semáforos, pero, además, permitirá desarrollar nuevas ideas de SmartCity e IoT, a partir de una base sólida y calidad de servicio. Si bien en el contexto nacional la utilización de fibra óptica parece aun demasiado lejana, a paso lento se vienen realizando cada vez una mayor cantidad de tendidos. Además, el advenimiento de la comunicación 5G permitiría mejorar los enlaces de comunicación, pero en este punto los costos asociados pueden ser más importantes.

2.1 ¿Como hacer viable al Proyecto?

El proyecto debe preservar el equipamiento disponible sin generar costos adicionales muy importantes. Bajo estas dos premisas se trabajó en la solución propuesta.

Para responder a la pregunta planteada se propone realizar primero un estudio de calidad y factibilidad, previniendo así que el municipio deba desembolsar dinero en una idea innovadora que después no pueda implementar. Asimismo, se deben plantear las soluciones y requerimientos que serán necesarios para que el usuario perciba y valore las mejoras propuestas.

Teniendo en cuenta lo mencionado anteriormente, la finalidad es que una vez implementado el proyecto se note una mejora para todos los stakeholders involucrados, desde los peatones, conductores, vecinos y hasta las mismas autoridades que lo hicieron posible.

Los peatones deben ver una mejoría en el momento de cruzar cuando la luz se lo permite. Además, los peatones con capacidades diferentes serán considerados y se les propondrá soluciones inclusivas.

Los conductores notarán una disminución en los tiempos de llegada y salida, reduciendo la famosa “ira de carretera” y favoreciendo que la circulación se desarrolle de una forma más cómoda y descongestionada. Las planificaciones de horarios podrán ser mejoradas y la ganancia en tiempo y por ende, disminución, debería ser considerable.

Por último, e igual de importante, es el cuidado del medio ambiente. Un control eficiente del tránsito tendrá consecuencias positivas en el cuidado de nuestro entorno. Un automotor que este menos tiempo ocioso, generará menos emisiones de carbono. Un automovilista menos alterado, generara menos polución sonora.

3 Solución propuesta

La figura 1 presenta el esquema genérico para la solución propuesta. Como se puede observar se presentan dos elementos constitutivos básicos:

- la red interna del municipio
- la conexión exterior de sensores relacionados con cada semáforo.

La conexión entre los mismos es a partir de un medio físico (red óptica, como plantea el gráfico, o eventualmente conexión 4 o 5 G).

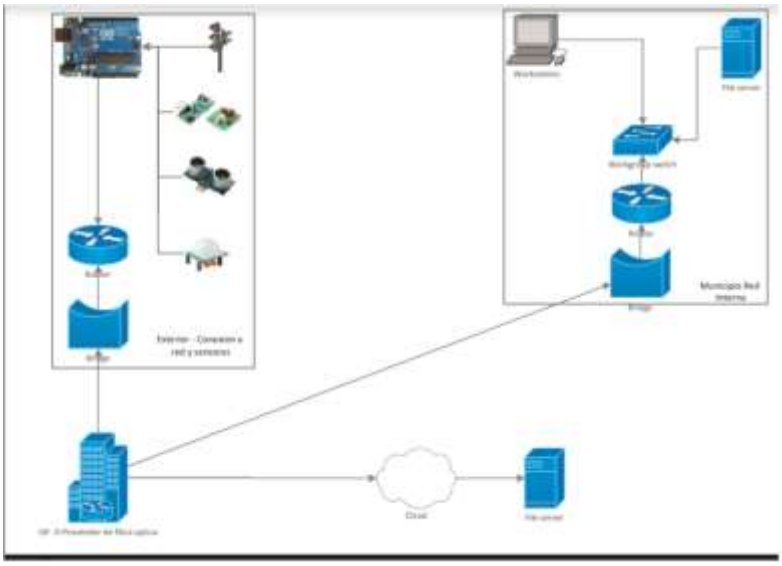


Fig. 1. Esquema gráfico de la solución propuesta.

3.1 Red de sensores

En cada semáforo se propone la instalación de los siguientes sensores:

- Interruptor para cruce peatonal,
- Sensores de proximidad,
- Sensores de movimiento.

El interruptor para cruce peatonal permite su activación por parte de un transeúnte que desee cruzar. En Argentina se permite el cruce de un peatón cuando el semáforo de paso vehicular está con la luz roja, en función del sentido de cruce del peatón. Esta solución tiene un inconveniente, los vehículos que doblan se encuentran con peatones cruzando. Esta solución genera situaciones donde el vehículo no respeta al peatón o que ante una determinada cantidad alta de peatones el tránsito que gira se vea afectado. La utilización de este tipo de botones de cruce permitirá activar un modo, donde todos los peatones de todas las esquinas afectadas por el semáforo podrían

cruzar, estando todos los vehículos detenidos. Esta solución se implanta en grandes ciudades como Tokio, Los Ángeles y Nueva York.

Se debe tener en cuenta, además, que pueden existir semáforos pensados solamente para el cruce peatonal. Esto es, detener el tránsito para que peatones crucen la calle sin que haya necesariamente en ese lugar un cruce de arterias. Estos semáforos deben permitir el paso de una persona cuando sea requerido. Es decir, semáforos que están en onda verde, hasta tanto sea pulsado un botón y no por tiempo de espera. La colocación de este tipo de interruptores tiene ventajas desde varias perspectivas.

Los sensores de proximidad son una alternativa cada vez mas viable para integrar a las personas con capacidades diferentes. Si bien disponer de semáforos que emiten sonidos para personas no videntes ya es una realidad en muchos municipios; hay alternativas superadoras cuyo costo no es elevado. A modo de ejemplo, se puede disponer de sensores de proximidad de bastones para personas no vidente. Estos sensores detectan al bastón del peatón y generan la señal necesaria para que el semáforo pase a luz roja, permitiendo así el cruce. De esta forma no es necesario pulsar un botón para generar el evento.

Esta solución puede adaptarse para personas con capacidad de movilidad reducida, obteniendo la misma funcionalidad que el ejemplo anterior. Los módulos RF de 433Mhz son muy populares por su bajo costo, facilidad de uso y su simple integración con un Arduino. Los mismos trabajan en modo emisor (FS1000A) y receptor (XY-MK-5V)[2]. Se pueden conectar receptores en puntos claves donde haya bajada para sillas de ruedas. Cada silla de rueda puede contar con un emisor, así al acercarse a una distancia adecuada, el receptor recibirá la señal del emisor y se activará la señal para el evento.

Para el control del tránsito vehicular, se propone el uso de sensores de movimiento. En este caso las alternativas son sensores tipo (SENSOR PIR) (SEN0171) + Sensor Ultrasonido Hc-sr04[3]. Este tipo de sensores se pueden utilizar para medir el tiempo en que un automóvil se encuentre detenido. Otra opción, es utilizar sensores ubicados en el asfalto y por presión determinar si un auto está o no en movimiento y por cuanto tiempo. La utilización de estos sensores permitirá evaluar la cantidad de automotores detenidos y de esta forma encontrar un tiempo óptimo para la luz verde del semáforo.

Una ventaja adicional de este tipo de sensores es alertar a los controladores de tránsito del municipio del mal funcionamiento de equipos o en su defecto de la posibilidad de un accidente. Estas posibles situaciones se presentan en caso que un sensor no emita cambios durante un tiempo prudencial.

Estos sensores se conectarán a una placa Arduino, con un shield Mini Ethernet W5100, necesario para la conexión entre el Mikrotik (hEX RB750Gr3) y la placa[4].

Cada panel de semáforos debe tener un código único identificatorio. El arduino que se conectará al panel del semáforo también será identificado unívocamente. Cada arduino tendrá conectados un conjunto de los sensores como los descriptos anterioremente.

La figura 2 describe gráficamente el conjunto de sensores y arduino. Se representa no solo la posición, si no el arco de funcionamiento de los sensores infrarrojos

2 Costo inferior a los 5US$

3 Costo inferior a los 10 US$.

4 Costo inferior a los 15 US$.

trabajando en conjunto con los sensores de ultrasonido, permitiendo conseguir datos que reflejen el estado del tránsito en tiempo real.

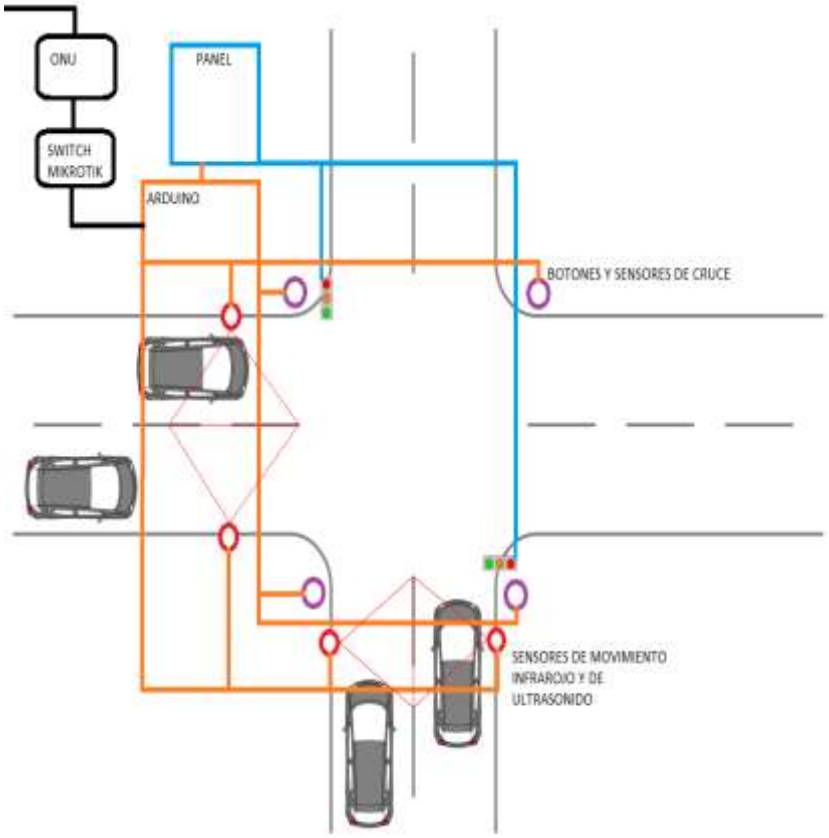


Fig. 2. Colocación estratégica de sensores.

A fin de mantener una correcta trazabilidad de componentes cada sensor deberá ser identificado a fin de permitir realizar tareas de mantenimiento mas efectivas. La conexión prevista es mediante cables UTP categoría 6, para minimizar el error en el envío de las señales respectivas. La figura 3 presenta un esquema de conexión del equipamiento y su vinculación con el municipio.

Como se indicó anteriormente, la elección de los elementos de hardware fue realizada tratando de minimizar al máximo los costos de instalación. Si bien el equipamiento necesario (Arduino + sensores) es por semáforo, cuando se analizan los costos se mantienen por debajo de los 100 US$ por unidad.

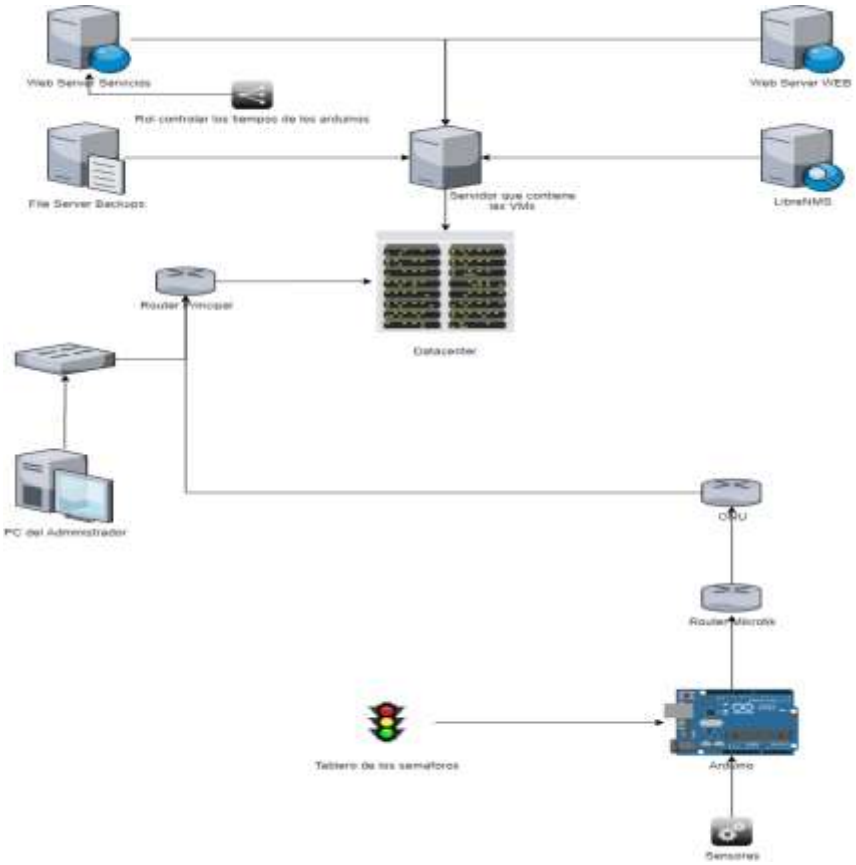


Fig. 3. Esquema de conexión y almacenamiento de datos.

3.2 Servidores

Una de las variantes analizadas para administrar la información del problema consiste en disponer de un servidor físico con cuatro máquinas virtuales. Dichas máquinas virtuales serán responsables de:

- BackEnd: se propone utilizar una versión de Ubuntu, con Java, Apache Maven y una BD MySQL
- FrontEnd: se propone utilizar Ubuntu con Angular
- LibreNMS: es un sistema de monitorización simple, orientado a la monitorización de infraestructuras de mediana envergadura. Esta herramienta permite monitorizar por SNMP diversos hosts y graficar el mapa de red según las conexiones, pudiendo crear alertas personalizadas sobre herramientas como telegram.
- Backup

3.3 Aplicación

El prototipo de aplicación tiene dos componentes básicos: un módulo automático de toma de decisiones y un sistema WEB de monitoreo de tránsito.

El módulo de toma de decisiones, a partir de la información recibida por el conjunto de sensores, y luego de su procesamiento determinará los tiempos de luz de un semáforo para que optimice el tránsito. Se debe tener en cuenta que la configuración definida para un semáforo afectará necesariamente los dispositivos circundantes. De esta forma, el modulo envía señales de configuración al entorno geográfico involucrado.

Este módulo no intervendrá de manera automática exclusivamente. Puede haber factores externos al tránsito puntual que necesiten alterar el funcionamiento de los semáforos. Así, desde el sistema de monitorización se pueden enviar señales manuales que afecten el comportamiento de las luces. Ejemplo de estas situaciones pueden darse en casos de accidentes o catástrofes.

Además, se dispone de un módulo de monitorización del sistema. Esta página web es el medio de comunicación entre el usuario y el sistema de control de semáforos. La figura 4 presenta un prototipo de interfaz tomando como base el municipio de La Plata. Se presenta la vista del Administrador del sistema. Consiste de un mapa donde se marcan los semáforos. Si se selecciona alguno de ellos se despliega el estado de funcionamiento actual del mismo.

Fig. 4. Vista parcial del sistema de monitorización

4 Resultados Obtenidos

Los prototipos de hardware y de software fueron desarrollados y probados en un entorno simulado. La configuración de la placa Arduino y el conjunto de sensores pudo ser desarrollado en un ambiente de laboratorio de forma rápida y efectiva.

El posterior enlace de comunicaciones no presentó mayores inconvenientes y, en general, en el entorno de pruebas el comportamiento del módulo de control se mostro eficiente.

Los resultados obtenidos fueron considerados satisfactorios. Sin embargo, la prueba final de este prototipo se puede lograr en un entorno parcialmente real donde se trabaje con una mayor cantidad de semáforos operando en tiempo real. De esta forma se podrá depurar con mayor certeza el comportamiento del módulo de toma de decisiones.

Aquí el proyecto no pudo realizar mayores avances debido a que por la situación conocida de pandemia no se pudo realizar convenio con algún municipio de la zona.

5 Conclusiones

El trabajo original fue propuesto como tesina de grado para obtener el título de Licenciado en Sistemas. A partir del análisis del estado del arte se pudo ver que existen varias soluciones propuestas y desarrolladas a nivel mundial.

En la mayoría de esas soluciones el presupuesto a invertir por parte del municipio es importante. Por dicho motivo, se analizó el problema desde un punto de vista innovador en base a lo económico, pero que pueda presentar mejoras notorias en el tránsito y que las mismas sean observadas por Stakeholders (transeúntes, automovilistas, administradores del municipio, etc.).

Para ello se propone la utilización de material de hardware que reúna varias características: la calidad, el bajo costo, disponibilidad, y además que permitan una conexión simple y rápida.

El impacto en la calidad de tránsito en un municipio resultaría muy evidente. Cualquier ciudad necesita un cambio en la forma en que funciona su control de tránsito. La aplicación de nuevas tecnologías es fundamental para crear ciudades inteligentes (Smart City).

Como trabajos futuros las posibilidades son variadas. Primeramente comprobar el funcionamiento de la solución en el campo real. Ante este contexto se deberían realizar ajustes en los módulos de software desarrollados y en los prototipos de hardware propuestos.

En una segunda etapa, se propone estudiar el comportamiento de la red de cámaras de video que normalmente cuentan los municipios y utilizar la misma como medio de comunicación de la red de sensores. Este acción permitiría abaratar aún más el costo final del proyecto.

Otra opción futura propone integrar la aplicación del sistema de control de tránsito con la aplicación de un sistema de estacionamiento. Aquí las posibilidades son varias. Si se colocan sensores de estacionamiento se podría controlar el principio y el fin del mismo de manera automática. Esto evitaría, lo que sucede muy a menudo, que

el automovilista pague de más, al terminar el estacionamiento y olvidar mandar el mensaje respectivo. Otra posibilidad concreta es poder brindar información en tiempo real a un conductor de los lugares libres para estacionar. El tiempo que se pierde y la polución que se genera buscando donde aparcar puede ser notoriamente reducido. Además, un potencial conductor puede ser disuadido de utilizar el vehículo si antes de iniciar su trayecto ya sabe que no dispondrá de lugares de estacionamiento.

Referencias

1. Here Mobility, https://www.here.com/platform/mobility-solutions
2. Smart City, https://smartcity.wien.gv.at/en/smart-traffic-lights/
3. Times, https://time.com/5502192/smart-traffic-lights-ai/
4. Surtrac, https://www.rapidflowtech.com/surtrac.
5. Ghazal, B., ElKhatib, K., Chachine, K., Kherfan, M.: Smart Traffic Light Control System. Conference: 2016 third International Conference on Electrical, Electronics, Computer Engineering and their Applicactions (EECEA) Abril 2016

Prototipo de sistema para la gestión de control de tránsito vehicular

Darío Propato[1], Marisa Daniela Panizzi[1], Rodolfo Bertone [2]
[1] Escuela Superior de Ingeniería, Informática y Ciencias Agroalimentarias. Universidad de Morón. Cabildo 134 (B1708JPD). Partido de Morón, Argentina.
[2] Facultad de Informática, Instituto de Investigación en Informática LIDI (III-LIDI), Universidad Nacional de La Plata, La Plata, Argentina
dariopropato@gmail.com; marisapanizzi@outlook.com; pbertone@lidi.unlp.edu.ar

Resumen. La construcción de rutas, caminos y el tránsito vehicular se incrementa año tras año en Argentina acarreando un aumento considerable en las infracciones y accidentes de tránsito. Esto genera una complicación y una demora en los controles de verificación vehicular, retrasando tanto los conductores como a las autoridades de tránsito. En este trabajo se presenta un prototipo para la gestión de tránsito vehicular con la incorporación del escaneo y reconocimiento de patentes vehiculares. El escaneo de patente aumenta la velocidad en el reconocimiento de los datos vehiculares y enriquece de información a las autoridades de tránsito. Se describe la experiencia en la definición, el diseño y el desarrollo del prototipo. Los resultados de las pruebas permiten demostrar la utilidad práctica de este tipo de solución y el aporte de valor agregado para los usuarios, principalmente las autoridades de tránsito que por propiedad transitiva beneficiará a los conductores.

Palabras clave: Prototipo de gestión, control vehicular, escaneo de patentes.

1 Introducción

Actualmente la construcción de rutas y caminos y el tránsito vehicular se incrementa año tras año acarreando un aumento considerable en las infracciones de tránsito y accidentes de tráfico. Las fuerzas de control policial, desplegadas en los controles vehiculares en su mayoría, recurren a controles manuales repetitivos que no agilizan la operatoria del control con la consiguiente demora tanto para los oficiales de control de tránsito como para los conductores que son sometidos a dicho control.

En el período 2007 y 2012 la asociación sin fines de lucro Luchemos por la Vida, comprobó una drástica reducción en la cantidad de actas labradas en los controles de tránsito en la Ciudad de Buenos Aires. La disminución de los controles va de contramano con las actuales estrategias para el logro de la seguridad vial que las experiencias de los países modelo en este tema han desarrollado y que Naciones Unidas recomienda para reducir sustancialmente los muertos y heridos en el tránsito [1].

Como se menciona en la Estadística de incidentes viales con fallecidos y lesionados en 2017 de la provincia de buenos aires es importante contar con datos

confiables sobre los incidentes de tránsito para evaluar el impacto y establecer estrategias que permitan reducirlos [2].

De acuerdo con el informe de Eficacia de los controles de tránsito del año 2002 las actas labradas representan el 16% del total de infracciones graves de todo tipo labradas durante el mes analizado. En la provincia de Buenos Aires se cometen unos 1067 millones de infracciones graves por mes, de las cuales se labran 36893 actas [3].

Desde octubre de 2017, bajo la modalidad Agentes 2.0 los agentes de tránsito de la Ciudad de Buenos Aires cuentan con celulares exclusivamente laborales, adaptados especialmente para optimizar sus funciones: digitalizar y mejorar el proceso de infracciones, coordinar cambios en sus recorridos y optimizar el trabajo de agente mediante el sistema de geoposicionamiento que facilita la ubicación de los Agentes en tiempo real [4].

En la Argentina existe un vehículo por cada 3,1 habitantes. Las estadísticas cuentan solo a los autos y utilitarios. El parque automotor creció un 6,4 con respecto a 2016 y subió un 29,8% con respecto a la primera medición, realizada en 2011 (10,24 millones) [5]. Esto incrementa considerablemente la cantidad de vehículos a controlar, la congestión vehicular y la consecuente demora en el control y labrado de actas en caso de cometerse infracciones.

De acuerdo con el último informe de la Super Intendencia de Seguros de la Nación (SSN) surge que en los doce meses del año 2018 se denunciaron 60832 casos de robo total de vehículos, lo que promedia 167 episodios por día o uno cada 9 minutos [6].

La identificación de los vehículos robados en los controles vehiculares en base a la estadística referenciada requiere de una base de datos que contenga la información actual.

Hoy en día las actas labradas por los agentes de tránsito se labran de forma manual. El agente de tránsito debe consultar al 911 la situación actual de los vehículos que se están controlando. Las imágenes tomadas de los vehículos en infracción se realizan con un dispositivo móvil que luego son adjuntadas al acta digitalizada. Los informes actuales se envían al Ministerio de Seguridad.

Argentina carece de sistemas integrados que permitan agilizar los controles de tránsito vehiculares de forma centralizada, rápida y efectiva. Tanto municipios, como provincias, recurren a métodos de control limitados por el conocimiento y el presupuesto que disponen. Se cuenta actualmente con la Licencia Nacional de Conducir, en la cual el conductor registra sus datos y asocia su licencia de conducir física a la aplicación Mi Argentina.

La autoridad fiscaliza la licencia digital y el código QR1[1] que muestra la misma, verificando su estado. A pesar de que fue publicada por el Ministerio de Modernización en el 2019 por la Disposición 39/2019, en muchas jurisdicciones las autoridades se ajustan a la letra de la norma y exigen que el conductor aporte el carnet tradicional cuando se lo solicitan en un control [7].

Con base en esta problemática de las demoras en los controles de verificación vehicular, que retrasan tanto los conductores como a las autoridades de tránsito, la aplicación de las Tecnologías de la Comunicación e Información (TICs) para optimizar los controles vehiculares y por consiguiente generar una reducción en los

[1] Código de Respuesta Rápida. Es un módulo para almacenar información en una matriz de puntos o en un código de barras bidimensional.

tiempos de verificación, nos permitió definir nuestro objetivo. Este consiste en la construcción de un prototipo de un sistema para la gestión de tránsito vehicular con la incorporación del escaneo y reconocimiento de patentes vehiculares.

Este artículo se estructura de la siguiente manera; se presentan los trabajos relacionados (sección 2), se describe el proceso de definición, diseño y desarrollo del prototipo de sistema para la gestión de tránsito vehicular (sección 3), se describen las pruebas realizadas (sección 4) y, por último, se presentan las conclusiones y trabajos futuros (sección 5).

2 Trabajos relacionados

Se realizó un estudio de mapeo sistemático (SMS, en inglés Systematic Mapping Study) según el proceso propuesto en [8] que permitió dar respuesta a la siguiente pregunta de investigación (PI): *¿Cuál es el estado del arte sobre las soluciones de gestión de control de tránsito?*

La Tabla 1 sintetiza las tareas realizadas en la actividad planificación del SMS.

Tabla 1. Tareas de la actividad Planificación del SMS.

Tareas	Descripción
Definir las Sub-Preguntas de investigación (PI1-PI4).	PI1: ¿Para qué se utilizan las aplicaciones de reconocimiento de patentes? PI2: ¿Sobre qué lenguaje fueron desarrolladas las aplicaciones de reco nocimiento de patentes? PI3: ¿Cuáles son los beneficios que le brindan las aplicaciones de reconocimiento al usuario final? PI4: ¿Qué tipos de investigación se utiliza?
Determinar las cadenas de búsqueda	((“license plate”) AND (“recognition”) AND (“mobile”) AND (“application”)) OR (((“matricula”) OR (“patente”)) AND (“reconocimiento”) AND (“móvil”) AND (“aplicación”))
Determinar los criterios de selección de los estudios.	Criterios de inclusión: • Artículos del 2010 hasta junio del 2020. • Artículos en el idioma inglés y español. • Artículos orientados al desarrollo de aplicaciones móviles para el reconocimiento de patentes y su implementación. • Artículos orientados al desarrollo de aplicaciones para la interacción de oficiales de tránsito y control vehicular. • Artículos duplicados: si hay varios artículos de un mismo autor que contemple la misma investigación, se considerara el más completo. Criterios de exclusion: • Artículos que no cumplan los criterios de inclusión. • Artículos sin revisión por pares. • Artículos que se encuentren en formato de resumen, presentaciones en power point y libros.

Definir las fuentes de datos.	IEEE Explore, Scopus, Academia, SemanticScholar, IJERT, Researchgate, Springer, SEDICI[2], AJAST, IJECE.
Determinar los tipos de publicación.	Artículos de congresos y artículos de revistas.
Definir el período.	2010 hasta junio del 2020.

Los Estudios Primarios analizados se encuentran en [9] junto con una tabla con los resultados por PI, por restricciones de espacio.

Después de analizar 28 estudios primarios, se concluye que:

- La mayoría de los estudios tiene foco principal en la mejora de los algoritmos de reconocimiento, dejando en un segundo plano, pero no por esto menos importante la utilización e implementación de estos para el control vehicular.
- Es muy baja cantidad de desarrollos para la región de América y los mismos hacen foco en la característica de mejora mencionada anteriormente.
- De los estudios primarios analizados, 26 estudios corresponden al tipo de investigación ¨propuesta de solución", 1 estudio corresponde al tipo "evaluación" y otro al tipo "validación".
- La mayoría de los estudios presentan soluciones desarrolladas en lenguaje Android, 20 en total.

Se analizó la literatura existente sobre sobre arte sobre las soluciones de gestión de control de tránsito. Esto permitió evidenciar la necesidad de considerar el escaneo de patente dado a que aumenta la velocidad en el reconocimiento de los datos vehiculares y enriquece de información a las autoridades de tránsito. Desde esta perspectiva, se plantea la construcción de un prototipo de gestión que permita la gestión de control de tránsito vehicular con la incorporación del escaneo y reconocimiento de patentes vehiculares.

3 Desarrollo

Dada la tendencia actual de la práctica de la industria del software del desarrollo de software híbrido [10], la cual se adoptó para el logro de la solución. Esto generó la combinación de métodos, prácticas y estándares, los cuales se detallan a continuación.

Para la especificación de los requisitos del prototipo se siguieron los lineamientos de un estándar tradicional, ISO/IEC/IEEE 29148 [11]. La documentación del proyecto se realizó con la herramienta CASE[3] Enterprise Architect (EA) [12]. Para las actividades de la construcción del prototipo se utilizó un modelo de ciclo de vida iterativo-incremental, respetando las disciplinas propuestas en el Proceso Unificado de Rational (RUP) [13]. El modelado del prototipo se logró mediante los diagramas

[2] SEDICI: Repositorio digital de la Universidad Nacional de La Plata, website: http://sedici.unlp.edu.ar/

[3] Ingeniería de Software Asistida por Computadora (en inglés, Computer Aided Software Engineering).

de UML[4] [14] que consideramos necesarios para la definición, diseño y desarrollo del prototipo [9]. Los diagramas logrados se detallan en la Tabla 2.

Tabla 2. Diagramas de UML según disciplinas del RUP.

Disciplina de RUP	Denominación del Diagrama.
Ingeniería de Requerimientos	Diagrama de casos de uso.
Análisis y Diseño	Diagrama de casos de uso, Diagrama de comunicación, Diagrama de clases, Diagrama entidad-relación, Diagrama, Diagrama de secuencia.
Implementación	Diagrama de despliegue, Diagrama de componentes.

El prototipo para la gestión de control de tránsito vehicular cubre las siguientes funcionalidades: el alta de los usuarios, los mismos, de acuerdo con distintos tipos de perfiles según la funcionalidad especifica que tengan en la aplicación, actualización de los perfiles y datos de estos. El escaneo de patentes función principal que permitirá agilizar el procesamiento de los datos. La carga de las infracciones de tránsito carga de denuncias por mal estacionamiento; ambas funcionalidades podrán ser registradas tanto por los agentes de tránsito como los usuarios que no pertenecen a los agentes de control. Además, contará con la generación de reportes de denuncias, infracciones por distintas agrupaciones, agente de tránsito, municipio por rango de fechas.

3.1 Especificación de requisitos

Los requisitos funcionales especificados en el documento de Especificación de Requisitos de Software (ERS) se presentan en el diagrama del modelo de casos de uso de la Figura 1. Por razones de falta de espacio, los requisitos no funcionales y las características generales de los usuarios que utilizarán el sistema se describen en [9].

En la Figura 2 se presenta el modelo conceptual del prototipo de sistema de información móvil que permitirá automatizar el proceso de control vehicular. En el mismo se visualiza a la autoridad de tránsito escaneando la patente del vehículo y realizando el control pertinente de la documentación y ejecutando un acta de infracción en el caso que sea necesario. Por su parte, un conductor puede denunciar un auto en infracción tomando una foto de este y finalmente el administrador encargado de la gestión de estas denuncias. Además, se visualiza en el mismo, como el conductor posee la documentación digitalizada en su dispositivo móvil. Las interfaces de conexiones entre el servidor central y los dispositivos móviles cumplen un rol central. Se visualiza como se concentran los componentes del prototipo de sistema y se incluyen las interfaces de usuario, el software desarrollado y el almacenamiento en base de datos, que se utiliza para darle la información necesaria al usuario.

[4] UML; Lenguaje de Modelado Unificado.

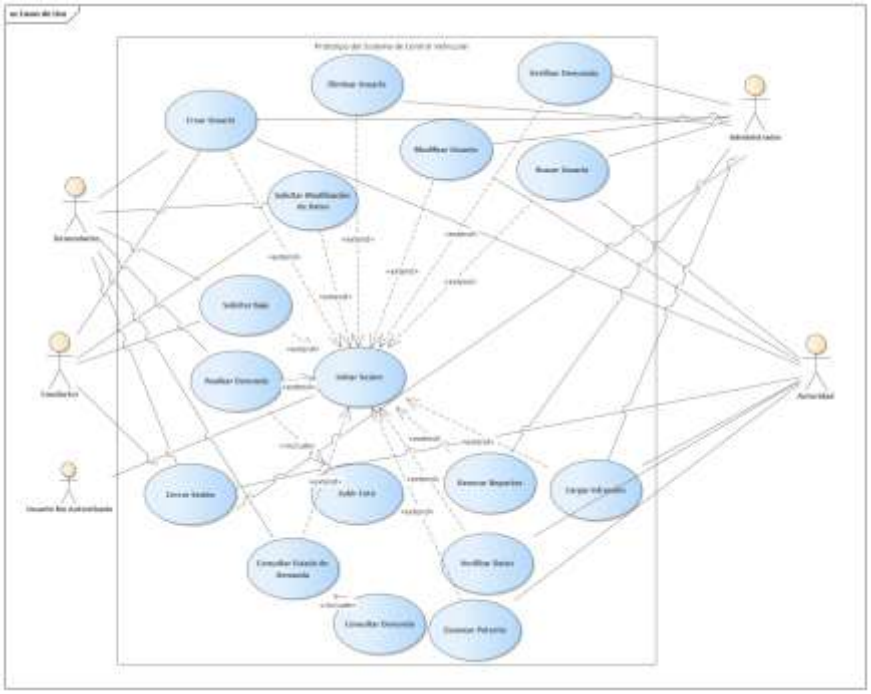


Fig. 1. Diagrama de casos de uso.

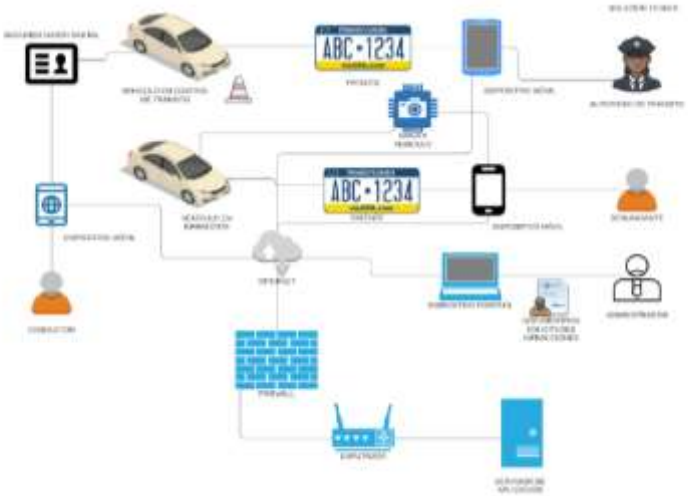


Fig. 2. Modelo conceptual del prototipo de alto nivel.

3.2 Análisis y Diseño

Se construyó el modelo de análisis, con el propósito de lograr una especificación detallada de los requisitos hacia una primera aproximación al diseño del prototipo. A modo de ejemplo, se presenta el Diagrama de Comunicación del Caso de Uso – Escanear patente [9].

Se construyeron los Diagramas de Secuencia para los Casos de Uso presentados en la Figura 1. En [9] se presenta a modo de ejemplo el Diagrama de Secuencia del Caso de Uso: Escanear patente y su explicación.

3.3 Implementación

Los Diagramas construidos en esta etapa del Proyecto contribuyeron en la decisión respecto a la organización de los componentes del prototipo. Estos son el Diagrama de Componentes y el Diagrama de Despliegue, ambos se encuentran en [9].

El desarrollo del prototipo se realizó mediante el lenguaje de programación C# con interfaz web, utilizando el IDE Visual Studio 2017. Se utilizaron librerías y frameworks como ADO.NET Entity Framework, jQuery y el motor de base de datos SQL SERVER 2012.

A continuación, se presentan algunas de las estadísticas que se pueden obtener el prototipo sistema para la gestión de control de tránsito vehicular (usuario Administrador). En la Figura 6 se presentan estadísticas de denuncias por localidad dentro un rango de fechas determinado. En la Figura 7 se presentan las estadísticas de infracciones por localidad dentro un rango de fechas determinado. La Figura 8 presenta las estadísticas de infracciones por categoría dentro de un rango de fechas determinado. En la Figura 9 se presentan las estadísticas de los controles de tránsito por localidad dentro de un rango de fecha. La Figura 10 presenta las estadísticas de los controles de tránsito por agente dentro de un rango de fechas determinado.

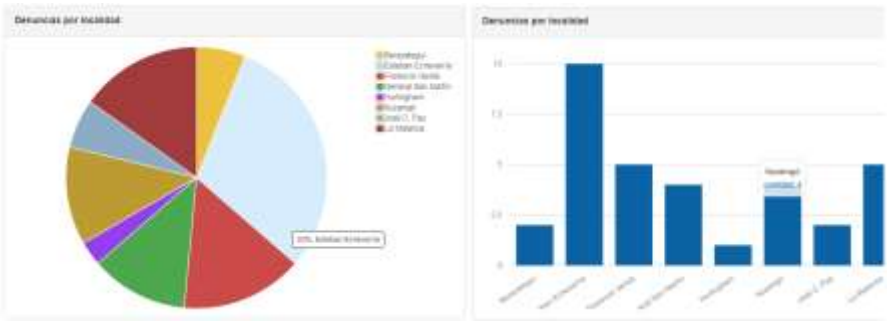


Fig. 6. Estadísticas de denuncias por localidad dentro de un rango de fechas determinado.

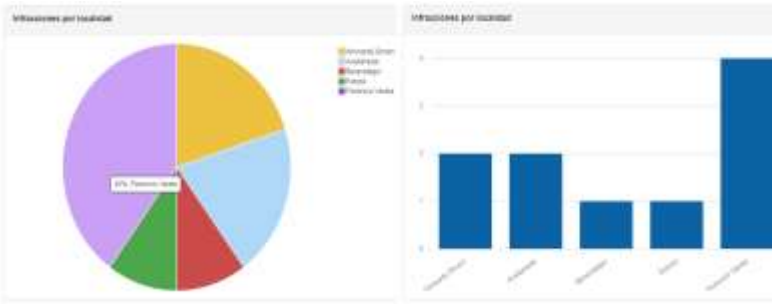


Fig. 7. Estadísticas de infracciones por localidad dentro de un rango de fechas determinado.

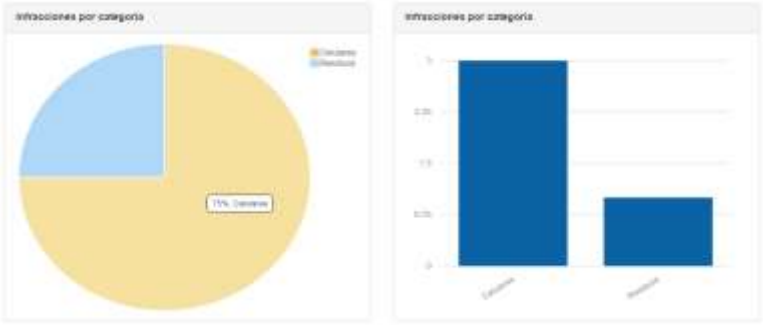


Fig. 8. Estadísticas de infracciones por categoría dentro un rango de fechas determinado.

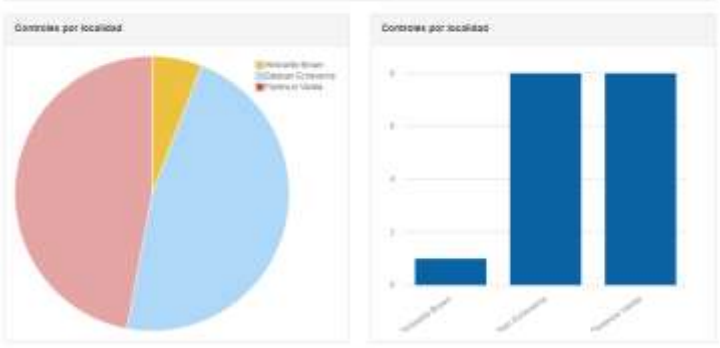


Fig. 9. Estadísticas de controles de tránsito por localidad dentro de un rango de fechas determinado.

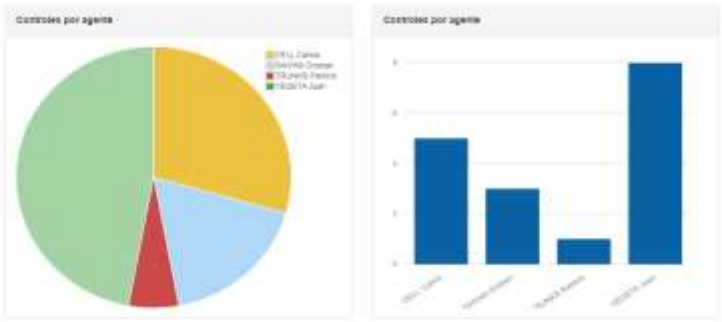


Fig. 10. Estadísticas de controles de tránsito por agente dentro de un rango de fechas determinado.

4 Pruebas

Se planificaron y desarrollaron un conjunto de pruebas unitarias y funcionales (integración). Se realizaron en total 29 pruebas de integración. En [6] se describe a modo de ejemplo las trazas de las pruebas de integración PI20 del CU9: Escanear patente y verificar que el usuario no está cargado en el sistema.

Los tipos de pruebas definidos han sido acertados para verificar el correcto funcionamiento del prototipo.

5 Conclusiones y trabajos futuros

En este trabajo se describe la experiencia en la definición, diseño y desarrollo de un prototipo de sistema para la gestión de control de tránsito vehicular.

La investigación permitió identificar que actualmente en Argentina no hay aplicaciones dedicadas a agilizar y optimizar los tiempos en los controles vehiculares, generando demoras considerables tanto para las autoridades de tránsito como para los conductores.

El proceso de control y reporte de denuncias se realiza actualmente de forma manual y descentralizada, involucrando exceso de personal para llevar a cabo tareas únicas. Adicional a esto, el personal no tiene acceso a la información en un tiempo considerablemente rápido, necesario en estas situaciones.

Esta solución es el puntapié inicial para optimizar controles vehiculares y el acceso a la información de forma centralizada, necesario en la actualidad con el incremento en la cantidad de vehículos que circulan actualmente en la República Argentina.

Se destacan como trabajos futuros: a) la generación de estadísticas y reportes que permitan evaluar comportamientos comunes de los conductores, alertando a las autoridades de tránsito de posibles comportamientos sospechosos que van surgiendo durante los controles, b) adicionar la intercomunicación en el sistema mediante las distintas autoridades de tránsito, Propuesta para mejorar y retroalimentar la

información entre los distintos responsables de la seguridad y c) integrar las bases de datos estatales tanto de las fuerzas de seguridad, como de los organismos del estado que contienen la información correspondiente a los usuarios conductores. Esto permite reducir tanto el volumen como el tiempo de carga de datos por parte de los usuarios, mitigando los errores que se puedan producir al ingresar la información.

Referencias

1. Aciti C., Marone J. A., Capra, J., & Capra, B. (2012, October). Una aplicación móvil para el reconocimiento automático de matrículas de automóviles argentinos. http://sedici.unlp.edu.ar/handle/10915/23802.
2. Ministerio de Salud, M. de S. (2019). Estadísticas de Incidentes Viales con Fallecidos y Lesionados 2017. Gobierno de La Provincia de Buenos Aires. Obetnido de http://www.gob.gba.gov.ar/UF/Informe.pdf.
3. Luchemos por la Vida (2002). Eficacia de los controles de Tránsito. Obtenido de http://www.luchemos.org.ar/revistas/articulos/rev22/pag08.pdf.
4. Cuerpo de Agentes de Tránsito. Obtenido de https://www.buenosaires.gob.ar/movilidad/plan-de-seguridad-vial/cuerpo-de-agentes-de-transito.
5. Asociación de Fábricas Argentinas de Componentes (05, 2018). Flota Circulante en Argentina 2017. Obtenido de www.afac.org.ar.
6. Nicolas Jasper. (2018). Anexo I, Vehiculos Expuestos a Riesgo. Recuperado de Rhttps://www.argentina.gob.ar/sites/default/files/ssn_20172018_vehiculo_expuesto_riesgo_anexo.pdf.
7. Clarín.com. Licencia de conducir digital: para qué sirve y por qué no reemplaza al carnet tradicional. Clarín. Recuperado del 20/02/2020, https://www.clarin.com/ciudades/licencia-conducir-digital-sirve-reemplaza-carnet-tradicional_0_KT2gpeaK.html.
8. Kitchenham, B. Budgen D. and Brereton, P. (2015). Evidence-Based Software Engineering and Systematic Reviews. Chapman and Hall/CRC. 1 st.Editon. New York, USA.
9. Propato D., Panizzi. Anexo – Prototipo de sistema para la gestión de control de tránsito vehicular. Disponible en: https://doi.org/10.6084/m9.figshare.15130482.v1
10. Kuhmann M., Tell P., Klünder J., Hebig R., Licorish S., MacDonell S. (Eds.): Complementing Materials for HELENA Study (Stage 2), online DOI: 10.13140/RG 2.2.1 1031 65288, published: 2018-11-28.
11. ISO/IEC/IEEE Std 29148:2011, IEEE Systems and software engineering -- Life cycle processes --Requirements engineering, IEEE Computer Society (2011). ISBN 978-0-7381-6591-2.
12. Sparx Systems. Enterprise Architect 13.5. (2017). Obtenido de https://sparxsystems.com/
13. Jacobson, I., Booch, G., & James, R. El Proceso Unificado de desarrollo de software. Addison Wesley (2000).
14. Rumbaugh, J., Jacobson, I., Grady, B. El Lenguaje Unificado de Modelado. Manual de Referencia. Segunda Edición. Pearson (2006).

ISBN 978–987–633–574–4

Gestión de presencialidad en la virtualidad para la Universidad Nacional de Río Negro

Lugani, Carlos Fabián

{clugani} @unrn.edu.ar

Universidad Nacional de Río Negro, Sede Atlántica
Laboratorio de Informática Aplicada

Resumen. Existe una nueva forma de tomar clases, trabajar, o simplemente estar presentes, en donde las personas se conectan mediante las tecnologías de las comunicaciones por videoconferencia, videollamada o algún tipo de conexión a sistemas o aplicaciones. Ante esta realidad se observan varias necesidades: la de comprobar la identidad de las personas que comienzan la conexión debido a que la organización desea dar al acceso sólo a las que deben hacerlo, la de comprobar si la conexión con esa persona se ha visto interrumpida, o se ha desconectado definitiva o momentáneamente; y además si existe la atención de la persona ante el sistema, videoconferencia o la conexión establecida. Esto será dado por un proceso de difícil resolución y existen grados de comprobación asociados a la dificultad de comprobación que irían desde simples pruebas y amplia confianza hasta sistemas más complejos que establezcan un grado de aseguramiento de la presencialidad que se podrán resguardar y ser auditados.

En este primer trabajo se define la problemática anterior y el desarrollo de un esquema sobre el cual se desarrollarán aplicaciones que se puedan utilizar en la Universidad Nacional de Río Negro para las necesidades de las plataformas existentes o futuras para ser utilizadas en forma efectiva por la organización.

Palabras clave: ausentismo, presencialidad, identificación.

1 Introducción

El presente trabajo se define como un esquema en donde se presentará la problemática y posibles cursos de acción o escenarios para desarrollar soluciones y definiciones que sirvan para establecer marcos de trabajo.

La presencialidad como acción de las personas de estar físicamente en un lugar ha sido modificada como resultado de la pandemia Covid y ha modificado en varias formas la forma de trabajar, presenciar clases, participar de reuniones y múltiples actividades. Esta presencialidad también se podría entender como el “estar en línea” o “estar disponible” a tiempo completo, pero se establece que los mecanismos o controles que se diseñen tengan asociados como requerimientos los tiempos estipulados de conexión necesaria o requeridos que se hayan previamente pautado. Siendo importante

el establecimiento de límites en la definición de las conexiones o disponibilidad del tiempo de las personas.

Actualmente los docentes se enfrentan a un nuevo paradigma que es la enseñanza sin presencialidad [(1)] y se ofrecen herramientas, recomendaciones, ideas que puedan serles útiles y en general formas de pensar o repensar la enseñanza para que sus aulas virtuales o cualquier tipo de estrategia sean enriquecidas. Aun así, existe la necesidad genuina de los docentes de requerir clases en donde los alumnos se conecten y tomen clases en forma remota necesitando de herramientas para comprobar que son los alumnos que tienen que estar conectados y de alguna forma que los mismos cumplen con ciertas pautas de presencialidad en esa virtualidad. Lamentablemente, los docentes no tienen las herramientas para verificar identidades o para comprobar la atención de los alumnos. En este sentido consiste el aporte que se pretende alcanzar con el presente trabajo.

Asimismo, las organizaciones también desean verificar identidades del personal tanto docente como no docente, que realizan tareas en sistemas, en reuniones, atención a personas u otras actividades por las cuales las personas se conectan a computadoras con acceso a Internet y a diversas aplicaciones y por las cuales cumplen con tareas que le han sido encomendadas. Por lo tanto, también las organizaciones desearían tener herramientas de control para poder hacer un seguimiento del presentismo, horarios de disponibilidad de personal, efectividad de atención y tiempos de respuesta.

Por otro lado, se destaca la evolución de herramientas de conectividad [(2)], la mejora en los sistemas de videoconferencia, existencia de pizarras compartidas, así como nuevas situaciones mixtas en donde un conjunto de personas están en forma presencial y otro conjunto están en forma virtual y participando síncronamente gracias a los medios de interacción que estos sistemas aportan y que son propios de la virtualidad. Así, un espacio virtual ha recibido varias denominaciones y se prefiere el nombre de Ambientes Virtuales de Enseñanza -Aprendizaje (AVEA) para el espacio virtual donde los miembros de una comunidad educativa interaccionan con la finalidad de desarrollar un proceso formativo, mediante la aplicación de las nuevas tecnologías de la información y la comunicación [(3)].

Por supuesto que lo anterior implica que las personas que se conecten y sean controladas acepten las condiciones que se plantean, si bien se prevén dificultades en cuanto a la disponibilidad o falta de disponibilidad de la tecnología necesaria para cumplir con los requerimientos de control como cámaras, micrófonos, conexiones estables a Internet y requerimientos de hardware más específicos, es indudable que las personas deberán contar con los medios físicos que aseguren la conexión y controles que se definan.

Se destaca que al realizar este trabajo se han buscado referencias académicas para las bases de su realización y se han encontrado, pero no se han encontrado para las secciones de desarrollo, se considera que el mercado o los productos de alguna forma pueden ofrecer variantes de comprobaciones de que los usuarios están conectados, pero no se ha establecido un esquema como el presentado en este trabajo en forma académica con lo cual no existen muchas referencias que hayan ayudado a definir este marco de trabajo.

2. Comprobación de identidades

La identidad personal es un conjunto de rasgos característicos de un individuo, por los cuales se puede decir que esa persona es realmente quien dice ser. Como primer paso se debe definir la identidad de la persona que está conectada al sistema que se desea controlar. Existen numerosos métodos para identificar a las personas, podemos diferenciar a los biométricos que dependen del cuerpo de una persona como reconocimiento facial, reconocimiento de voz, venas de los dedos, geometría de la mano, cadencia de movimientos del cuerpo, etc. Por el otro lado, se puede identificar a una persona a través de contraseñas o preguntas que sólo la persona puede conocer. También existen dispositivos que una persona puede poseer como tokens de seguridad (también conocidos como token de autenticación o token criptográfico, son dispositivos portátiles de alta tecnología que generan una clave de forma aleatoria e irreemplazable, y que están asociados a un sistema de firma digital) o certificados digitales que están instalados en computadoras o celulares y que entregan un número o código que sirve para autenticar a la persona que posee ese dispositivo de hardware o teléfono celular.

En resumen, existen tres grandes conjuntos de formas de comprobar la identidad de una persona, (a) por algo que la persona conoce y recuerda como una contraseña, (b) por algo que la persona posee como un token o certificado digital, o (c) por algo que la persona es, como su huella dactilar o su rostro.

En primera instancia se puede definir que de acuerdo al nivel de control que se quiera establecer o el nivel de confianza en que el control será efectivo, será la necesidad de utilizar un mecanismo mas avanzado de autenticación y por lo tanto para sistemas que deban comprobar y registrar en forma segura la identidad de una persona, se necesiten mas niveles de autenticación o la combinación de dos tipos.

Por lo tanto, cualquiera sea la metodología, es conveniente establecer una asociación del tipo: nivel de autenticación requerido, nivel de información, activos de información a los que se accede o a valores (en caso de información contable), a categorías de información confidencial o procedimientos por ejemplos en casos de un proceso jurídico o legal. Es de especial cuidado y se deben analizar otras implicancias en el caso de por ejemplo exámenes en una Universidad que se tomen a través de Internet, ya que la autenticación de la persona y la asociación de un examen realizado por ella constituye un documento con implicancias legales.

También se debe tener en cuenta actos de votación que se pueden dar y otros actos en donde se requiere que una persona sea verificada de alguna forma porque será parte de una decisión que quedará documentada la coordinación.

Se desarrolla una tabla (*Tabla 1 – Actividad relacionada con nivel de autenticación para la Universidad Nacional de Río Negro*) para presentar casos posibles, sin que estos quieran determinar todos los casos posibles que son necesarios en una organización. Al desarrollar los mismos se tiene en cuenta los requerimientos que se observan en la Universidad Nacional de Río Negro, ya que este esquema se considera para su evaluación y uso en esta institución.

Actividad relacionada con:	Actividad	Posible nivel de autenticación requerido
Docencia	Presentismo en actividad programada ON LINE (Docente)	Nivel 1: Usuario y Contraseña Nivel 2: Certificado digital
Docencia	Presentismo en actividad programada ON LINE (Alumno)	Nivel 1: Usuario y Contraseña Nivel 2: Certificado digital
Docencia	Presentismo en examen parcial o presentación de trabajo práctico ON LINE (Docente y Alumno)	Nivel 1: Usuario y Contraseña Nivel 2: Certificado digital
Docencia	Presentismo en examen final ON LINE (Docente y Alumno)	Nivel 1: Usuario y Contraseña Nivel 2: Certificado digital Nivel 3: Reconocimiento facial
No docencia	Presentismo en puesto de trabajo ON LINE	Nivel 1: Usuario y Contraseña Nivel 2: Certificado digital
No docencia	Coordinación de trabajo en puesto de trabajo ON LINE	Nivel 1: Usuario y Contraseña Nivel 2: Certificado digital Nivel 3: Reconocimiento facial
Investigación	Evaluación de proyectos, integrantes de proyectos, becas.	Nivel 1: Usuario y Contraseña Nivel 2: Certificado digital Nivel 3: Reconocimiento facial
Comunidad de la UNRN	Participación en reuniones informativas	Nivel 1: Usuario y Contraseña
Comunidad de la UNRN	Participación en consejo sin voto	Nivel 1: Usuario y Contraseña Nivel 2: Certificado digital
Comunidad de la UNRN	Participación en consejo con voto	Nivel 1: Usuario y Contraseña Nivel 2: Certificado digital Nivel 3: Reconocimiento facial

Tabla 1. Actividad relacionada con nivel de autenticación para la Universidad Nacional de Río Negro

La definición de identificación de personas ante el sistema que se vea involucrado siempre debería ser programada y diseñada previamente. Lo que se plantea en este trabajo es el diseño de un grado de requerimientos, nivel de autenticación o la suma de varios métodos de autenticación. Se debe siempre tener en cuenta el valor de la información y las implicancias legales de las acciones que las personas pueden realizar en la virtualidad, asimismo se deben revisar los procesos administrativos asociados que se encuentren involucrados en busca de responsabilidades definidas, actividades requeridas que no pueden faltar o no ser realizadas, controles que deben realizar las personas y dejar registros de su presencia y ejecución, y en general la revisión de las tareas de las personas para identificar aquellas actividades que deben ser analizadas para su realización en la virtualidad y sobre las cuales se puede efectivamente comprobar que han sido realizadas por esas personas.

Este esquema propuesto se lo puede calificar como modelo de identidad de personas, cada componente de comprobación de identidad equivale a un porcentaje que se va sumando para dar un total de comprobación de identidad. Así se puede decir que un numero de 100, sería la comprobación completa de la identidad de una persona a través de diferentes métodos. De acuerdo a la necesidad o requerimiento de un proceso o aplicación se establece que sea un porcentaje determinado de comprobación de identidad, lo cual es algo lógico de realizar ya que a mayor responsabilidad o implicancias de la participación de una persona, se le requerirá mayores comprobaciones de su identidad.

3. Comprobación de presencialidad

Se define la comprobación de presencialidad a un registro que depende de la característica del sistema a que se está conectado la persona, pero que permite siempre dejar un rastro que la persona estaba participando y atento a lo que sucedía en la actividad.

Por ejemplo, si existe una reunión en donde una persona está disertando ON LINE sobre un tema en particular (presentando un gráfico sobre el almidón de maíz y la variación de otras materias primas en la producción de alimentos:) y aparece una ventana en la pantalla de los participantes en donde debe seleccionar una opción de tres presentadas (En este momento se está hablando sobre Almidón de Maíz / Aditivos / Arroz) teniendo 10 segundos para presionar una de las tres opciones y luego desapareciendo la ventana. Este registro de cada participante es suficiente muestra que estaba prestando atención y por sobre todo que estaba en ese momento y que pudo interactuar con el sistema.

Pueden definirse múltiples o diversas comprobaciones de presencialidad y el objetivo es definir estrategias de comprobación de que la persona se encuentre presencialmente delante de la computadora.

Se definirán múltiples comprobaciones de presencialidad que se pueden enumerar como:

1. Comprobación de tipo Opción a completar con pregunta con vencimiento y desaparición de la pregunta (descripto como ejemplo anteriormente).

2. ¿Comprobación de tipo “Esta Ud. ahí?” mediante chat en que se registra el tiempo de pregunta y el tiempo de respuesta de la persona.
3. Comprobación por foto (para lo cual la cámara debe estar habilitada). En este tipo de comprobación se deben realizar múltiples fotos de la persona que se encuentra delante de la computadora y se puede relacionar con la identificación biométrica (reconocimiento facial).
4. Comprobación de aplicaciones. Se debe analizar la aplicación en uso en la computadora, las aplicaciones abiertas y aplicación en foco durante la conexión, teniendo que ser el resultado un porcentaje mayor de la aplicación que se requiera en la conexión.
5. Encuesta o cuestionario de contestación rápida. Preparada con antelación y a disposición por un tiempo determinado dentro de la actividad comprueba que la persona esta activa y responde en un tiempo también determinado a la misma.
6. Comprobación a través de votación. La persona que diserta solicita que los participantes realicen una votación o selecciones en particular una opción, con un sistema que registra a través del tiempo las selecciones de las personas y las relaciona con las personas conectadas. Entregando un listado de las conexiones efectivas realizadas o los participantes que efectivamente realizaron la premisa tal como fue solicitada.

Además, se establece una relación entre la comprobación de la presencialidad con la comprobación de identidades en que también debe ser definida para cada sistema o aplicación. Al principio de la comunicación se establece que la identidad es la primera acción que se realiza y luego pasa a comprobarse que la persona se encuentra disponible y presente en la comunicación, pero pueden requerirse más comprobaciones de identidad luego de cierto tiempo. Se establece que luego de dos horas de conectado se requerirá de una comprobación de identidad que asimismo servirá de comprobación de presencialidad. Si cualquier situación hace que la comprobación no sea exitosa, se registrará el hecho y se finaliza la conexión.

4. Sistema de auditoria

Se considera que la comprobación de identidad primero y luego las sucesivas comprobaciones de presencialidad dejarán como resultado una cantidad de registros que serán la comprobación de las acciones que las personas realicen en los sistemas o conexiones.

Esta cantidad de registros requieren un sistema de gestión que tenga algunas características como las de:

- Identificación de personas que forman parte del sistema de autenticación con resguardo especial de claves, firmas digitales y datos biométricos
- Acceso a la información sobre actividades y necesidades de identificación en particular
- Participación de personas en actividades en forma general o particular
- Registros de votación

- Controles efectivos registrados en los sistemas

El almacenamiento de la información de la identidad de las personas, y sus actividades en los sistemas, conexiones contendrá información confidencial y sensible. Se debe resguardar este tipo de información ya que existe una legislación específica en la República Argentina (Ley 25.326 – Protección de Datos Personales) [(4)].

5. Conclusiones

Luego de haber establecido ciertos parámetros y definiciones, se ha realizado una aproximación a un esquema de administración de la autenticación de personas y controles de su presencialidad en actividades virtuales. Si bien el origen de este trabajo es la necesidad de una organización en mejorar los procesos con que se cuentan y tener registros que se puedan comprobar a través de sistemas informáticos; este esquema se puede extender en forma normal a otro tipo de organizaciones que actualmente se encuentran en situaciones similares. También este esquema se puede extender al concepto de Ciudades Inteligentes, mediante estas herramientas, las ciudades pueden contar con realizar transformaciones digitales o incorporar tecnología utilizando el componente de identificación del ciudadano, realizar reuniones informativas con comprobación de personas, o comprobar la participación de los ciudadanos en plebiscitos. Asimismo, se puede aplicar el funcionamiento de este esquema para ampliar los alcances de actividades como las de Fábricas Inteligentes (Smart Factories) o Viviendas Inteligentes (Smart Homes) ya que los componentes que se desarrollan pueden servir tanto para comprobar la identidad de personas como la presencialidad.

Se menciona también que inmediatamente a la definición de este esquema se debe comenzar a desarrollar aplicaciones asociadas las cuales al funcionar en entornos reales será beneficioso para el mejoramiento del presente esquema pudiendo ampliar sus alcances y comprobando su funcionamiento.

Finalmente, se destaca que este esquema es un proceso de apoyo a la virtualización que se está llevando a cabo en todos los ámbitos, siendo valioso el aporte para asegurar que la misma se apoye sobre trabajos formales que establezcan definiciones y sobre experiencia en la formulación de estrategias que se apliquen para solucionar la problemática detectada.

6. Referencias

1. Mazza, D. (2020). Lo que la pandemia nos deja: una oportunidad para pensarnos como docentes. Citep. Centro de Innovación en Tecnología y Pedagogía. [Sitio web] http://citep.rec.uba.ar/covid-19-ens-sin-pres/

2. Medina Uribe, Jury Carla ; Calla Colana, Godofredo Jorge ; Romero Sánchez, Phill Arnold (2019) Las teorías de aprendizaje y su evolución adecuada a la necesidad de la conectividad Revista de la Facultad de Derecho y Ciencia Política de la Universidad Alas Peruanas, ISSN-e 2313-1861, ISSN 1991-1734, págs. 377-388

3. Mario E. Díaz Durán, Mariela Svetlichich. Nuevas Herramientas Tecnológicas en la Educación Superior PROYECCIONES - Nº 11 - Año XI (2016) pp. 93 – 149 [Sitio web] https://revistas.unlp.edu.ar/proyecciones/article/download/6485/5565/
4. Ley 25.326 Protección de Datos Personales Sancionada: Octubre 4 de 2000 y Normas complementarias Disposición 7 / 2010 Dirección Nacional de Protección de Datos Personales/ Decreto Reglamentario 1558 / 2001 [Sitio web] https://www.argentina.gob.ar/aaip/datospersonales