AVANCES Y EXPERIENCIAS

Innovadoras

en Computación e Informática

EDITORIAL
UNIVERSIDAD CESMAG

AVANCES Y EXPERIENCIAS

Innovadoras

en Computación e Informática

Compilación y Edición
Javier Alejandro Jiménez Toledo
Cesar Collazos
Robinson Andrés Jiménez Toledo
Oscar Revelo Sánchez
Manuel Ernesto Bolaños González

Avances y experiencias innovadoras en computación e informática / Javier Alejandro Jiménez Toledo, Cesar Collazos, Robinson Andrés Jiménez Toledo, Oscar Revelo Sánchez, Manuel Ernesto Bolaños González, Compiladores. -- 1. Ed. – San Juan de Pasto : Universidad Cesmag
2020.
400 p. : color. (Serie Investigación)

Incluye referencias bibliográficas al final de cada capítulo

ISBN: 978-958-5504-82-0
E-ISBN: 978-958-5504-81-3
DOI: 10.15658/CESMAG20.10090122

1. CIENCIA DE LA COMPUTACIÓN 2. INNOVACIÓN TECNOLÓGICA 3. GEOINFORMÁTICA 4. INTELIGENCIA ARTIFICIAL 5. SERVICIOS BASADOS EN LOCALIZACIÓN 6. VIDEOJUEGOS I. Jiménez Toledo, Javier Alejandro, Collazos, Cesar, Jiménez Toledo, Robinson Andrés, Revelo Sánchez, Oscar, Bolaños González, Manuel Ernesto, Compiladores II. Título.

CDD
621.392
22. ED.

CEP – Universidad Cesmag. Biblioteca Remigio Fiore Fortezza.

Avances y experiencias innovadoras en computación e informática

Primera edición, 2020

© Javier Alejandro Jiménez Toledo, 2020
© Cesar Collazos, 2020
© Robinson Andrés Jiménez Toledo, 2020
© Oscar Revelo Sánchez, 2020
© Manuel Ernesto Bolaños González, 2020

© Universidad Cesmag
Editorial Universidad Cesmag
Carrera 20A # 14-54
Tel: +572 7216535 ext: 265 y 377
E-mail: editorial@unicesmag.edu.co
Website: www.unicesmag.edu.co
San Juan de Pasto, Nariño, Colombia
CP: 520003

© Grupo de Investigación Tecnofilia
Carrera 20A # 14-54
Tel: +572 7216535 ext: 336 y 265
E-mail: jajimenez@unicesmag.edu.co
San Juan de Pasto, Nariño, Colombia
CP: 520003

APA

Jiménez, J., Collazos, C., Jiménez, R., Revelo, O., y Bolaños, M. (2020). *Avances y experiencias innovadoras en computación e informática.* San Juan de Pasto, Colombia: Editorial Universidad Cesmag. doi: 10.15658/CESMAG20.10090122.

El pensamiento que se expresa en esta obra es responsabilidad exclusiva de los autores y no compromete la ideología de la Universidad Cesmag.

Se permite la citación del texto nombrando la fuente.

Todos los derechos reservados. Esta publicación no puede ser reproducida totalmente y en partes por ningún medio mecánico, fotoquímico, electrónico, magnético, digital, fotocopia o cualquier otro, sin el permiso previo por escrito de la editorial o sus autores.

ISBN: 978-958-5504-82-0
E-ISBN: 978-958-5504-81-3
DOI: 10.15658/CESMAG20.10090122

Rector
Fray Daniel Omar Sarria Tejada OFM. Cap.

Director editorial
Javier Alejandro Jiménez Toledo

Edición
Diana Milena Betancourth Castillo

Edición impresa y digital
Impreso y hecho en Colombia
Printed and made in Colombia

Diseño de cubierta y diagramación
Angelica Mayag Chud - angelicamayag@gmail.com

Autores

Jiménez Toledo Javier Alejandro.
Héctor Fabio Ocampo Arbeláez.
Carlos Andrés Delgado Saavedra.
Óscar A. Martínez.
Alixón Franco Salazar.
Germán Andrés Holguín Londoño.
Carlos Fernando González Guzmán.
José María Muñoz Botina.
Fredy Santacruz.
Alexandra Ruiz Gaona.
William Joseph Giraldo Orozco.
José Luis Arciniegas.
Javier Suarez Jaramillo.
Mateo Córdoba.
John Evert Barco Jiménez.
Camilo Lagos Mora.
Manuel Ernesto Bolaños González.
César Alberto Collazos Ordóñez.
Arturo Eraso Torres.
Christian Camilo Angulo Mirama.
Santiago Álvarez Álvarez.
Rider Geovanny Benavides Bárcenas.
Elizabeth Del Rosario Guerrero. Arciniegas.
Berenice Castillejos López.
Carlos Arturo Torres Gastelú.
Héctor Alirio Guerrero.
Ricardo Javier Hernández Revelo.
Yuri Mercedes Bermúdez Mazuera.
Edwin Gamboa.
Luis Fernando Quintero Castaño.
Jorge Leonardo Zapata Castaño.
Andrés Mauricio Bueno Peña.
Javier Simón Naranjo Herrera.
Nelson Antonio Jaramillo Enríquez.
Miller Andrés Mera Inga.
Lisbeth Jurany Zemanate Zuñiga.
Gabriel Elías Chanchi Golondrino.
Katerine Márceles Villalba.
Jefferson Armando Peña Torres.
Víctor Buchelí.
Raúl Ernesto Gutiérrez de Piñerez
Fabio A. González O.
Manuel Pastrana.
Hugo Armando Ordóñez.
Sandra Guañarita.
Luisa Fernanda Londoño Celis.
Alfonso Joaquín Patiño Galarza.
Diego Fernando Criollo Estrada
Jairo Antonio Guerrero García.
Sandra Vallejo Chamorro.
Ronald Mauricio Gelvez Ruiz.
Luz Marina Santos Jaimes.
Mallory Julieth Muñoz Gutiérrez.
Liceth Marcela Domínguez Pinzón.
Daniela Pérez Oliveros.
Cesar Augusto Castelblanco Celis.
David Alejandro Albornoz.
Sebastián Alejandro Moncayo.
Samir Ruano Hoyos.
Gloria Amparo Thomé Ponce.
Giovanni Hernández Álvarez.
Álvaro Martínez.
Robinson Andrés Jiménez Toledo.
Franklin Jiménez.

Agradecimientos

Al apoyo incondicional a los procesos investigativos de Fray Daniel Omar Sarria Tejada OFM. Cap. de la Universidad Cesmag, así como también a la Sco2, la Universidad Mariana, la Universidad de Nariño, los autores y al personal de apoyo que hicieron posible la publicación de este libro.

Tabla de contenido

Prólogo

p. 11

A comparison between different artificial intelligence algorithms in the diagnostic of the diabetic retinopathy

Hector Fabio Ocampo Arbeláez
Carlos Andrés Delgado Saavedra

p. 12

Aerial Wide - Area Image Storage and Retrieval for Deep Learning Training

Oscar A Martinez
Alixon Franco
German A Holguin

p. 25

Análisis Patrones de Competencias Egresados Programa Ingeniería de Sistemas

Javier Alejandro Jiménez Toledo
Carlos Fernando González Guzmán
José María Muñoz Botina

p. 41

Aprendizaje por Descubrimiento Utilizando OVA en el Desarrollo del Laboratorio de Física Mecánica

José María Muñoz Botina
Fredy Santacruz
Carlos Fernando González Guzmán

p. 59

Aproximación metodológica para el desarrollo de sistemas interactivos basado en triadas de modelos

Alexandra Ruiz Gaona
William Joseph Giraldo
José Luis Arciniegas

p. 73

Comparación de controladores MPC y PID aplicados al proceso térmico de una máquina extrusora de poliestireno expandible

Javier Suarez Jaramillo
Mateo Córdoba
John Barco Jiménez
Camilo Lagos

p. 90

Creación de un algoritmo de cifrado simétrico denominado Sarch

Arturo Eraso Torres
Christian Camilo Angulo Mirama
Santiago Álvarez Álvarez

p. 107

Efectos blockchain como estrategia de incorporación con criptomonedas en mercados de capitales

Rider Geovanny Benavides Bárcenas
Elizabeth Del Rosario Guerrero Arciniegas

p. 132

El portafolio digital como medio para promover la identidad digital académica y profesional en estudiantes universitarios

Berenice Castillejos López
Carlos Arturo Torres Gastelú

p. 146

El software como una herramienta didáctica de las Matemáticas Universitarias

Alirio Guerrero
Ricardo Hernández

p. 161

EnglishPlay: Videojuego para el apoyo a la enseñanza del inglés

Carlos Andrés Delgado Saavedra
Yuri Mercedes Bermúdez Mazuera
Edwin Gamboa
Luis Fernando Quintero Castaño
Jorge Leonardo Zapata Castaño
Andres Mauricio Bueno Peña
Javier Simón Naranjo Herrera

p. 179

Estudio de caracterización de la Web Colombiana

Arturo Eraso Torres

p. 196

Evaluación del componente práctico en la enseñanza de Programación I en el programa de Ingeniería de Sistemas de la Universidad de Nariño

Manuel Ernesto Bolaños González
Nelson Antonio Jaramillo Enríquez

p. 213

Evaluación Heurística de usabilidad sobre el Sistema Gestor de Base de Datos (SGBD) de Sevenet de la industria licorera del Cauca (ILC)

Miller Andrés Mera Inga
Lisbeth Jurany Zemanate Zuñiga
Gabriel Elías Chanchí Golondrino
Katerine Márceles

p. 231

How to adapt deep learning models to a new domain: Biomedical relation extraction case

Jefferson Amado Peña Torres
Víctor Bucheli
Raúl Ernesto Gutiérrez De Piñerez
Fabio A. González O.

p. 248

Inception Deck un camino hacia el desarrollo de Sof Skills requeridas por el sector TI en Colombia

Manuel Pastrana
Hugo Ordoñez
Sandra Guañarita

p. 263

La Elicitación de Requisitos mediante la técnica Inception Deck en las empresas de Santiago de Cali

Manuel Alejandro Pastrana Pardo
Hugo Armando Ordoñez
Sandra Guañarita

p. 269

Marco de Referencia para el diseño de la Presentación de Localizaciones Geográficas

William Joseph Orozco Giraldo
Luisa Fernanda Londoño Celis

p. 276

PARLANA: Sistema Experto para determinar el nivel de inglés de un estudiante según el Marco Común Europeo de Referencia (MCER) para las Lenguas en los estudiantes de la Facultad de Ingeniería de la Universidad de Nariño.

Alfonso Joaquín Patiño Galarza
Dario Fernando Criollo Estrada
Jairo Antonio Guerrero García
Sandra Marleni Vallejo Chamorro

p. 291

Procedimiento para el Tratamiento de Datos Climáticos utilizando la herramienta Pentaho

Ronald Mauricio Gelvez Ruiz
Luz Marina Santos Jaimes

p. 306

Propuesta de un conjunto de heurísticas para el diseño de portales web para niños

Mallory Julieth Muñoz Gutiérrez
Liceth Marcela Domínguez Pinzón
Daniela Pérez Oliveros
Gabriel Elías Chanchí Golondrino
Katerine Márceles

p. 323

Sistema de Gestión del seguimiento del Progreso Académico

Cesar Augusto Castelblanco Celis
William JosephOrozco

p. 341

Sistema software para la ejecución de pruebas de usabilidad bajo el enfoque de mouse tracking

David Alejandro Albornoz
Sebastian AlejandroMoncayo
Samir Ruano Hoyos
Gabriel Elías Chanchí Golondrino
Katerine Marceles

p. 356

Un docente, enseña o educa: la ingeniería con pensamiento sistémico

Gloria Amparo Thomé Ponce
Manuel Ernesto Bolaños González
Nelson Antonio Jaramillo Enríquez

p. 369

Una aproximación a métricas de productividad para equipo en desarrollo ágil de software

Giovanni Hernández
Álvaro Martínez
Robinson Jiménez
Franklin Jiménez

p. 383

Prólogo

Avances y experiencias innovadoras en computación e informática es un libro resultado de investigación que recopila los principales estudios en temáticas como la Inteligencia Artificial, Deep Learning, minería de datos, Machine Learning, HCI, sistemas de control, algoritmos, Blockchain, informática educativa, videojuegos, pensamiento computacional, usabilidad, ingeniería de software, geolocalización, sistemas expertos, entre otras.

Los resultados de investigación presentados en este libro son consecuencia de los procesos realizados por la Sociedad Colombiana de Computación (Sco2) en su objetivo de contribuir con el fortalecimiento de la comunidad colombiana en computación, mediante la participación activa de estudiantes, profesores, profesionales e investigadores en el ámbito nacional e internacional.

La SCo2 organiza el Congreso Colombiano de Computación con el apoyo de las instituciones de educación superior desde el año 2005 en Cali en el marco del CLEI 2005, en Bogotá (2007), Medellín (2008), Bucaramanga (2009), Cartagena (2010), Manizales (2011), Medellín (2012) en el marco del CLEI 2012, Armenia (2013), Pereira (2014), Bogotá (2015), Popayán (2016), Cali (2017), Cartagena (2018) y Pasto (2019).

Este libro recopila mediante un riguroso proceso de selección los principales estudios de la edición 14ª del Congreso Colombiano de Computación (14CCC) realizada en la ciudad de San Juan de Pasto entre el 25 y el 27 de septiembre de 2019 por la SCo2, la Universidad Cesmag, la Universidad Mariana y la Universidad de Nariño.

Javier Alejandro Jiménez Toledo

A comparison between different artificial intelligence algorithms in the diagnostic of the diabetic retinopathy

Hector Fabio Ocampo Arbeláez (a)
Carlos Andrés Delgado Saavedra (b)

Una comparación entre diferentes algoritmos de Inteligencia Artificial para el diagnóstico de la Retinopatía Diabética

Resumen

El diagnóstico temprano de las enfermedades es un proceso crítico para evitar consecuencias irreversibles. Este trabajo se centra en analizar diferentes estrategias para el diagnóstico de la retinopatía diabética, basadas en información obtenida de imágenes procesadas del fondo de ojo. Esta enfermedad debe detectarse en una etapa temprana, ya que de lo contrario el paciente presentará pérdida irreversible de visión parcial o total. De acuerdo a diferentes estudios, el diagnóstico se puede mejorar con la ayuda de un sistema de apoyo a la decisión.

Para el diagnóstico de la retinopatía diabética, se utilizaron datos con 19 características de las imágenes procesadas del fondo de ojo, proporcionados por la base de datos Messidor. Este estudio se centra en el análisis de precisión de redes neuronales (RN) y máquinas de vectores de soporte (SVM), debido a que la revisión de literatura muestra que son las más utilizadas en el diagnóstico de enfermedades usando síntomas o datos obtenidos de exámenes médicos.

La metodología de comparación consiste en realizar varios experimentos con RN y SVM evaluando diferentes escenarios generados a partir de datos de personas sanas y con retinopatía diabética, observando la precisión y costo computacional de ambas técnicas.

Palabras clave: Inteligencia artificial, propagación de espalda, Retinopatía Diabética, enfermedades diagnósticas, Algoritmos Genéticos, Redes Neuronales, máquina de soporte vectorial.

Abstract

Early disease diagnosis is a critical process for it is essential in order to avoid irreversible consequences. This work focuses on designing and testing strategies for the categorization of diabetic retinopathy, based on information from processed ocular-fundus images. This disease must be detected at an early stage, since otherwise the patient will present partial or total loss of vision. In different studies, the diagnosis can be improved with the help of a decision support system.

For the diagnosis of diabetic retinopathy, we used data with 19 characteristics from the processed ocular-fundus images, provided by the Messidor database [6] and the project An ensemble-based system for automatic screening of diabetic retinopathy. Also, we selected Neuronal Networks and Support Vector Machines (SVM) as techniques due to the literature review of the artificial intelligence (AI) in the review, these two algorithms presented the best results in terms of accuracy.

Our methodology consists in comparing three AI strategies: Backpropagation neural networks, evolutive neural networks and SVM with linear kernel. The main result indicates that Backpropagation neural networks presents the best results in terms of accuracy but it has a considerable computational cost in training unlike other techniques.

Keywords: Artificial intelligence, BackPropagation, Diabetic Retinopathy, Diseases diagnostic, Genetic Algorithms, Neural Networks, Vector Support Machine.

I. INTRODUCTION

The early diagnosis of a disease is important in order to begin a proper treatment of it on time. A later diagnosis represents a high risk to the patient and possible irreversible consequences. Our case study is the diabetic retinopathy, a disease that appears as a consequence of diabetes. To be more specific, it is due to the retention of blood in the vessels of the retina, which is known as retinal ischemia. Our bodies try to compensate the problem by creating new blood vessels, which are easy to destroy and grow without control, contributing to the loss of the patient's vision [6]. This disease represents approximately the 2.6% of the blindness cases in the world [14].

Technology can support different disciplines including medicine. Medical informatics is a discipline that takes different elements from medicine and informatics, like the analysis of clinic histories, medical formal languages and AI, in order to create tools to support the decision making process concerning, for example, the disease diagnosis based on the analysis of a set of symptoms of a disease[15,3,4].

Medical experts often diagnose diseases based on symptoms a patient has and on different exams that are performed in the human body. However, different diseases may present a similar symptomatology, in that case, broad knowledge about it is required. Unfortunately, it is possible that the doctors don't have enough experience treating the disease [10]. Then, the diagnosis in some cases will not be correct, thus a decision support system could be required[1,9].

Diabetic retinopathy can be diagnosed by studying the anomalies of the vessels in the eye. It is done by a test known as ocular-fundus examination, which consists in taking an image of the entire eyeball. The diagnosis is based on studying whether there are exudates, which are deformed vessels, or if there is any micro-bleeding presen[7].

In the literature, it is possible to find different experiences regarding the diagnosis of diabetic retinopathy when using the Artificial Intelligence approach. We stand out the work of Salvador and others. They processed ocular-fundus images through the techniques of color decomposition, contrast enhancement and intensity inversion in order to extract features of the vessels structure. Then, they trained neural networks to detect bleeding or swollen vessels, their accuracy is 95% [13]. Antal and others used a approach based on features extracted from the output of several retinal image processing algorithms obtained from the Messidor Database[5] . After ,they determined the presence of the disease using a machine learning classifiers where Support Machine Vector (SVM) has the best results, they get with Receiver Operating Characteristic analysis(ROC) 90% in sensitivity, 91% specificity and 90% accuracy and 0.989 AUC. Some authors[8,16] indicate that SVM is the most appropriate technique for the diagnosis of this disease, but in literature we can found different AI strategies as diffuse logic[2], convulotional neural networks[12], evolutionary neural networks[11] with accuracy more than 80%.

In Section 2, the characteristics extracted from the ocular-fundus images are presented. In Section 3, we explain the approach of neural networks with training strategies with evolutionary algorithms and back-propagation. The SVM approach is explained in section 4, as well as the reasons why lineal and exponential kernels were selected. In section 5, the results are presented and a discussion about them is made. Finally, conclusions are presented in section 6.

II. DATA FEATURES

The Messidor database[5] provides 19 features extracted from a bank of 1200 ocular-fundus images. They are explained in 1.

Table. 1. *Features extracted from the eye-fundus images.*

Feature	Description
X_0	*Imagen quality (0 = bad; 1= good).*
X_1	*Diagnosis of the disease made before processing, (1 = several RD o 0 = not several RD).*
$X_2 \ldots X_7$	*Integer that determines the number of micro-aneurysms detected, with a confidence level of 50% to 100 %, respectively.*
$X_8 \ldots X_{16}$	*Integer that represents number of pixels of exudates, with a confidence level of 20% to 100%, respectively.*
X_{17}	*Relative value that determines the distance from the center of the macula to the center of the optical disc.*
X_{18}	*AM/FM Classification.*

In order to understand the data, an analysis about it has been made. The figure 1 shows a scatter diagram of all the parameters, where you can see which data have maximum and minimum values. For example, X_8 which represents the number of pixels of exudates with a confidence level of 20%, is the parameter that has the highest value when compared to all others. Another observation that can be made is that the parameter set $X_2 \cdots X_7$, which represents the microaneurysms, is decreasing in a linear way as the percentage of confidence increases and the parameter set $X_8 \cdots X_{16}$ that represents the exudates, decreases exponentially as the confidence percentage increases.

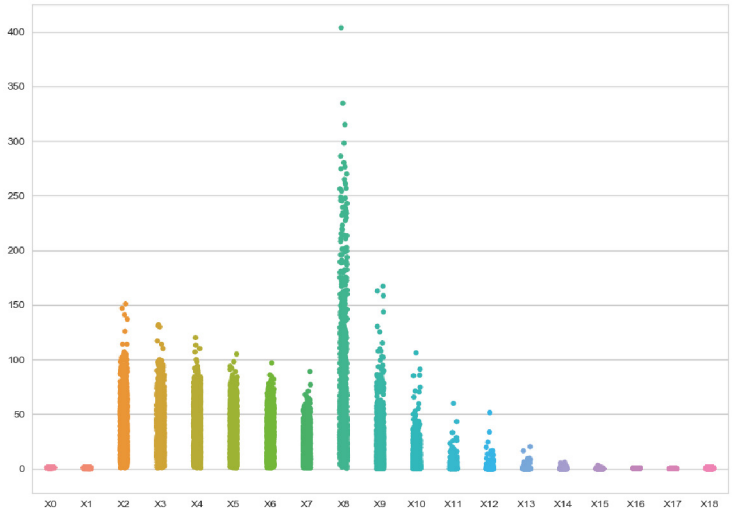


Figure 1. *Scatter diagram of the data.*

III. NEURAL NEWTWORK APPROACH

3.1 Backpropagation algorithm

We used a neuronal network with a hidden layer varying the number of neurons in that layer between 9 and 29. We also carried out tests with two and three layers, but no improvement was found in the results.

3.2 Training with a genetic algorithm

For the implementation of the evolutionary algorithm, it is important to determine the genotype and phenotype of the chromosome and how it will be represented computationally for its execution. The phenotype will be the representation of the weights and bias matrices used in the Neural Network.

In neural networks, there was the need to determine the form of the weights in the hidden layer (HLW) and in the output lawyer (OLW). These are two matrices with size M number of features plus N number of neurons in the hidden layer

$$HLW(M * N) = \begin{bmatrix} W_{11}\ W_{12}\ W_{13}\ W_{14}\ W_{15} \dots\ W_{1N} \\ W_{21}\ W_{22}\ W_{23}\ W_{24}\ W_{25} \dots\ W_{2N} \\ . \\ . \\ . \\ W_{M1}\ W_{M2}\ W_{M3}\ W_{M4}\ W_{M5} \dots\ W_{MN} \end{bmatrix}$$

$$OLW_{(1*M)} = [P_{11}\ P_{12}\ P_{13}\ P_{14}\ P_{15} \dots\ P_{1M}]$$

The genotype or chromosome (Crh) is an array with size:

$$tamano = M * N + M + 2 \qquad (1)$$

Which has the following indexation:

- The first $M * N$ positions represent the weight matrix of the hidden layer.
- The following M positions is the weights of the output layer.
- The position before the last position is the bias of the hidden layer.
- The last position is the bias of the output layer.

$$Crh = [W_{11} \dots W_{1N} \dots W_{MN}\ P_{10} \dots P_{1M}\ B1\ B2]$$

IV. SVM APPROACH

It is a technique that separates data with hyperplanes, it allows to categorize them. SVM uses support vectors that represent the position of the hyperplanes.

The hyperplane can behave in different ways depending on the function of the Kernel and can be linear, exponential, polynomial, sigmoidal, among others.

To decide which Kernel to implement in the algorithms, the information of the data source must be analyzed and its behavior must be observed. Our data has 19 parameters which are analyzed below.

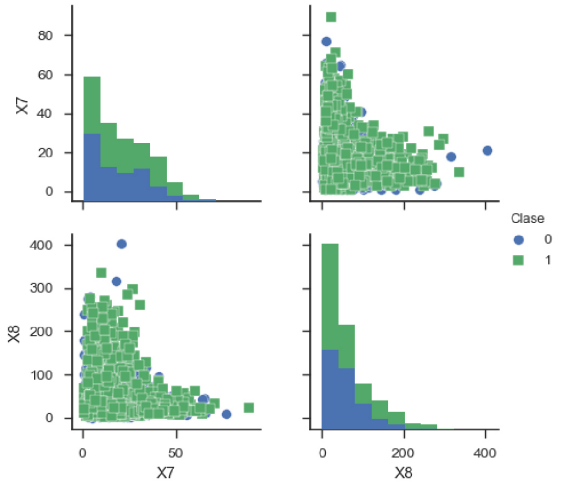


Figure 2. *Diagram of comparison of features X_7 and X_8, we found negative correlation*

Figure 2 shows that the two variables have a descending exponential behavior. This translates into a clue with which it can be inferred that an exponential can be used for classification.

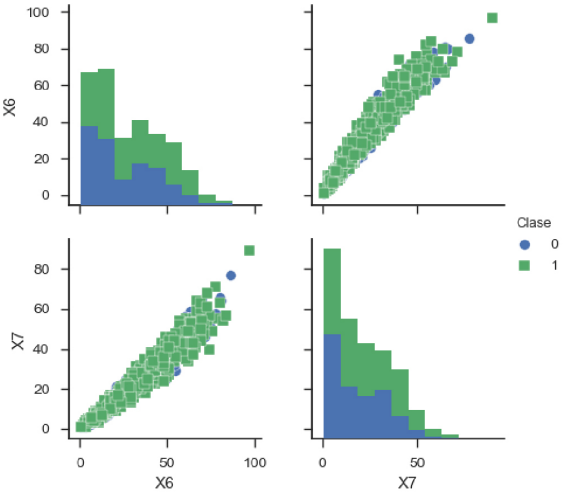


Figure 3. *Diagram of comparison of features X_6 and X_7, we found positive correlation*

In order to finish with the comparison images, there is the figure 3, which evidently shows a high linear behavior between the two variables, validating that the correlation between these two variables is positive and, at the same time, leaving an indication that A linear kernel can be used. Analyzing the results of the figures 2 and 3 we can conclude that a linear kernel and an exponential kernel are appropriate strategies to solve the problem.

Analyzing the results of the figures 2 and 2 we can conclude that a linear kernel and exponential kernel are appropriate strategies to solve the problem.

V. RESULTS AND DISCUSSION

5.1 Backpropagation

The tests were performed by varying the number of neurons in the hidden layer between 9 and 29. To analyze the behavior of the configurations, the mean square error (ECM) is plotted through the iterations of the training, thus obtaining the figure 4, which shows that the configuration of 19 neurons is the one that provides the best result. It is interesting to note that it is the same number of features.

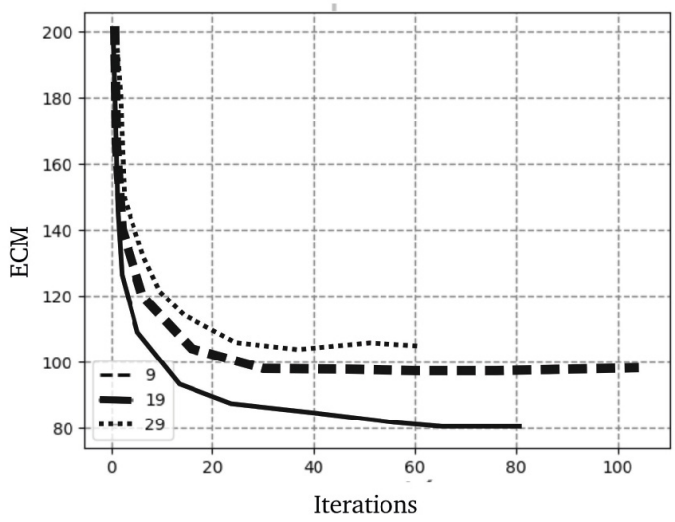


Figure 4. *Accuracy of neural networks with 9, 19 and 19 neurons in hidden layer.*

5.2 Genetic algorithms

For the tests of the genetic algorithms, it is required to vary some parameters as the initial population, mutation percentage, if scaling is used or not, among other. Then it can generate a very large number of combinations that make the tests non-viable,

therefore these must be limited to generate a more discrete set of combinations between Datum 1 and Datum 2, which are shown in the Table. 2.

Table. 2. *Combinations of parameters in neural networks with genetic algorithms*

Parameter	Datum 1	Datum 2
Initial population	*50*	*100*
Replacement	*Insertion*	*Elite*
Mutation percentage	*8%*	*20%*
Scaling	*Yes*	*No*

A total of 32 different configurations were analyzed, from which the 4 of them that have the best results are selected. Table. 3 shows the corresponding parameters.

Table. 3. *Parameters of the evolutionary algorithm tests*

Test number	Initial population	Mutation percentage	Replacement	Scaling
1	*50*	*8%*	*Insertion*	*No*
2	*50*	*8%*	*Elite*	*No*
3	*50*	*20%*	*Insertion*	*No*
4	*50*	*20%*	*Elite*	*No*

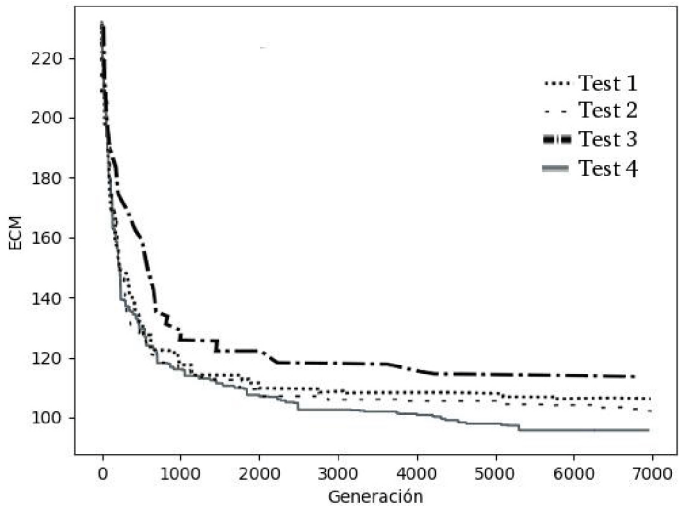


Figure 5. *Accuracy of the neural networks with evolutionary algorithms.*

5.3 SVM

For the SVM tests we use the linear and exponential kernels.

Table. 4. *SVM tests with linear and exponential kernel.*

Kernel	Precision	Sensitivity	Specificity
Lineal	*72.7%*	*61.5%*	*87.7%*
Exponential	*53.3%*	*51.6%*	*20.9%*

After analyzing the information obtained, it can be inferred that many of the characteristics in the data source behave linearly.

5.4 Discussion

The figure 6 shows the curves that correspond to each of the three AI strategies: neural networks with propagation, neural networks with genetic algorithms and SMV. The curve shows that the model of Backpropagation as the best with respect to the other two with an accuracy of 94%.

According to the computational costs of the different strategies, backpropagation requires more computing time than the other two, however it is the one that has the best results.

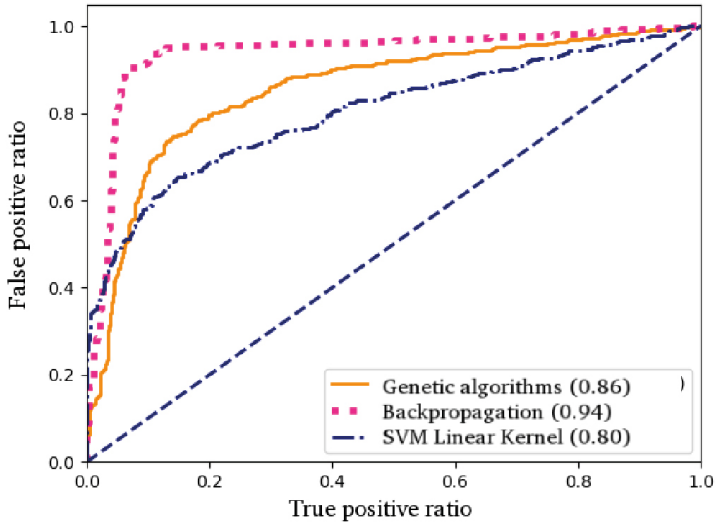


Figure 6. *ROC analysis between the three AI strategies.*

V. CONCLUSIONS

SVM training is 60 times faster than neural network training using Backpropagation and evolutionary algorithms, since, while one takes 5 minutes, the other takes approximately 5 hours. This was evidenced during the tests carried out and is due to the fact that the Backpropagation strategy requires a large amount of time of computation, while the evolutionary strategy has a slow convergence speed.

When performing the tests some important patterns were observed: the resulting Backpropagation models always converge to the fact that the hidden layer weights are in a range of numbers of [-1,1] and the weights of the output ayer are between [-7,7], that compared to the models resulting from evolutionary algorithms, the weights show a great variation. With this you avoid staying in local optimals and you can perform a better search in the solutions space.

Backpropagation reached the best accuracy with 94%, however, this technique depends a lot on the initial model, which implies a possible disadvantage, since being linked to the initial model does not explore the solution domain above all and ends up being generally on local optima. Still, it can be said that Backpropagation gives answers with good precision and useful models for decision support systems.

The genetic algorithms obtained a lower precision with 86%, but given the quality of having a great variability it is possible to explore more the search spaces and not to fall in local optima, besides having a high stochastic component it can be said that the algorithms Genetics are closer to reaching a precision of 100%, since it can not be determined that it will be found across generations.

SVM obtained the lowest accuracy with 80 %, but this has a great advantage compared to the other two, which is that SVM is deterministic, which implies that it does not take many tests to generate a model that can be used in decision support systems.

When deciding which technique to use in a decision support system, 3 environments can be generated: the first one would be where the best precision is needed in a moderate time, here the decision would be to implement Backpropagation since it corresponds to the best precision. In second place is an environment where a quick implementation is needed, since you do not have time for many tests and you can sacrifice a bit the accuracy of the system, here the decision would be SVM. Finally, it would be a research environment in which there is a lot of time for testing, here the decision could be genetic algorithms, in which a large number of tests can be generated, with different configuration options for the model.

According Meyer and others[10], medical diagnoses can be affected by variables such as overconfidence, showing that the diagnostic accuracy is 55.3%, which compared to the accuracy obtained in the current project is lower. In addition, Carrion and other[1] the accuracy of professionals can be improved by using a decision support system. From the above it can be concluded that the algorithm of Neural Networks trained by Backpropagation can be used by decision support systems, so that this will serve as professional help to improve its accuracy or compare its own opinion with the system and have another point of view.

REFERENCES

[1] *Alonso-Carrión, L., Sánchez-Lora, F.J., Hinojosa-López, C., Porras-Vivas, J.J., Cuenca-Guerrero, A., Abarca-Costalago, M.: Precisión del diagnóstico en medicina interna e influencia de un sistema informático en el razonamiento clínico. Educación Medica 16(3), 184-187 (2015). https://doi.org/10.1016/j.edumed.2015.09.012, http://dx.doi.org/10.1016/j.edumed.2015.09.012*

[2] *Barkana, B.D., Saricicek, I., Yildirim, B.: Performance analysis of descriptive statistical features in retinal vessel segmentation via fuzzy logic, ANN, SVM, and classifier fusion. Knowledge-Based Systems 118, 165-176 (2017). https://doi.org/10.1016/j.knosys.2016.11.022, http://dx.doi.org/10.1016/j.knosys.2016.11.022*

[3] *Coiera, E.: Guide to medical informatics, the internet and telemedicine. Chapman & Hall, Ltd. (1997)*

[4] *Cruz-Gutiérrez, V., Sánchez-López, A.: Un sistema experto difuso en la Web para diagnóstico de diabetes. Research in Computing Science 107, 145-155 (2015), http://www.rcs.cic.ipn.mx/2015 107/Un sistema experto difuso en la Web para diagnóstico de diabetes.pdf*

[5] *Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., Charton, B., Klein, J.C.: Feedback on a publicly distributed database: the messidor database. Image Analysis & Stereology 33(3), 231-234 (Aug 2014). https://doi.org/10.5566/ias.1155, http://www.iasiss.org/ojs/IAS/article/view/1155*

[6] *Fauci, A.S., et al.: Harrison's principles of internal medicine, vol. 2. Mcgraw-hill New York (1998)*

[7] *Klein, R., Klein, B.E.K., Moss, S.E., Davis, M.D., DeMets, D.L.: The Wisconsin Epidemiologic Study of Diabetic Retinopathy: II. Prevalence and Risk of Diabetic Retinopathy When Age at Diagnosis Is Less Than 30 Years. Archives of Ophthalmology 102(4), 520-526 (04 1984). https://doi.org/10.1001/archopht.1984.01040030398010, https://dx.doi.org/10.1001/archopht.1984.01040030398010*

[8] *Kumar, S.J.J., Madheswaran, M.: An improved medical decision support system to identify the diabetic retinopathy using fundus images. Journal of Medical Systems 36(6), 3573-3581 (2012). https://doi.org/10.1007/s10916-012-9833-3*

[9] *Kunhimangalam, R., Ovallath, S., Joseph, P.K.: A clinical decision support system with an integrated EMR for diagnosis of peripheral neuropathy. Journal of Medical Systems 38(4) (2014). https://doi.org/10.1007/s10916-014-0038-9*

[10] *Meyer, A.N., Payne, V.L., Meeks, D.W., Rao, R., Singh, H.: Physicians' diagnostic accuracy, confidence, and resource requests: A vignette study. JAMA Internal Medicine 173(21), 1952-1961 (2013). https://doi.org/10.1001/jamainternmed.2013.10081*

[11] *Mookiah, M., Acharya, U.R., Martis, R.J., Chua, C.K., Lim, C., Ng, E., Laude, A.: Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: A hybrid feature extraction approach. Knowledge-Based Systems 39, 9 - 22 (2013). https://doi.org/https://doi.org/10.1016/j.knosys.2012.09.008, http://www.sciencedirect.com/science/article/pii/S0950705112002651*

[12] *Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P., Zheng, Y.: Convolutional neural*

networks for diabetic retinopathy. Procedia Computer Science 90, 200 - 205 (2016). https://doi.org/https://doi.org/10.1016/j.procs.2016.07.014, http://www.sciencedirect.com/science/article/pii/S1877050916311929, 20th Conference on Medical Image Understanding and Analysis (MIUA 2016)

[13] Rosales-silva, A.J., Gallegos-funes, F.J.: Detection and classification of NonProliferative Diabetic Retinopathy using a Back-Propagation neural network Detección y clasificación de Retinopatía Diabética no Proliferativa usando una Red Neuronal de Retropropagación. Revista Facultad de Ingeniería Universidad de Antioquia pp. 70-85 (2015)

[14] mundial de la salud, O.: Diabetes, datos y cifras mundiales. http://www.who.int/mediacentre/factsheets/fs312/es/ (2017), última visita 9 mayo del 2017

[15] Sikchi, S.S., Sikchi, S., Ali, M.: Artificial intelligence in medical diagnosis. International Journal of Applied Engineering Research 7(11), 2012 (2012)

[16] Taveira-Gomes, T.: machine learning on the diabetic retinopathy debrecen data set data set. https://rpubs.com/tiagotaveira/debrecen

Biografía Autor(es)

a. Ingeniero de Sistemas. Docente, Universidad del Valle. E-mail: hector.ocampo@correounivalle.edu.co. Cali, Colombia.

b. Magister en Ingeniería con énfasis en Ingeniería de Sistemas. Docente, Universidad del Valle. E-mail: carlos.andres.delgado@correounivalle.edu.co. Cali, Colombia.

Aerial Wide - Area Image Storage and Retrieval for Deep Learning Training

Oscar Andrés Martinez (a)
Alixon Franco Salazar (b)
Germán Andrés Holguín Londoño (c)

Almacenamiento y consulta de imágenes aéreas de área amplia para el entrenamiento de modelos de aprendizaje profundo

Resumen

En el contexto de la agricultura de precisión, la habilidad de procesar imágenes que cubren vastas extensiones de tierra es crucial para el análisis, diagnóstico y detección temprana de problemas que afecten a los cultivos y la seguridad alimentaria en general.

En este trabajo presentamos un método para el almacenamiento, procesamiento y consulta de imágenes aéreas de área amplia. El propósito principal del sistema es permitir que las imágenes estén disponibles para el entrenamiento de algoritmos de aprendizaje profundo. Cuando el área de interés es extensa, el conjunto de imágenes capturadas por Drones no cabrán en la memoria de una máquina convencional; por lo tanto el sistema propuesto utiliza un motor de base de datos relacional compatible con el modelo de almacenamiento de datos distribuido. Las imágenes se capturan con varias cámaras utilizando uno o múltiples UAV's, estas imágenes se guardan junto con sus metadatos, como son las coordenadas GPS y las marcas de tiempo. La consulta de imágenes se puede hacer por coordenadas GPS o por etiquetas que pueden ser creadas por un algoritmo de aprendizaje profundo o por un etiquetador humano.

El sistema propuesto responde a las consultas del usuario y entrega las imágenes aledañas a un punto determinado con la corrección fotogramétrica para fines de alineación. En la sección de resultados se muestra la efectividad del método y se presentan algunos ejemplos de las imágenes obtenidas ante algunas consultas.

Palabras clave: Almacenamiento masivo de imágenes, recuperación de imágenes, imágenes de área amplia.

Abstract

In the context of precision agriculture, the ability to process images covering vast amounts of land is considered crucial to the analysis, diagnostic, and early detection of problems affecting crops and food safety in general. In this work, we present a method for the storage, processing, and query of aerial wide-area images. The primary purpose of our system is to make these images available for the training of deep learning algorithms. When the area of interest is extensive, the set of images will not fit into the RAM of a single machine. Therefore, our method uses a relational database engine compatible with the distributed data warehouse model. Images are captured with possibly several cameras using one or more UAVs. Images are stored in the database along with their metadata such as GPS coordinates and timestamp. The retrieval can be done by issuing queries by GPS coordinates or by labels created either by a deep learning algorithm or a human tagger. The output image is composed of the stitching of all available neighboring raw image patches. Our system responds to the user query by retrieving the requested wide-image section, properly ortho-corrected for photogrammetric purposes. In the results section, we show the effectiveness of our method and present some retrieval examples.

Keywords: Image Massive Storage, image retrieval, wide-area images.

I. INTRODUCTION

Precision agriculture is an umbrella term covering a large set of methods used for the optimization of productive processes inside farms [18], [7]. Process optimization allows for the reduction in resources, increased quality, and larger profit margins, in almost every kind of crop such as fruits, and vegeTable.s. Several studies, [8], [28], [21], [3] show how the global demand for food has been constantly increasing due to the increase in the population, however, the total area of land dedicated to crops is limited, hence its efficiency must be maximized.

1.1 Problem Background

One of the essential methods to track and control the performance of a crop is the use of aerial images [11]. More recently, with the popularization of UAVs (Unmanned Aerial Vehicles), aerial images are becoming ubiquitous [24], [5]. As opposed to satellite images, UAVs images can be obtained at a greater frequency, with much higher resolution, and at a fraction of the cost. UAVs are more flexible to program and automate, and they are easily deployed as stand-alone devices or in clusters [23], [25]. Furthermore, satellite images require perfect atmospheric conditions, and it takes time before a suiTable. frame can be shot [20].

Traditionally, agricultural land covers a large area [16], and the whole image of such area is known as a wide-area image. Capturing high resolution images of wide-areas constitute a challenge for the memory management of those devices in charge of store and process the collected data [15], [9], [19]. Several data structures and storage methods have been used in the literature, such relational and non-relational databases, local storage arrays, or any other ad-hoc architectures, [13], [12], [10], [1]. An aerial wide-area image is a composition, or stitching, of single snapshots, or tiles, taken from UAVs around a particular area of interest. The resulting image is orthorectified for photogrammetric purposes.

The system needs to be designed to allow query images by it's GPS coordinates. The problem with querying a non-hashed field such as GPS coordinates is the complexity. A naive implementation of the GPS search might take up to $O\ n^2$ which is not suiTable. for an extensive set of images. This problem is similar to retrieving images by its content, which is known as Content-Based Image Retrieval CBIR. In [14] and [2] authors propose a data structure based on RTrees and/or KD-Trees. In [27] authors use Locally Sensitive Hashing to reduce the complexity of the search. Hashing is also used in other works such [26].

To reduce the complexity, we will show in section 2.2, that our methods follow a similar strategy as in [17], and [1] where authors use a Hilbert SpaceFilling Curve in other to transform the space of the coordinates into a linearly searchable space.

1.2 Hilbert Space-Filling Curves

Hilbert space-filling curves (HSFC) are a mapping curve from $R^{NxN} \longrightarrow R$.

Assuming a squared space and splitting it into four equal squares, space can be filled by walking through the four centroids of the subdivisions. This is shown in figure 1.

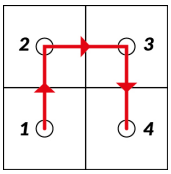


Fig. 1. *First iteration of HSFC.*

Recursively, it subdivision can be further divided into another four spaces. In order to connect adjacent spaces, they can be rotated 90 degrees each time. Figure 2 and figure 3 show iterations second and third respectively. The order of an HSFC is given by the number of performed iterations.

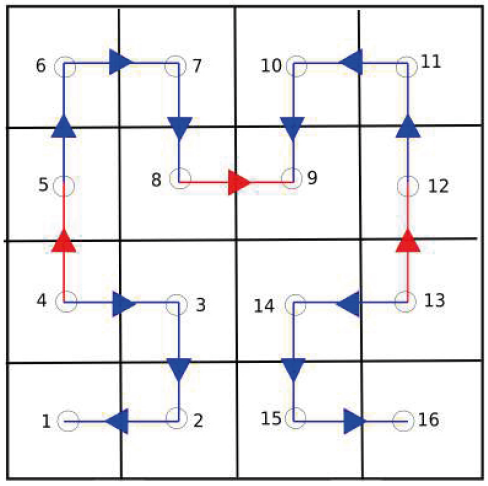


Fig. 2. *Second iteration of HSFC*

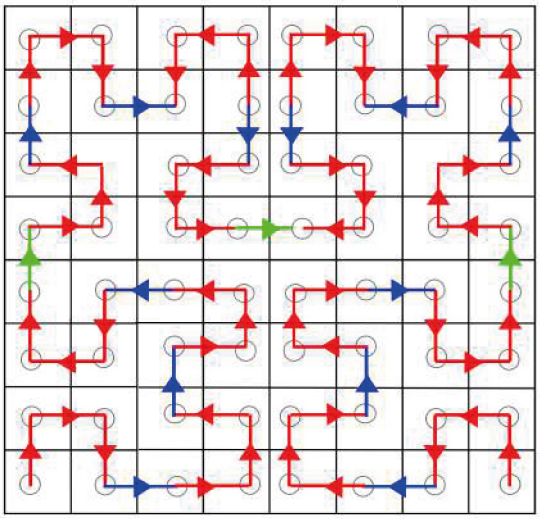


Fig. 3. *Third iteration of HSFC*

This curve represents an ordering of the data that can be linearly searched. It has been shown that, [22], [1], [17], Hilbert space/filling curves (HSFC) are an efficient method since searching can typically be done in $O\ (log\ n)$, which is the typical complexity of searching in a sorted linear space.

Mapping points from a plane ω to a n-order HSFC requires an affine transformation that can be easily achieved by partitioning ω in 2^n squared intervals.

There are four combinations to form the sequence of the HSFC that can be represented in the equivalent matrix form as can be seen in equations (1) ,(2), (3), and (4), where $(x1, x2)$ are the complex vector representation of a point in the unit square, and Q_n represents the respective subsquare for mapping.

$$Q0\begin{pmatrix} x1 \\ x2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x1 \\ x2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (1)$$

$$Q1\begin{pmatrix} x1 \\ x2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x1 \\ x2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (2)$$

$$Q2\begin{pmatrix} x1 \\ x2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x1 \\ x2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (3)$$

$$Q3\begin{pmatrix} x1 \\ x2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} x1 \\ x2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad (4)$$

1.3 Proposed System

In this work, we propose a general system to capture, process, store, and retrieve wide-area images from agricultural land with the purpose of analyzing, diagnose, and/or classify different performance and health metrics in the crop. Stored images may also contain meta-data such GPS coordinates, timestamps, and other spectral channels such as NIR (Near Infra-Red), [4], [6]. The vast amount of information available will allow our system to be used by automated deep learning multi-class classifiers. Farmers will be able to access data in near real-time to determine the state and health of their crops.

In the rest of the paper, we will describe our system in section 2. The experimental set up is explained in detail in section 3.1, results are presented in section 3.2. Finally, conclusions are presented in section 4.

II. METHODS

In this section, we start by describing our system for managing wide-area images. Our system is composed of three main functionalities: storage, image retrieval, and training.

2.1 Storage

We define the storage functionality, as the set of methods and algorithms that allow images to be added to the system. Images are captured by UAVs following a programmed flight path that guarantees at least a 30% overlap between neighboring images. Images in the scope of this project are considered to be multispectral, since they include, in addition to the regular RBG channels, at least the NIR channel. Images include also their respective GPS coordinates. The widearea image is composed of hundreds, or

even thousands, of these snapshots, so in general, they cannot be stored in a single computer. Furthermore, since more than one UAV can be taking snapshots, our system uses a distributed, SQL-like, data warehouse model, that provides the required flexibility and robustness. Images are stored in the HDFS file system provided by the data warehouse. Figure 4 is a schematic depiction of the capture and storage system.

There is an SQL Table. hosted by the data warehouse. Every image creates a record in this Table.. Most relevant fields in the Table. are: 1) link to the image file in the HDFS, 2) GPS coordinates, 3) timestamps, 4) Hilbert curve index, 5) Hilbert linked-list header, 6) Hilbert linked-list tail, 7) classification tags, 8) some descriptors, among others.

The capture and storage system, is run in a local computer. This system has been implemented in Python. The main functionality provided by this module consist of: 1) download images from the drone, 2) allows to load images into the data warehouse using FTP and creating their respective records, 3) allows to query the data warehouse, and 4) allows to delete or overwrite images from the data warehouse and their respective records in the SQL Table..

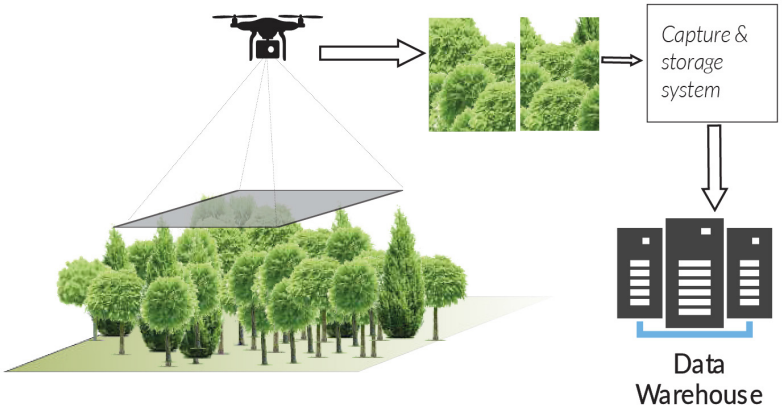


Fig. 4. *An Schematic for the capture and storage system*

2.2 Image Retrieval

The second main functionality in our system is the image retrieval. This is perhaps the most important feature in our system. As we mentioned before, the retrieval can be efficiently done using a Hilbert Space-Filling Curve (HFSC). An HSFC is a line running over an entire N-dimensional space passing through each point, only once. Storing and retrieving images by their GPS coordinates will be as complex as searching in a 1-D array of elements. The advantage with this technique is that neighbors with similar GPS locations are in the same section of the array, making image retrieval by GPS coordinates or certain tags more efficient. This is, using the HFSC, the query can be accomplished in $O\ (log\ n)$ as opposed to $O\ (n^2)$ as in a 2-D linear search.

In order to construct an HSFC, it is required to design the curve order beforehand. Such order must avoid collisions and clearly separate each image from its neighbors. In our system, we know the flight plan of the UAV beforehand. Therefore, we choose an HSFC order of about two orders of magnitude the number of expected snapshots in the wide-area image. This way, there will be an index in the curve for every GPS coordinate in the wide-area image. These GPS coordinates are known as the centroids.

Subsequently, a mapping function in quaternions is defined. Therefore, all GPS coordinates will transform into the HSFC. Since adjacency is invariant to the mapping, the resulting HFSC index will be close to the actual position. Querying for a non-centroid GPS coordinate will be as complex as searching in a sorted 1-D array, $O(\log n)$.

Therefore, a wide-area image, is an ordering of individual snapshots over a Hilbert Space-Filling Curve. There are three main cases to consider during the queries by GPS: 1) when the resulting image window is within a single image file, 2) when the resulting image window is composed by parts of contiguous images, and 3) when the resulting image window is composed by several images. For the first case, after the correct image is found, the only remaining step is the ortho-normalization of the image. However, for the second and third cases, a panoramic photo stitching is also required.

2.3 Processing workflow

Figure 5 shows the block diagram describing the behavior of the system when there is a coordinate query. The first step is to receive the data, which is, a GPS coordinate and a squared window. Input coordinates are converted to UTM system and normalized for processing. The second step is to convert the normalized UTM coordinates into an HSFC's index, mapping 2d coordinates into a 1d array and preserving locality.

The third step is to search the nearest index obtained in step 2 on to the stored HSFC array. Because the array is sorted, searching the nearest value complexity is $O(\log(n))$.

Once the nearest index is retrieved, Image with queried GPS is located and downloaded finishing the process.

If the window is bigger than the image size, then the system will query for the adjacent image. This query is easily achieved by fetching the image in the next and previous HSFC index and validating that images are neighbors by finding and matching key points.

Finally, retrieved images are stitched and the result is presented to the user.

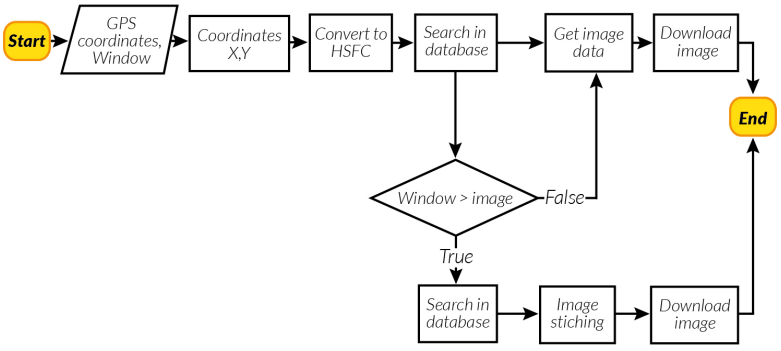


Fig. 5. *Processing workflow*

2.4 Training

Images in the database also contain tags. Those tags are basically string-based labels assigned to each snapshot and are based on the image content. Many classification

algorithms have been studied in order to discriminate aerial image content. As we mentioned before, the main goal in this project is to allow a Deep Learning algorithm, to access the database in an efficient way. Once the HSFC is created, an external tool can crawl the database efficiently in O (n). Although the focus of this paper is the capture and storage system, it is worth mentioning that the external system we use to tag snapshots can also be used by human taggers, as in human-in-the-loop systems. Once tags are in place, they can be organized similarly as the GPS coordinates using their own HSFC in order to allow for queries by tag.

III. RESULTS AND DISCUSSION

3.1 Experimental Setup

For experimental purposes, a land area of 20000 m2, from a semi-rural area in the city of Pereira, Colombia, was photographed using a commercial drone with a regular RGB camera. The land was divided into 159 centroids, each one corresponding to an image of 5472 × 3648 pixels taken from a 50 m height. The centroids for the capture are shown in figure 6.

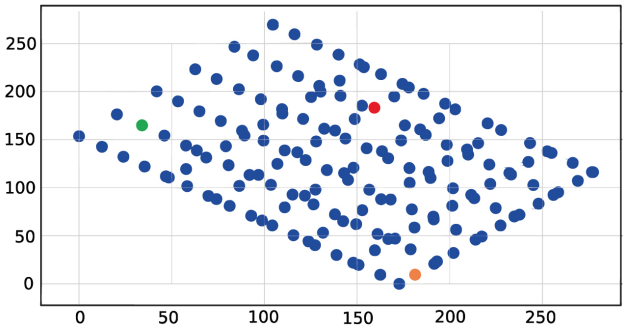


Fig. 6. *Centroids for land image capture*

For demonstration purposes, corresponding images for colored points in figure 6 are shown in figure 7.

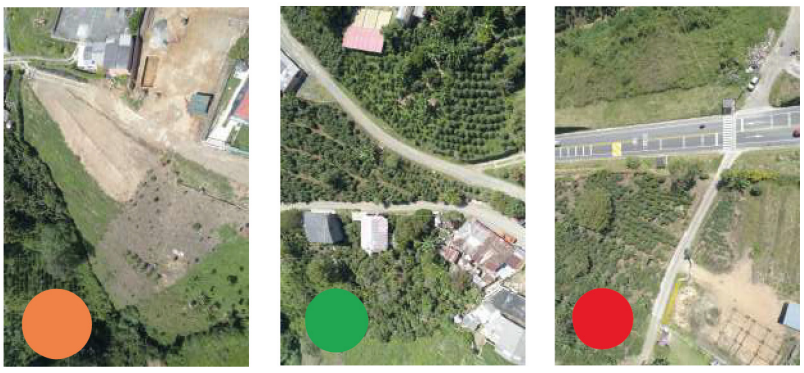


Fig. 7. *Aerial images corresponding to GPS marked points*

GPS coordinates altitude, latitude, and longitude need to be transformed into Universal Transverse Mercator System (UTM) in order to obtain a planar coordinate system, similar to the Cartesian but keeping distances between points.

This data set is used to test the storage and retrieval system. Randomly generated GPS coordinates within the scope of the area of interest are then used to test for querying. However, this data set is still small to test the performance of the query algorithm. Therefore, we augmented the GPS data set with interpolated coordinated within the wide-area image, allowing us to perform a large-scale experiment.

The experiment consists of measuring execution time while querying the database for GPS coordinates regardless of their condition as centroids in the HSFC. This execution time is then compared against a method that finds the centroid with the minimum distance to the query.

3.2 Results

System for wide-area image storage and retrieval is tested by loading two datasets with aerial images from different locations. Each dataset contains 120 aerial images. For test purposes, images data Table. in the storage system is populated with 100.000 GPS points interpolated from current dataset.

Figure 8 show the behavior of both methods when the size of the problem increases. In both figures, the x-axis is the number of GPS coordinates in the dataset, which in normal conditions will be equal to the number of images in the database. The y-axis is the average time of 20 queries to the system. It can be easily seen how the HSFC outperforms the minimum distance method.

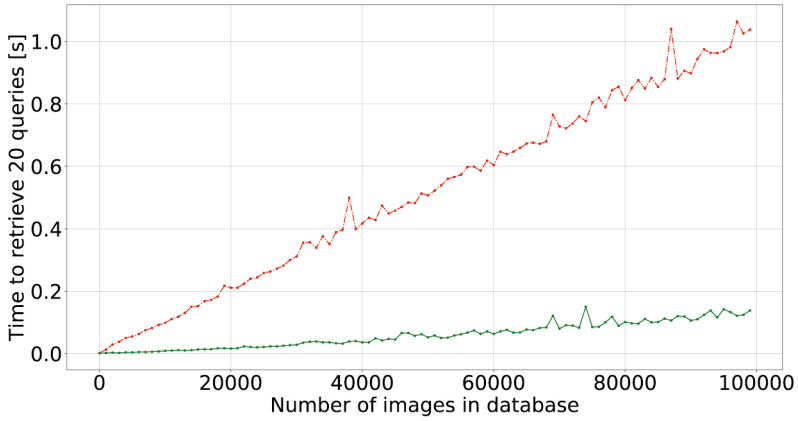


Fig. 8. *Results of near GPS finding query*

When several images are needed in order to create a panorama view, the system downloads the neighboring images by consulting database with the previous and the next HSFC indexes already retrieved by the current query. Figure 9 shows images in adjacent HSFC indexes, queried by using "next index" field in Database for each image.

Figure 10 shows the process of key points matching between two neighboring images in our dataset. Matches are found using SIFT descriptors and RANSAC for outlier suppression. Then, a panorama composition is created, as it is shown in figure 11.

Fig. 9. *Query with adjacent indexes*

Once images are stitched, there can be built a bigger panorama image, in order to get a wide area image. Figure 11 shows four retrieved images stitched. In figure 12 wide image can be seen by retrieving 14 images from the database. For better results is needed lens correction in order to get images with fixed lens distortion.

Fig. 10. *Matching keypoints in two neighbouring images.*

3.3 Discussion

This work has been developed in order to address the complexities of handling a wide-area image composed by a very large set of single snapshots. Due to the size of this particular kind of sets, it is not possible to handle these images in local massive storage systems. Therefore, we proposed a whole system that can reside in the framework of a

distributed data warehouse. Experiments have shown that the HSFC is a significant improvement over a conventional minimum distance search for answering queries about GPS coordinates, or tags, or even other features in the image space. Once the query is resolved, the output for the system is a panoramic composition resulting from the stitching and orthocorrection of several snapshots. Experiments have also shown the correctness of these procedures.

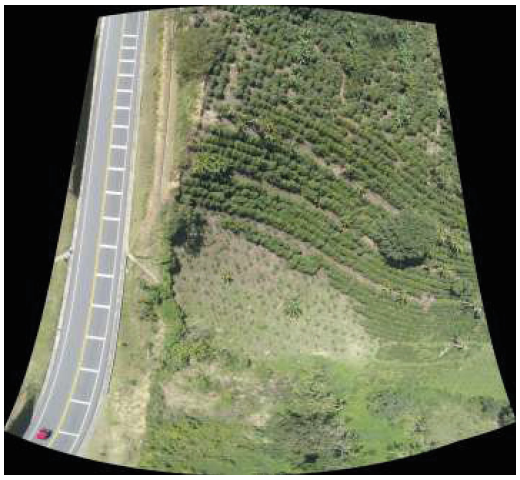


Fig. 11. *Panoramic composition of two neighbouring images.*

Fig. 12. *Final panoramic composition of several neighbouring images.*

IV. CONCLUSIONS

A system for the efficient management, storage, and retrieval of a wide-area image has been presented. Our methodology allows for the storage of multiple photogrammetric snapshots belonging to the same wide-area. One or multiple UAVs have captured these individual images. Our system is intended for coordinated use with machine learning and specifically deep learning algorithms. For that reason, we have presented a system that behaves efficiently in retrieving images by either GPS coordinate or tags.

Results in image retrieving for panorama construction showed the locality preservation by Hilbert curves, the ordered indexes allow to download overlapping images in $O(1)$ complexity.

We have shown how our system is able to retrieve an image composition of interest, around a particular GPS coordinate, using the Hilbert transformation, achieving an efficient time complexity of $O(\log n)$. It is also shown that the same technique is useful in content-based image retrieval.

Our near future work includes testing different methods for tagging images. This work is intended for the coordinated use with deep learning external systems. We pretend to extend an experiment base on deep learning algorithms and other methods for tagging including, but not limited to, active learning and/or human in the loop classifiers.

V. ACKNOWLEDGEMENT

This work is part of the project 6-19-6 funded by Vicerrectoría de Investigaciones, Innovación y Extensión, and the support from the Electrical Engineering Department, undergraduate and master's programs, at Universidad Tecnológica de Pereira.

REFERENCES

[1] *Akgun, T., Altunbasak, Y., Mersereau, R.M.: Super-resolution reconstruction of hyperspectral images. IEEE Transactions on Image Processing 14(11), 1860-1875 (Nov 2005). https://doi.org/10.1109/TIP.2005.854479*

[2] *and, J.H.: A hybrid tree approach for efficient image database retrieval with dynamic feedback. In: Object recognition supported by user interaction for service robots. vol. 1, pp. 263-266 vol.1 (Aug 2002). https://doi.org/10.1109/ICPR.2002.1044680*

[3] *Arnholt, M., Batte, M.T., Prochaska, S.: Adoption and use of precision farming technologies: A survey of central ohio precision farmers. The Ohio State University Department of Agricultural, Environmental and Development Economics, Report Series: AEDE-RP-0011-01 (2001)*

[4] *Chu, W., Zheng, X., Ding, D.: Image2weather: A large-scale image dataset for weather property estimation. In: 2016 IEEE Second International Conference on Multimedia Big Data (BigMM). pp. 137-144 (April 2016). https://doi.org/10.1109/BigMM.2016.9*

[5] *Das, J., Cross, G., Qu, C., Makineni, A., Tokekar, P., Mulgaonkar, Y., Kumar, V.: Devices, systems, and methods for automated monitoring enabling precision agriculture. In: 2015 IEEE International Conference on Automation Science and Engineering (CASE). pp. 462-469 (Aug 2015). https://doi.org/10.1109/CoASE.2015.7294123*

[6] *Das, J., Cross, G., Qu, C., Makineni, A., Tokekar, P., Mulgaonkar, Y., Kumar, V.: Devices, systems, and methods for automated monitoring enabling precision agriculture. In: 2015 IEEE International Conference on Automation Science and Engineering (CASE). pp. 462-469 (Aug 2015). https://doi.org/10.1109/CoASE.2015.7294123*

[7] *Deepika, G., Rajapirian, P.: Wireless sensor network in precision agriculture: A survey. In: 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS). pp. 1-4 (Feb 2016). https://doi.org/10.1109/ICETETS.2016.7603070*

[8] *Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C.: Food security: The challenge of feeding 9 billion people. Science 327(5967), 812-818 (2010). https://doi.org/10.1126/science.1185383, http://science.sciencemag.org/content/327/5967/812*

[9] *Goh, S.T., Tan, K.L.: Mosaic: A multi-feature access method for large image databases. In: Bench-Capon, T.J., Soda, G., Tjoa, A.M. (eds.) Database and Expert Systems Applications. pp. 862-871. Springer Berlin Heidelberg, Berlin, Heidelberg (1999)*

[10] *Hani, A.F.M., Paputungan, I.V., Hassan, M.F., Asirvadam, V.S., Daharus, M.: Development of private cloud storage for medical image research data. In: 2014 International Conference on Computer and Information Sciences (ICCOINS). pp. 1-6 (June 2014). https://doi.org/10.1109/ICCOINS.2014.6868433*

[11] *Hunt, E.R., Daughtry, C.S.T., Mirsky, S.B., Hively, W.D.: Remote sensing with simulated unmanned aircraft imagery for precision agriculture applications. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(11), 4566-4571 (Nov 2014). https://doi.org/10.1109/JSTARS.2014.2317876*

[12] *Li, C., Yang, W.: The distributed storage strategy research of remote sensing image based on mongo db. In: 2014 Third International Workshop on Earth Observation and Remote Sensing Applications (EORSA). pp. 101-104 (June 2014). https://doi.org/10.1109/EORSA.2014.6927858*

[13] *Li, R., Lu, J., Zhang, Y., Lu, Z., Xu, W.: A framework of large-scale and real-time image annotation system. In: 2009 International Joint Conference on Artificial Intelligence. pp. 576-579 (April 2009). https://doi.org/10.1109/JCAI.2009.33*

[14] *Liang, Z.: Fast image retrieval method based on visual word tree word. In: 2011 10th International Symposium on Distributed Computing and Applications to Business, Engineering and Science. pp. 211-215 (Oct 2011). https://doi.org/10.1109/DCABES.2011.44*

[15] *Mikulik, A., Chum, O., Matas, J.: Image retrieval for online browsing in large image collections. In: Brisaboa, N., Pedreira, O., Zezula, P. (eds.) Similarity Search and Applications. pp. 3-15. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)*

[16] *Mora, A., Vemprala, S., Carrio, A., Saripalli, S.: Flight performance assessment of land surveying trajectories for multiple uav platforms. In: 2015 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS). pp. 1-7 (Nov 2015). https://doi.org/10.1109/RED-UAS.2015.7440984*

[17] *Nguyen, G., Franco, P., Mullot, R., Ogier, J.: Mapping high dimensional features onto hilbert curve: Applying to fast image retrieval. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). pp. 425-428 (Nov 2012)*

[18] *Pedersen, S.M., Lind, K.M.: Precision Agriculture - From Mapping to Site Specific Application, pp. 1-20. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-68715-5_1, https://doi.org/10.1007/978-3-319- 68715-5 1*

[19] *Remias, E., Sheikholeslami, G., Zhang, A., Syeda-Mahmood, T.F.: Supporting content-based retrieval in large image database systems. Multimedia Tools and Applications 4(2), 153-170 (Mar 1997). https://doi.org/10.1023/A:1009614331352, https://doi.org/10.1023/A:1009614331352*

[20] *Ribera, J., Chen, Y., Boomsma, C., Delp, E.J.: Counting plants using deep learning. In: 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP). pp. 1344-1348 (Nov 2017). https://doi.org/10.1109/GlobalSIP.2017.8309180*

[21] *Robert, P.: Precision agriculture: a challenge for crop nutrition management. In: Progress in Plant Nutrition: Plenary Lectures of the XIV International Plant Nutrition Colloquium. pp. 143-149. Springer (2002)*

[22] *Sagan, H.: Space-filling curves. Springer Science & Business Media (2012)*

[23] *Shih-Ming Huang, Ching-Chun Huang, C.C.C.: Image registration among uav image sequence and google satellite image under quality mismatch. In: 2012 12th International Conference on ITS Telecommunications. pp. 311-315 (Nov 2012). https://doi.org/10.1109/ITST.2012.6425189*

[24] *Tri, N.C., Van Hoai, T., Duong, H.N., Trong, N.T., Van Vinh, V., Snasel, V.: A novel framework based on deep learning and unmanned aerial vehicles to assess the quality of rice fields. In: Advances in Information and Communication Technology. pp. 84-93. Springer International Publishing, Cham (2017)*

[25] *van der Wal, T., Abma, B., Viguria, A., Prévinaire, E., Zarco-Tejada, P.J., Serruys, P., van Valkengoed, E., van der Voet, P.: Fieldcopter: unmanned aerial systems for* crop monit*oring services. In: Stafford, J.V. (ed.) Precision agriculture '13. pp. 169-175. Wageningen Academic Publishers, Wageningen (2013)*

[26] *Wang, W., Wang, S.: A method using locality-sensitive hashing for large-scale content-based image retrieval. In: 2009 Chinese Control and Decision Conference. pp. 1816-1820 (June 2009). https://doi.org/10.1109/CCDC.2009.5192277*

[27] *Weihong, W., Song, W.: A scalable content-based image retrieval scheme using locality-sensitive hashing. In: 2009 International Conference on Computational Intelligence and Natural Computing. vol. 1, pp. 151-154 (June 2009). https://doi.org/10.1109/CINC.2009.124*

[28] *Zhuo, Y., Sun, G., Ke, Y., Shen, C., GuoxuLiu, J.: Comparison of two led spectrum engineering approaches for horticulture lighting. In: 2018 15th China International Forum on Solid State Lighting: International Forum on Wide Bandgap Semiconductors China (SSL-China: IFWS). pp. 1-4 (Oct 2018). https://doi.org/10.1109/IFWS.2018.8587320*

Biografía Autor(es)

a. MSc(c) Ingeniería eléctirca. Docente, Universidad Tecnológica de Pereira, Colombia. E-mail: osanmartinez@utp.edu.co. Pereira, Colombia.

b. Ingeniero electricista. Docente, Universidad Tecnológica de Pereira, Colombia. E-mail: alifranco@utp.edu.co. Pereira, Colombia.

c. PhD(c) en Ingeniería. Docente, Universidad Tecnológica de Pereira, Colombia. E-mail: gahol@utp.edu.co. Pereira, Colombia.

Análisis patrones de competencias egresados Programa Ingeniería de Sistemas

Javier Alejandro Jiménez Toledo (a)
Carlos Fernando González Guzmán (b)
José María Muñoz Botina (c)

Resumen

Actualmente, la información juega un papel importante en el desarrollo de la sociedad, las empresas cuentan con información almacenada en diferentes bancos de datos y/o repositorios, y en algunas ocasiones, esta información se encuentra estática sin posibilidad de poder ser analizada para la toma de decisiones que afecten a la organización.

Esta investigación, contó con repositorios internos y externos con que cuenta la IUCESMAG, dentro de las fuentes internas se manejaron repositorios de la Oficina de Proyección Institucional (OPI) y egresados del Programa de Ingeniería de Sistemas. En el caso de las fuentes externas COPNIA, y el Observatorio Laboral de Empleabilidad – OLE. Cruzando información de estos repositorios se creó un modelo de datos, tipo estrella en MaríaDB, transformado la información en un Data mart y utilizando el análisis multidimensional relacional – ROLAP, transformar dinámicamente las consultas de los usuarios en consulta SQL.

Utilizando la metodología CRISP-DM, se aborda el estudio de egresados de la IUCESMAG, caso Ingeniería de Sistemas que consta de 6 fases o etapas donde cada una de ellas presentan concurrencia entre sí, la tarea utilizada fue la predictiva donde se basó en relaciones de la variable dependiente con variables independientes, utilizando como método de clasificación la técnica de árboles de decisión. Los resultados logrados en este proceso servirán para la toma de decisiones por parte del programa en el fortalecimiento de competencias, y poder así identificar el perfil profesional y ocupacional de los futuros ingenieros de sistemas en el contexto local como regional.

Palabras clave: Data Mining, egresados, repositorios, arboles de decisión, RapidMiner.

Analysis patterns of graduate competences Program Engineering of Systems

Abstract

Currently, information plays an important role in the development of society, companies have information stored in different data banks and / or repositories, and in some cases, this information is static without the possibility of being analyzed for data collection. of decisions that affect the organization.

This research included internal and external repositories of the IUCESMAG, internal repositories of the Office of Institutional Projection (OPI) and graduates of the Systems Engineering Program were managed within the internal sources. In the case of external sources COPNIA, and the Employment Observatory of Employability - OLE. Crossing information from these repositories a data model was created, star type in MaríaDB, transformed the information in a Data mart and using the multidimensional relational analysis - ROLAP, dynamically transforming the queries of the users in SQL query.

Using the CRISP-DM methodology, the study of graduates of the IUCESMAG is addressed, in the case of Systems Engineering, which consists of 6 phases or stages where each of them has concurrence with each other, the task used was the predictive one, where it was based on relations of the dependent variable with independent variables, using the technique of decision trees as a classification method. The results obtained in this process will be used for the decision making of the program in the strengthening of competences, and thus be able to identify the professional and occupational profile of future systems engineers in the local and regional context.

Key words: Data Mining, graduates, repositories, decision trees, RapidMiner.

I. INTRODUCCIÓN

Actualmente, vivimos en un mundo donde la información asume un papel importante en el desarrollo de la sociedad, las diferentes organizaciones cuentan con información almacenada en diferentes bancos de datos y/o repositorios, y en algunas ocasiones, esta información se encuentra estática sin posibilidad de poder ser analizada para la toma de decisiones que afecten a la organización. El uso de diferentes tecnologías como es el internet y los diferentes sistemas de información que poseen las organizaciones ha hecho que la búsqueda de patrones de datos sea una acción continua para lograr la calidad y éxito en los diferentes procesos de las organizaciones.

Las condiciones de calidad garantizan que un programa académico corresponde a su naturaleza en un contexto institucional que le favorece y puede ser ofrecido a la sociedad. El Ministerio de Educación Nacional – MEN exige que "la información de sus graduados contemple variables ajustadas a un proceso de calidad tales como: salarios, ubicación laboral, tipo de contratación, entre otros datos y por otro lado su impacto social, cultural, económico y académico". [1]

Por consiguiente, determina que las Instituciones de Educación Superior (IES) necesitan realizar un continuo seguimiento del desempeño de sus graduados con el propósito de determinar los niveles de pertinencia de la formación ofrecida, sector económico, emprendimiento, prácticas profesionales, entre otras, que corresponden con el logro de los fines definidos autónoma e institucionalmente en cada proyecto educativo. [2]

En un estudio se ha señalado que uno de los factores más preocupantes en las IES en Colombia, se refiere al seguimiento de la situación de sus graduados, desde el punto de vista de su desarrollo personal, laboral, social y su impacto en el medio [3]. Por otra parte, la Institución Universitaria CESMAG se demoró en fijar políticas y operacionalizarlas a través de sistemas de información eficientes alrededor de los graduados, trayendo como consecuencia la dificultad para comunicarse con ellos, ya sea por el cambio de trabajo, de ciudad, de dirección o de celular, o por fallecimiento, entre otros datos. En tal sentido el programa de Ingeniería de Sistemas, para efectos de acreditación debe periódicamente obtener información de los egresados del programa para efectos de conocer sobre diferentes aspectos tales como: información básica, formación, creación de empresa, empleabilidad, apoyo institucional y del programa, competencias y habilidades, confrontándolas con la información de empleadores en el sector local.

El estudio se realizó a partir diferentes repositorios tanto internos como externos o fuentes de información quevposee la IUCESMAG, dentro de las fuentes internas se manejaron dos repositorios de datos en formato xls, como fueron el de la Oficina de Proyección Institucional (OPI) y el repositorio de egresados de Programa de Ingeniería de Sistemas. En el caso de las fuentes externas se contó con el repositorio de COPNIA, con información de matrícula profesional e información básica del egresado y el repositorio del Observatorio Laboral de Empleabilidad – OLE. Cruzando información de los repositorios se creó un modelo de datos, basado en un modelo relacional tipo estrella en MaríaDB, donde se incluye una tabla de hecho y varias tablas de dimensión, transformado la información en un Data mart utilizando el análisis multidimensional relacional – ROLAP, transformando dinámicamente las consultas de los usuarios en consulta SQL [4], estas consultas se relacionan mediante tablas cruzadas con el objetivo de entregar información a los diferentes usuarios que deseen acceso a ella.

Basándose en el Descubrimiento del Conocimiento en Base de Datos y utilizando Data Mining, se descubre patrones ocultos, ya en el proceso de Minería de Datos, se utilizó la metodología CRISP-DM [5], que es un modelo que describe la manera como a nivel técnico se aborda el estudio de egresados de la IUCESMAG, caso Ingeniería de Sistemas que consta de 6 fases o etapas donde cada una de ellas presentan concurrencia entre sí, la tarea de minería de datos utilizada fue la predictiva donde se basó en relaciones de la variable dependiente con variables independientes, utilizando como método de clasificación la técnica de árboles de decisión. Los resultados logrados en este proceso servirán para la toma de decisiones por parte del programa en el fortalecimiento de competencias, entendiéndose estas como Todo lo que un estudiante debe saber relacionado con una disciplina particular y su aplicación significativa para recrear (o crear) conocimiento a partir de lo adquirido. Los referentes para seleccionar este saber son de tres tipos: Universal de la disciplina, el de la cultura regional y el referente de los intereses y motivaciones individuales [7] y habilidades, así como identificar el perfil profesional y ocupacional de los futuros ingenieros de sistemas en el contexto local como regional.

II. METODOLOGÍA

La metodología utilizada en el proceso de minería de datos fue la CRISP-DM (Cross Industry Standard Process for Data Mining), el cual es un modelo, mediante el cual se describe la manera como a nivel técnico se comprende y aborda el tema objeto de estudio, igualmente entre las tareas de minería de datos más importantes están: la clasificación, la segmentación o Clustering, la asociación y finalmente los patrones secuenciales [8], que en este caso es el de egresados en la IUCESMAG, caso Ingeniería de Sistemas. consta de 6 fases o etapas que no solamente se pueden comportar de forma lineal sino concurrente [9]. A continuación, se describe las actividades que se tuvieron en cuenta durante el proceso investigativo de egresados.

2.1 Compresión del negocio.

En ella se encuentran las tareas más significativas que se deben desarrollar y analizar, convertidas en objetivos y requisitos necesarios para aplicar el concepto de minería de datos [10].

La problemática de la investigación se desarrolla en la IUCESMAG, tomando como caso de uso el Programa de Ingeniería de Sistemas, en él se realizará un seguimiento de la información de egresados desde el año 2007 hasta el primer periodo del 2016, antes y durante la investigación se pudo comprobar que la información que se manejaba en las dependencias encargadas presentaba algunos inconvenientes tales como: información incompleta es decir existían registros faltos de datos con respecto a una característica del egresado, de la misma manera esa información no era la más actualizada en cuanto a: datos básicos del egresado, formación académica, empleabilidad, de tal manera que los datos no se ajustaba a una realidad plena de los egresados, ya sea porque la obtención de la información se realizó hace mucho tiempo o las fuentes de datos no eran las más confiables.

Por otra parte, esta información se encontraba almacenada en diferentes archivos y en algunos casos en diversas fuentes de obtención de datos, trayendo como consecuencia un manejo separado de la información en la relación egresado – empleador- institución, lo anterior conlleva que se haga necesario realizar un estudio pertinente de minería de datos.

Para poder llegar a determinar los objetivos del negocio, se formuló el problema objeto de estudio de la investigación el cual es: ¿Cómo esTable.cer el grado de aceptación a nivel social de los graduados del programa de Ingeniería de Sistemas de la Institución Universitaria CESMAG en los aspectos personal, laboral y profesional ya sea en el contexto local y regional? Es así como los objetivos de la investigación a nivel técnico se resumen en los siguientes:

- Generar un solo repositorio de información a partir de diferentes fuentes tanto internas como externas logrando que este repositorio se encuentre actualizado en un gran porcentaje.

- Presentar información oportuna al momento en que esta sea consultada por las diferentes dependencias y/o entidades que la requieran.

- Analizar y visualizar los resultados de la investigación de forma gráfica o estadística y poder así ser mejor interpretados por los usuarios finales que en este caso pueden ser personal administrativo, directivo y docente.

- Proponer una metodología para los programas que sea fácil de manejar, adecuada y ajustada a la realidad en el manejo de información de egresados presente en la IUCESMAG que permita operar con procesos de depuración, actualización, minería y visualización de datos.

Con respecto a la evaluación de la situación problémica del estudio de egresados, en principio hay que responder los siguientes interrogantes que se presentan en la IUCESMAG: ¿Cuál es el conocimiento previo disponible en cuanto a egresados?, actualmente los programas de la IUCESMAG, cuentan con información de egresados que provienen de diferentes fuentes tanto internas es decir información propia y externas como aquellas que el Ministerio de Educación provee, de la misma forma aunque se cuenta con esta información existen pocos estudios al respecto que plasmen resultados, metodologías y procesos en la obtención de información. ¿se cuenta con cantidad de datos requerida para resolver el problema?, frente a esta pregunta, aunque los datos pueden están presentes, no se cuenta con una metodología apropiada para que esa cantidad de información sea la necesaria y precisa al momento de realizar el estudio. De igual manera se debe clasificar y organizar la información con el propósito de lograr resultados más coherentes y estructurados acordes a la realidad. ¿se cuenta con aplicaciones de minería de datos para el manejo de la información de egresados?, frente a esta pregunta se puede afirmar que los datos de egresados se manejan de forma digital pero en formato de hoja de cálculo y en algunos caso archivos planos, esta información se procesa de forma sistematizada mediante fórmulas estadísticas o filtros que no garantizan una transparencia de los datos, la minería de datos es una técnica que afianza el comportamiento, validación, depuración y entrega de resultados relacionables de n variables en n dimensiones llegando a alcanzar un conocimiento más profundo y evidenciado.

El plan de proyecto a realizar durante la investigación en el estudio de egresados del Programa de Ingeniería de sistemas tuvo como etapas fundamentales las siguientes:

- Análisis de los diferentes repositorios almacenados en diferentes fuentes tanto internas como externas.

- Unificación de un solo repositorio tanto en fuentes internas como externas.
- Selección de campos asociados a las variables objeto de estudio.
- Generación de un único repositorio de información de egresados. Para las anteriores etapas se tuvo en cuenta como herramienta de apoyo la hoja de cálculo, formulas y filtros.
- Segmentación de la información del repositorio en archivos que contienen la información referente a cada una de las variables objeto de estudio.
- Exportación de los archivos de Excel en formato CSV para ser trabajado en bases de datos.
- Creación de la base de datos en WorkBench 6.3 CE utilizando script sql
- Depuración de registros en la base de datos.
- Diseño E-R de la base de datos y normalización
- Exportación base de datos formato sql.
- Conexión de la base de datos y almacenamiento en forma de repositorios por cada una de las tablas de egresados utilizando RapidMiner.
- Limpieza, transformación, reemplazo, búsqueda, ordenamiento de registros de las tablas de la base de datos que fueron aplicadas al modelamiento de minería de datos, mediante la técnica de árboles de decisión.
- Exportación del archivo de Excel de empleadores en formato CSV para ser trabajado en bases de datos.
- Creación de la base de datos en Workbench 6.3 CE utilizando script sql
- Depuración de registros en la base de datos.
- Exportación base de datos formato sql.
- Resultados parametrizados con cruce de las variables entre los datos de egresados y los de empleadores mediante el análisis multidimensional.

2.2 Comprensión de datos

Esta etapa se relaciona con la obtención de datos primarios y fuentes de acceso de la información, que conducen a tener una visión y realidad del comportamiento en cuanto a la información de egresados se refiere.

Inicialmente, se empezó determinado las fuentes de información que la IUCESMAG y específicamente el Programa de Ingeniería de Sistemas cuenta, para ello se diferenció en dos (2) fuentes que fueron INTERNAS Y EXTERNAS. En el caso de la primera fuente, se manejaron dos repositorios de datos guardados en formato XLS, que fueron: Oficina

de Proyección Institucional – OPI que contiene información básica de egresados y está conformada por 472 registros y 16 campos. Igualmente se tiene el repositorio de datos encuesta a egresados realizada por el Programa de Ingeniería conformado por 172 registros y 170 campos contiene información básica del egresado, formación, empleabilidad, grado de satisfacción hacia la institución y el programa. Para el caso de la segunda fuente, se manejó el repositorio del Consejo Profesional Nacional de Ingeniería – COPNIA, con información de matrícula profesional e información básica del egresado. Cuenta con 13 campos y 56 registros y como segundo repositorio está el Observatorio Laboral de Empleabilidad – OLE como un sistema de información, que se toma toda la información que brinda herramientas valiosas para analizar la pertinencia de la educación a partir del seguimiento a los graduados y su empleabilidad en el mercado laboral, contribuyendo de esta forma al mejoramiento de la calidad de los programas académicos ofrecidos [6], el cual para efectos de la investigación consta de 289 registros y 170 campos.

Se cruzó información básica del egresado en el repositorio interno con el objeto de poder llegar a contar con la información completa y necesaria, en algunos casos se hizo necesario poder agregar nuevos registros de información de egresados que no existen en el repositorio base que en este caso es del OPI la información común entre ambos repositorios es la información básica.

El repositorio base es OLE y se compara con el repositorio de COPNIA colocando nueva información o suprimiendo alguna de ellas de la cual no se tiene mayor cantidad de datos.

Estructura	*Identificación*	*Variable*	*Tipo*
Datos Básicos	*Cédula*	*Cedula*	*Integer*
	Nombre	*Nombre*	*Varchar*
	Apellido	*Apellido*	*Varchar*
	Fecha de Grado	*Fecha_grado*	*Date*
	Género	*Genero*	*Varchar*
	Residencia	*Residencia*	*Varchar*
	Estado Civil	*Estado_civil*	*Varchar*
Formación	*Cédula*	*Ced*	*Integer*
	Formación	*Formación*	*Varchar*
	Tarjeta Profesional	*Tarjeta_prof*	*Varchar*
	Certificación Idioma	*Cert_idioma*	*Varchar*
Crear Empresa	*Cédula*	*Ced*	*Integer*
	Crear Empresa	*Crear_emp*	*Varchar*
	Primer Trabajo	*Primer_trab*	*Varchar*
	Relación Empresa Carrera	*Emp_carrera*	*Varchar*
Empleo	*Cédula*	*Cedula*	*Integer*
	Ocupación	*Ocupacion*	*Varchar*
	Personal a Cargo	*Personal_cargo*	*Varchar*
	Sector de la Empresa	*Sector_empresa*	*Varchar*
	Relación Institución Empresa	*Institución_empresa*	*Varchar*

Apoyo Institucional	*Cédula*	*Ced*	*Integer*
	Formación	*Formación*	*Varchar*
	Orientación	*Orientación*	*Varchar*
	Emprendimiento	*Emprendimiento*	*Varchar*
	Educación Virtual	*Ed_virtual*	*Varchar*
	Educación Posgradual	*Est_posgrado*	*Varchar*
	Cursos	*Cursos Varchar*	*Varchar*
	Encuentro de Egresados	*Encuentro_eg*	*Varchar*
	Vinculación Laboral	*Vinculación_lab*	*Varchar*
Programa	*Cédula*	*Cedula*	*Integer*
	Idiomas	*Idiomas*	*Varchar*
	Institución	*Institución*	*Varchar*
	Programa	*Programa*	*Varchar*
	Empleabilidad	*Empleabilidad*	*Varchar*
	Perfil Profesional	*Perfil_profesional*	*Varchar*
	Plan de Estudios	*Plan_estudios*	*Varchar*
	Actualización del Plan de Estudios	*Actualizacion_plan_estudios*	*Varchar*
	Práctica Profesional	*Practica_profesional*	*Varchar*
Competencias	*Cédula*	*Ced*	*Integer*
	Comunicación	*Comunicación*	*Varchar*
	TIC	*TIC*	*Varchar*
	Resolución	*Resolución*	*Varchar*
	Análisis	*Análisis*	*Varchar*
	Convivencia	*Convivencia*	*Varchar*
	Decisión	*Decisión*	*Varchar*
Habilidades	*Cédula*	*Cedula*	*Integer*
	Desarrollo Web	*Web*	*Varchar*
	Emprendimiento	*Emprendimiento*	*Varchar*
	Lecto Escritura	*Lectoescritura*	*Varchar*
	Decisiones	*Decisiones*	*Varchar*
	Liderazgo	*Liderazgo*	*Varchar*
	Solución de Problemas	*Solución_problemas*	*Varchar*
	Planificación	*Planificación*	*Varchar*
	TIC	*TIC*	*Varchar*
	Base Tecnológica	*Base_tecnológica*	*Varchar*
	Base de Datos	*Base_datos*	*Varchar*
	Ingeniería de Software	*Ingeniería_software*	*Varchar*
	Gestión de Proyectos	*Gestión_proyectos*	*Varchar*
	Aplicaciones Telemáticas	*Aplicaciones_telematica*	*Varchar*
	Aplicaciones Nube	*Aplicaciones_nube*	*Varchar*
	Minería de Datos	*Data_mining*	*Varchar*
	Auditoría de Sistemas	*Auditoria_sistemas*	*Varchar*
	Programación Frame Works	*Programación_frameworks*	*Varchar*

Fig. 1. *Tabla de variables egresados*
Fuente. Esta investigación.

Cabe mencionar que durante este proceso se pudo apreciar una serie de inconvenientes generados en su mayoría por el uso no constante de registros en los diferentes repositorios tanto de fuentes internas como externas lo que conlleva un proceso minucioso de comparación de datos que, aunque se emplean como técnicas la utilización y manejo de fórmulas y filtrados, los tiempos de entrega de resultados de los repositorios son dispendiosos.

Posteriormente se realizó el cruce de información entre las dos fuentes obteniendo un solo repositorio final el cual contiene 290 registros y 92 campos.

Como la cantidad de campos es muy extensa y existe información (campos) que no se ajustan a la realidad de la investigación se hace una depuración, dejando únicamente lo campos que son necesarios para el estudio los cuales se encuentran divididos en 8 estructuras del repositorio final de egresados como lo muestra la figura nro.1.

Se importan los datos de cada una de las estructuras en formato CSV, delimitado por comas, al administrador grafico de base de datos WorkBench, donde mediante instrucciones SQL se crea la base de datos de egresados, con un diseño relacional tipo estrella donde se construye a partir del modelo entidad relación, en él se distingue entre tablas de hechos (facts), que contienen los datos para el análisis rodeadas de las tablas de dimensiones que permiten agrupar los hechos en función de los valores de la dimensión, para el caso particular con 1 entidad de hecho y 7 entidades derivadas como se plasma en la figura nro.3.

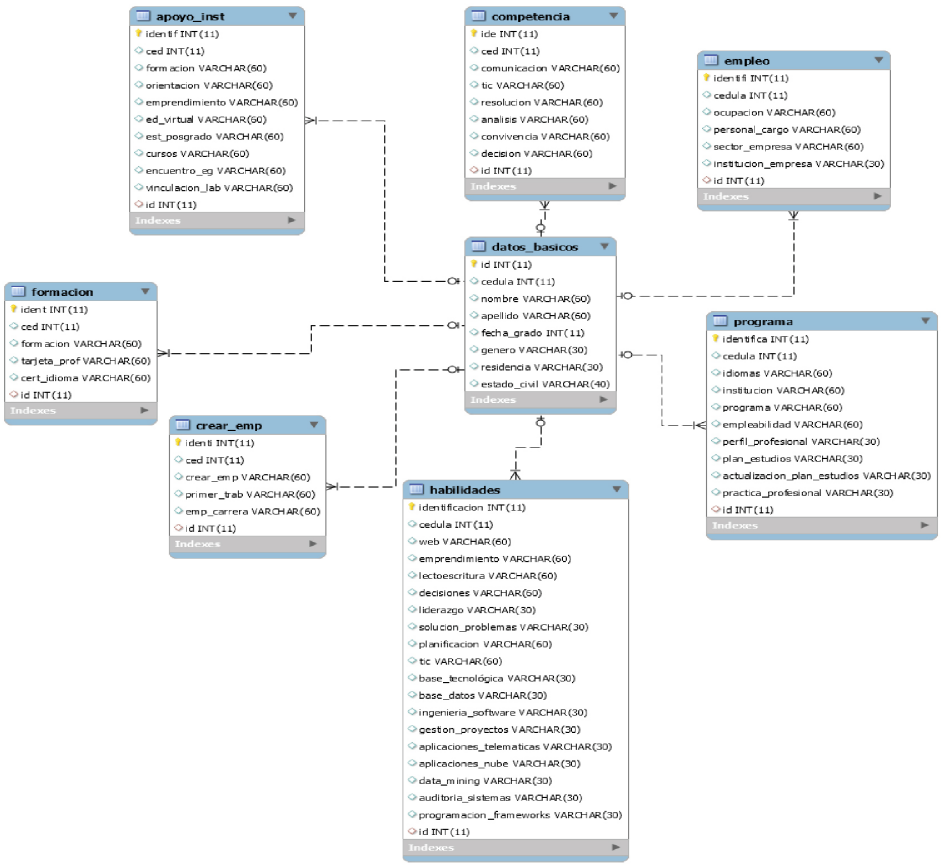


Fig. 2. *Diseño relacional*
Fuente. Investigación propia.

Se utilizaron scripts para la creación de la base de datos en el sistema gestor de bases de datos MariaDB, A partir de instrucciones SQL, se depura la base de datos buscando datos nulos (NULL), con el propósito de insertar (INSERT) los registros con información pertinente, igualmente se actualizo la información de algunos registros a partir de la instrucción (UPDATE), y se eliminó información duplicada en el caso de la identificación: cédula, nombres, apellidos del egresado mediante la instrucción (DELETE). Como punto importante para la utilización de WorkBench, se necesitó la instalación de WampServer, versión 2,2 el cual corre como servidor en la maquina local a través de la dirección URL, http://localhost y poder trabajar igualmente la base de datos desde PhpMyAdmin, versión 3.51.

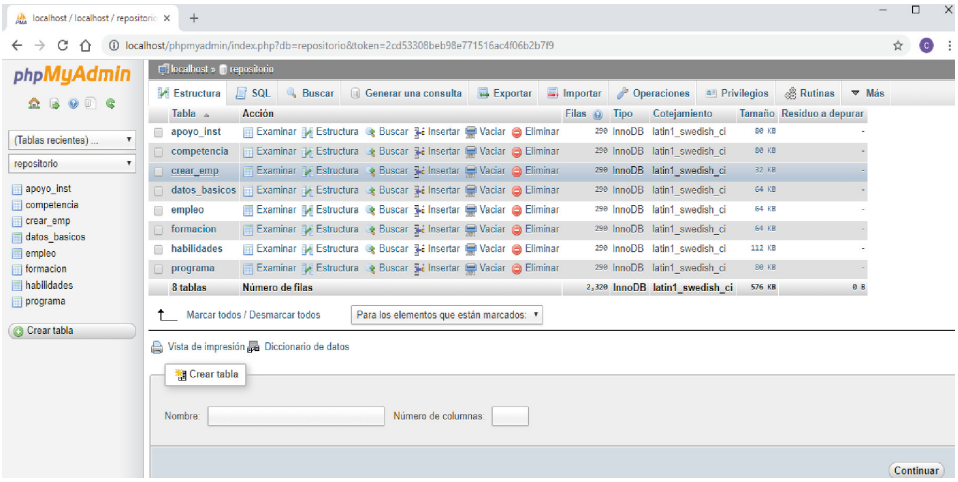


Fig. 3. *Aplicativo de base de datos workbench*
Fuente. Investigación propia.

Así mismo se procede al diseño relacional de la base de datos que tiene como tabla centralizada los datos básicos del egresado, esta tabla mantiene una relación de uno a varios con las 7 tablas restantes formando un esquema dimensional en estrella, como se aprecia en la figura nro. 3. A partir de instrucciones SQL, se depura la base de datos buscando datos nulos (NULL), con el propósito de insertar (INSERT) los registros con información pertinente, igualmente se actualizo la información de algunos registros a partir de la instrucción (UPDATE), y se eliminó información duplicada en el caso de la identificación: cédula, nombres, apellidos del egresado mediante la instrucción (DELETE).

Como punto importante para la utilización de WorkBench, se necesitó la instalación de WampServer, versión 2,2 el cual corre como servidor en la maquina local a través de la dirección URL, http://localhost y poder trabajar igualmente la base de datos desde PhpMyAdmin, versión 3.51. Figura nro. 4.

2.3 Preparación de los datos

En esta fase se procede a la preparación de los datos para poder llegar a aplicarlos a la técnica de minería de datos, es decir se realiza una limpieza de los mismos utilizando para ello la herramienta de minería llamada RapidMiner, ver figura 4., se selecciona la opción de depuración previa conexión con un origen de datos (DataSet)

que en este caso puede ser una tabla o la relación entre dos o más tablas del repositorio de base de datos de egresados, una vez realizado este proceso se procede a transformar (reemplazar, ordenar, renombrar, cambiar el tipo de datos, copiar, filtrar) y se activa la transformación, este proceso permite tener una base de datos coherente y estructurada con información completa en todos sus campos. De igual manera el proceso de limpieza permite remover aquellos datos con baja calidad en su estructura e igualmente datos correlacionales de las variables relacionadas entre dos o más tablas que no son significativos para la obtención de resultados.

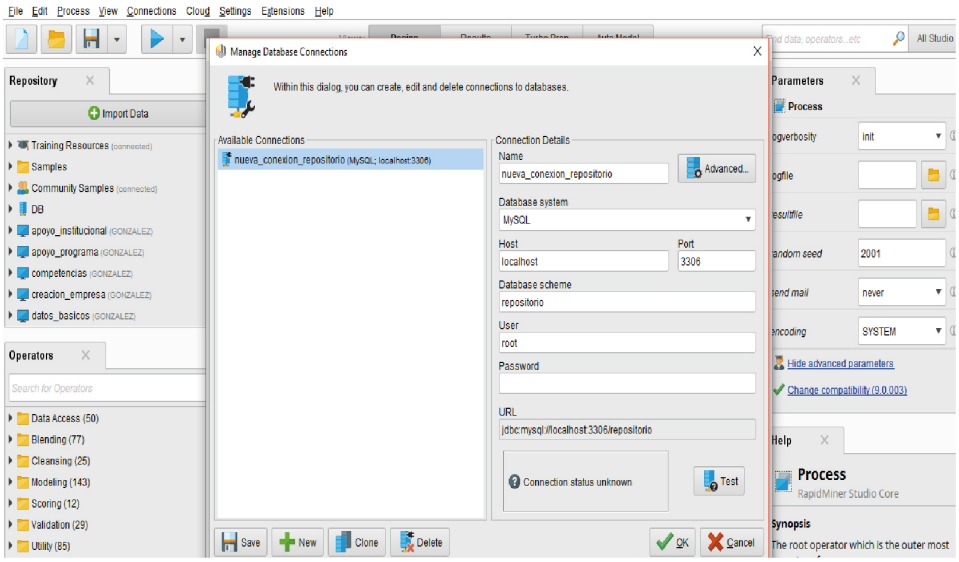


Fig. 4. *Interfaz Rapid Miner.*
Fuente. Investigación propia.

Posteriormente se procede a crear un nuevo repositorio con el nombre base_datos_egresados se abre la conexión nueva_conexion_repositorio, se crea un DataSet denominado datos_basicos_egresados, y se procede a realizar la limpieza de los datos como se visualiza en la figura nro.5

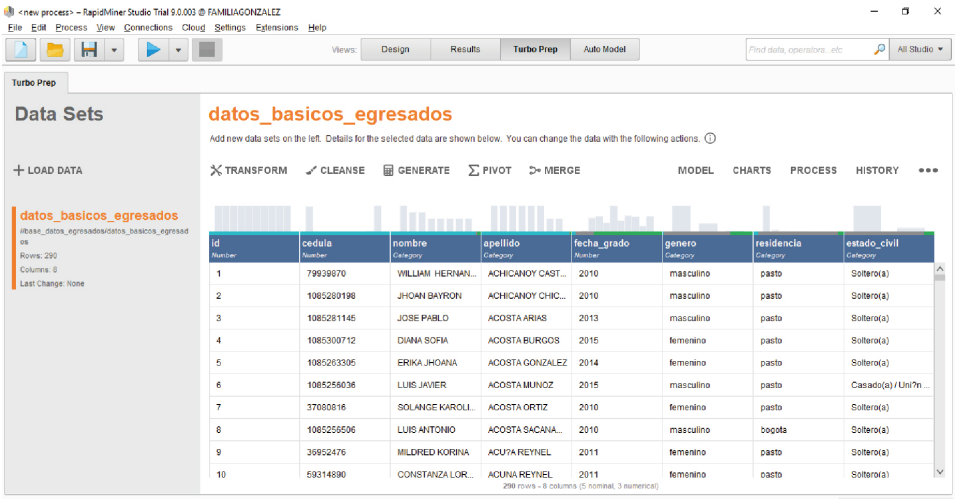


Fig. 5. *Depuración base de datos.*
Fuente. Investigación propia.

Se procede a la transformación de los mismos seleccionando cada uno de los campos que se desean realizar las acciones tanto de renombrar, cambio de tipo de datos, remoción de datos, copiar, filtrar, generación de rangos y ordenación. Posteriormente, se realiza la limpieza de datos nuevamente que se encuentren duplicados y se exporta el DataSets donde se crea en RapidMiner un nuevo proceso el cual utilizando como técnica de clasificación 5 árboles de decisión por cada cada proceso con 5 jerarquías se genera el árbol gráfico y el algoritmo correspondiente como se aprecia en la figura nro. 6, cabe mencionar que las ventajas de un árbol de decisión, entre muchas están la interpretación de la decisión adoptada, la comprensión del conocimiento en la toma de decisiones, el comportamiento respecto a una determinada decisión y la reducción en el número de variables independientes.

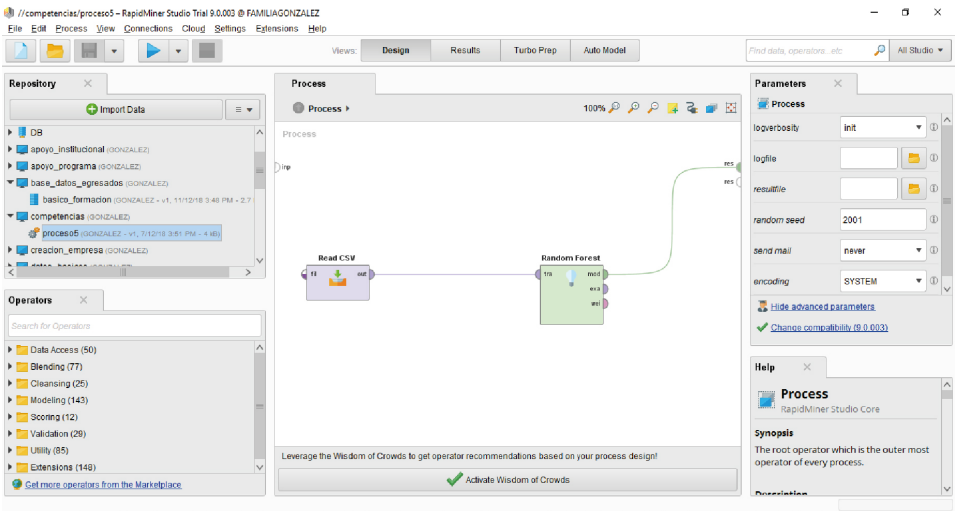


Fig. 6. *GUI. Proceso.*
Fuente. Investigación propia.

2.4 Modelado

En esta fase se determina la técnica de modelado como es la descriptiva utilizando los árboles de decisión, Decision tree, basado en el algoritmo Quinlan's C4.5 (CART), este algoritmo genera árboles de decisión a partir de los datos mediante particiones realizadas recursivamente. Se tienen en cuenta todas las pruebas posibles que pueden dividir el conjunto de datos y se selecciona la prueba con la que se haya obtenido mayor ganancia de la información, para cada atributo discreto, se considera una prueba con k resultados, siendo k el número de valores posibles que puede tomar el atributo[11]. Teniendo en cuenta que dicha técnica debe ser la mas adecuada frente al seguimiento de egresados, se contó con los datos suficientes para realizar dicho análisis que en este caso fueron 290 registro con 92 campos, se cumple igualmente con todos los requisitos del problema que se está en concordancia con las las estructuras de variables plateadas como son datos básicos, formación, empleo, crear empresa, habilidades, competencias, apoyo de la institución y apoyo del programa. Por último, la técnica de modelado utilizada maneja como regla un grado de exactitud superior al 76% al momento de seleccionar los cruces de variables donde por cada una de las estructuras debe existir una etiqueta de comparación la cual sirve como patrón de consulta de datos y verificación frente a las variables que en su mayoría son binomiales.

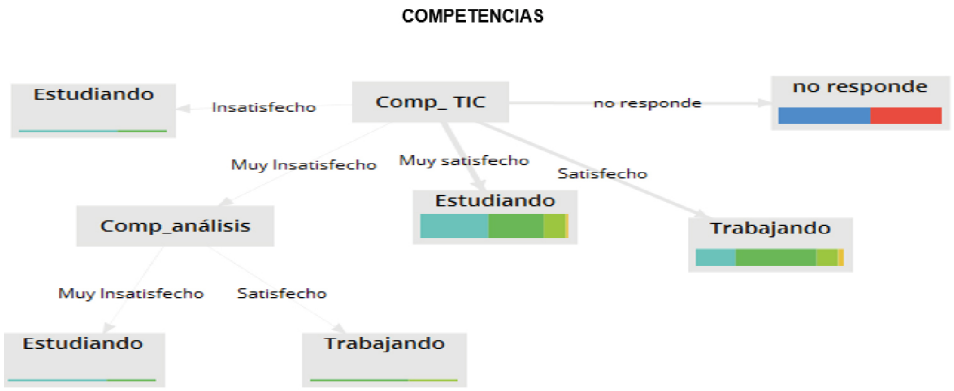


Fig. 7. *Árbol de decisión.*
Fuente. Investigación propia.

```
Comp_ TIC = Insatisfecho: Estudiando {no responde=0, Estudiando=2, Trabajando=1, Buscando trabajo=0, Otra actividad=0, trabajando=0}
Comp_ TIC = Muy Insatisfecho
|   Comp_análisis = Muy Insatisfecho: Estudiando {no responde=0, Estudiando=2, Trabajando=1, Buscando trabajo=0, Otra actividad=0, trabajando=0}
|   Comp_análisis = Satisfecho: Trabajando {no responde=0, Estudiando=0, Trabajando=2, Buscando trabajo=1, Otra actividad=0, trabajando=0}
Comp_ TIC = Muy satisfecho: Estudiando {no responde=0, Estudiando=55, Trabajando=45, Buscando trabajo=18, Otra actividad=2, trabajando=0}
Comp_ TIC = Satisfecho: Trabajando {no responde=0, Estudiando=22, Trabajando=45, Buscando trabajo=12, Otra actividad=3, trabajando=0}
Comp_ TIC = no responde: no responde {no responde=45, Estudiando=0, Trabajando=0, Buscando trabajo=0, Otra actividad=0, trabajando=35}
```

Fig. 8. *Algoritmo*
Fuente. Investigación propia.

Se debe aclarar que para el caso de los empleadores como es un repositorio muy pequeño en cantidad de registros, menos de 24, no se puede realizar los procesos de minería de datos, pero si se puede generar comparaciones con los resultados finales del repositorio de egresados.

2.5 Evaluación

En esta fase se evalúa el modelo, el cual se tiene en cuenta el cumplimiento de los criterios para el éxito del análisis de la información de egresados.

III. RESULTADOS Y DISCUSIÓN

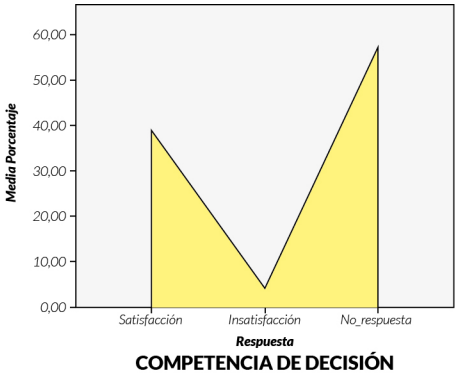
Para efectos de este artículo, solo se tomó la visión que tienen los egresados sobre el Programa de Ingeniería de Sistemas, en cuanto al fortalecimiento de las competencias que debe tener todo profesional en ingeniería de sistemas al momento de empezar su vida laborar a nivel empresarial o como gestor de empresa.

Table 1. *Análisis de las competencias de los egresados*

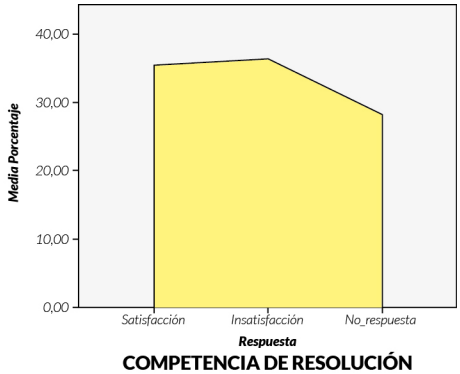
Competencia	Satisfechos	Insatisfechos	No Responden
Tecnología Información y Comunicación	*69.24%*	*69.24%*	*69.24%*
Análisis	*66.67%*	*66.67%*	*66.67%*
Toma de Decisiones	*38.83%*	*38.83%*	*38.83%*
Resolución Problemas	*35.42%*	*35.42%*	*35.42%*
Convivencia	*68.73%*	*68.73%*	*68.73%*

COMPETENCIA TIC

COMPETENCIA ANÁLISIS



COMPETENCIA DE DECISIÓN



COMPETENCIA DE RESOLUCIÓN

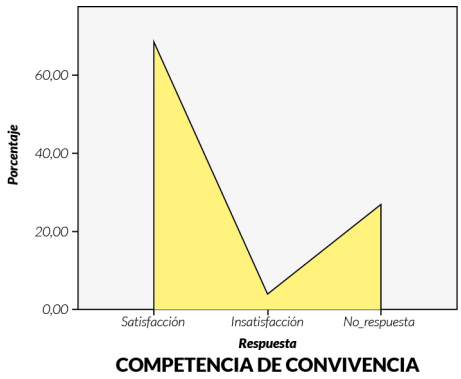


Fig. 9. *Interpretación gráfica competencia estudios de egresados. Fuente. Investigación propia.*

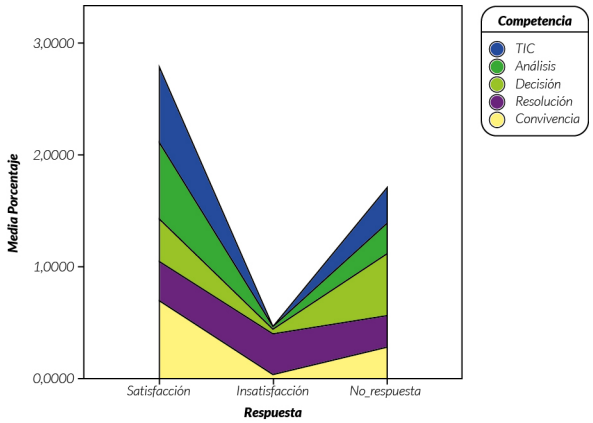


Fig. 10 *Consolidado respuestas competencias Programa de Ingeniería de Sistemas. Fuente. Investigación propia.*

V. CONCLUSIONES

Más del 50% (media), de los egresados al preguntarles sobre la importancia del fortalecimiento de sus competencias en la academia relacionadas a TIC, ANÁLISIS Y CONVIVENCIA al momento de pasar a una instancia profesional y laboral afirman que presentan un grado de satisfacción alto y muy alto, es decir si se da desde la academia un fortalecimiento en el uso de sistemas de base, aplicaciones y herramientas básicas para el desarrollo de la ingeniería de sistemas, así mismo en el uso de lenguajes de programación, paquetes estadísticos y software de diseño, el índice de egresados que no respondieron las 3 preguntas de TIC, ANÁLISIS Y CONVIVENCIA, se percibe un mayor porcentaje en la competencia de TIC, y el de menos porcentaje en la competencia de CONVIVENCIA, y es así como en esta competencia se garantiza una validez en sus resultados.

Los porcentajes generados de las 3 competencias anteriores, esTable.cen que la mayoría de sus egresados si están contentos y fortalecidos con el desarrollo de sus competencias las cuales son de un valor significativo a nivel profesional, con respecto al grado de insatisfacción el porcentaje más alto corresponde al de la competencia de CONVIVENCIA, con un valor cercano al 4% lo que se hace necesario fortalecerla durante la academia y lograr así que sea de un valor alto agregado de los futuros profesionales del programa de ingeniería de sistemas de la IUCESMAG.

En cuanto a la TOMA DE DECISIONES Y RESOLUCIÓN DE PROBLEMAS los valores de satisfacción se encuentran por debajo de la media y el grado de insatisfacción más alto lo tiene la competencia de RESOLUCIÓN DE PROBLEMAS CON UN 36.4%, Se recomienda fortalecer dicha competencia en los estudiantes del programa de ingeniería de sistemas, que les permita conocer la problemática de su entorno y formular soluciones ingenieriles precisas y prácticas frente a las necesidades de las organizaciones y los clientes.

REFERENCIAS

[1] *MINISTERIO DE EDUCACIÓN NACIONAL. Sistema de información Observatorio Laboral [en linea]. <http://www.egresadoscolombia.edu.co/html/1732/w3-property-value-36267.html>[citado en 13 de diciembre de 2016].*

[2] *ASCUN, A. C. (2006). Seguimiento de egresados de instituciones de egresados de Educación Superior.*

[3] *NORTE, U. C. (2007). Revista Virtual No. 22.*

[4] *Perez, Carlos. Estudio de patrones de interacción entre los estudiantes y la plataforma de teleenseñanza en la UPM. Escuela Técnica de Superior de Ingeniería y Sistemas de Telecomunicación.*

[5] *Galán Victor. Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universtiario. Universidad Carlos III de Madrid. Escuela Politecnica Superior Ingeniería en Informática. 2015.*

[6] *MINISTERIO DE EDUCACIÓN NACIONAL (2007). Graduados Colombia Observatorio Laboral para la Educación. [en línea]. <http://www.mineducacion.gov.co/1621/propertyvalue-35696.html>.[citado en 15 de febrero de 2017].*

[7] *Campos Mesa, M.d.C., Ries, F. y Castillo Andrés, Ó.d. (2011). Análisis de las competencias adquiridas y utilizadas por los egresados maestros en Educación Física. RICYDE. Revista Internacional de Ciencias del deporte, VII (24), 216-229.*

[8] *Agrawal, R., Srikant, R. (1994). Fast Algorithms for Mining Association Rules. En: 20th International Conference on Very Large Data Bases, VLDB 1994, Santiago de Chile. pp. 487 a 499.*

[9] *Oldemar Rodríguez Rojas. Metodología para el Desarrollo de Proyectos en Minería de Datos CRISPDM, 2010.*

[10] *Rodriguez Oldemar. Metodología para el desarrollo de proyectos en minería de datos CRISP-DM. 2010.*

[11] *Quinlan. J.R. 1994. Tesis: Algoritmo C4.5 : C 4.5 para atributos continuos y discretos.*

Biografía Autor(es)

a. Doctor en Ciencias de la Electrónica. Vicerrector de Investigaciones, Universidad Cesmag. jajimenez@unicesmag.edu.co. Pasto, Colombia.

b. Magister en Docencia Universitaria. Docente tiempo completo, Universidad Cesmag. cfgonzalez@unicesmag.edu.co. Pasto, Colombia.

c. Magister en Educación. Docente tiempo completo, Universidad Cesmag. jmmunoz@unicesmag.edu.co. Pasto, Colombia.

Aprendizaje por descubrimiento utilizando OVA, en el desarrollo del laboratorio de Física Mecánica

Jose Maria Muñoz Botina (a)
Fredy Santacruz (b)
Carlos Fernando González Guzmán (c)

Resumen

Las TIC, han estado presentes en el campo de la educación, son una herramienta importante en el mejoramiento de los procesos de enseñanza– aprendizaje. El uso de una estrategia de aprendizaje por descubrimiento apoyado en OVA contribuye un mejor aprendizaje en el laboratorio de Física Mecánica de la Universidad de Nariño.

La propuesta fue el resultado de aplicar un paradigma mixto con tres grupos de ingenierías (sistemas, electrónica y civil) y un cuarto por estudiantes de primer semestre de Física. Se implementó el uso de un OVA, aplicando la metodología ASSURE, previamente el docente idea una metodología de enseñanzaaprendizaje que promueve el trabajo colaborativo y el uso de herramientas TIC.

Finalmente, la estrategia de aprendizaje no solo es la aplicación del OVA, sino también la propuesta de un mapa conceptual como técnica para apropiación de conocimientos teóricos y experimentales en situaciones reales y de esta forma poder evidenciar un verdadero aprendizaje significativo en el grupo.

Palabras clave: TIC, OVA, aprendizaje colaborativo, aprendizaje significativo, laboratorio de física.

Learning for discovery using OVA, a in the development of the laboratory of Mechanical Physics

Abstract

ICT have been present in the educational field, favoring and strengthening teaching-learning processes. The use of a learning strategy for discovery supported by OVA contributes to better learning in the Mechanical Physics laboratory of the University of Nariño.

The proposal was the result of applying a mixed paradigm with three groups of engineering (systems, electronics and civil) and a quarter by students of first semester of Physics. The use of an OVA was implemented, applying an ASSURE methodology, previously the teacher devised a teaching-learning methodology that promotes collaborative work and the use of ICT tools.

Finally, the learning strategy is not only the application of the OVA, but also the proposal of a conceptual map as a technique for appropriation of theoretical and experimental knowledge in real situations and in this way to be able to demonstrate a true meaningful learning in the group.

Keywords: ICT, OVA, collaborative learning, meaningful learning, physics laboratory.

I. INTRODUCCIÓN

Los procesos de enseñanza – aprendizaje han evolucionado constantemente a lo largo de la historia en el ámbito educativo, se consolidan y fortalecen con el uso del componente tecnológico, entendiéndose como una herramienta que apoya el proceso pedagógico y didáctico en el salón de clases, en este orden de ideas la Organización de las Naciones Unidas para la Educación, Ciencia y Cultura (UNESCO) afirma: Las TIC pueden contribuir al fortalecimiento y la gestión de la planificación educativa democrática y transparente, las tecnologías de la comunidad pueden ampliar el acceso al aprendizaje, mejorar la calidad y garantizar la integración en el desarrollo del conocimiento [13].

El uso y apropiación de las TIC contribuyen a un acceso global de la educación, a lograr una igualdad en el acceso a la información y mejorar la calidad en el ejercicio de la educación en cuanto al proceso de aprendizaje de los estudiantes y el desarrollo profesional de los docentes universitarios. El entorno educativo en su contexto general y más específicamente en el universitario, supone innovadoras formas de aprendizaje, con nuevas estrategias didácticas, lo que proporciona nuevos retos en el contexto educativo.

Igualmente, de acuerdo con Ramírez, Escalante & Salazar "en Latinoamérica, ya no únicamente, son evidentes las diferencias económicas y culturales sino también aquellas en función de la modernidad, principalmente, en el ámbito científico-tecnológico" [10]. Frente a esto queda claro que el problema no radica en adicionar más ciencia o tecnología o por el contrario estar al margen de ellas, se hace necesario generar capacidades reales en nuestras sociedades para asumir esta realidad y apropiarse críticamente de ella.

En lo concerniente al uso de las TIC, como una herramienta para fortalecer y mejorar el proceso de enseñanza aprendizaje en el aula de clases, el uso y la aplicación de las diferentes innovaciones tecnológicas generan nuevas formas de comunicación que se traducen en canales que difunden diferentes prototipos de comportamiento social tales como las actitudes, valores y formas de organización [11]. Además, el trabajo práctico acompañado de una buena disposición de herramientas didácticas genera motivación, resignificación de conocimientos, oportunidades de enseñar diferentes y ofrecer recursos que no se encuentran en otros ámbitos. Todo lo anterior, es el resultado de aplicar óptimamente las TIC dentro del proceso educativo actual.

Es así como la educación no es ajena a esta evolución y cambios y en lo relacionado a los procesos de enseñanza - aprendizaje busca nuevas formas y mecanismos donde el docente ya no promueva un aprendizaje de tipo netamente memorístico sino por el contrario ampliar el conocimiento de los estudiantes desarrollando y promoviendo la imaginación y recurriendo a los procesos de investigación derivados del acercamiento al contexto cotidiano.

En el aprendizaje significativo el alumno tiene la posibilidad de relacionar los conocimientos aprendidos con aquellos que son nuevos generando nueva información con una estructura cognoscitiva dirigida al proceso de aprendizaje, es asi como se puede expresar que aprender de forma significativa, consiste en que el estudiante puede comprender, elaborar y asimilar diferentes diferentes conceptos en el proceso de aprendizaje, se puede llegar a afirmar que el aprendizaje significativo es una combinación del componente cognitivo como afectivo, logrando así que el aprendizaje por parte del estudiante sea personalizado, todo el aprendizaje en el salón de clases puede ser situado a lo largo de dos dimensiones independientes: la dimensión repetición-aprendizaje significativo y la dimensión recepción – descubrimiento. En el pasado se generó mucha confusión al considerar axiomáticamente a todo el aprendizaje por recepción y a todo el aprendizaje por descubrimiento como significativo [12].

Particularizando en el aprendizaje significativo por descubrimiento, se entiende como una estrategia que permite lograr que el docente brinde la posibilidad al estudiante de aprender por sí mismo, no existe una comprensión real de alumno con su contexto cotidiano se necesita que se aplique el conocimiento a otros contextos y situaciones donde el aprender sea la descripción e interpretación de diferentes situaciones relacionadas con factores relevantes, seleccionando y aplicando reglas, métodos y poder construir así el estudiante sus propias conclusiones."[3].

De la misma manera, "cuando los alumnos son impulsados a trabajar de forma autónoma en la fase preexperimental para, entre otras cosas, estimular su creatividad, se les incita a que sean independientes, diseñen trabajos, planteen hipótesis, las comprueben o las rechacen" [9], es decir los estudiantes tienen la posibilidad de crear su propio conocimiento, de generar sus propias conclusiones, de formular resultados y encontrar soluciones.

La implementación de OVA, como una herramienta de TIC, estimula sanamente la actividad mental de los estudiantes, ya que es un material de soporte digital y carácter educativo diseñados y creados en pequeñas unidades con temas como el contenido educativo, el evaluativo y el complementario, tiene la característica de poder ser reutilizable en el proceso de aprendizaje [14], de la misma manera lo define como: materiales de soporte diverso con el objeto de facilitar un medio de aprendizaje más completo y por medio de diferentes vías, de modo que unos se complementen y refuercen a los otros a través de texto, videos, galerías de imágenes, animaciones, ejercicios y cuestionarios autoevaluables, ejercicios de puzle, simulaciones de laboratorio, gráficos, etc.[6]

Los OVA utilizan diferentes tipos de entornos, y medios multimedia durante su implementación lo que permite llamar la atención del estudiante y motivación al momento de realizar su proceso de aprendizaje.

Los OVA son recursos digitales estructurados de una manera significativa, auto-contenibles y reutilizables, asociados a un propósito educativo y el cual debe contar con algunos de los componentes internos, tales como: contenidos, actividades que promuevan el aprendizaje y elementos de contextualización. De esta forma se desarrollan espacios en donde los escenarios de apropiación de nuevos conocimientos y experiencias por parte del estudiante le permitan generar procesos sencillos de análisis, reflexión y apropiación [1].

II. METODOLOGÍA

La estrategia de aprendizaje significativo por descubrimiento y la utilización de una herramienta en OVA, requiere el uso y apropiación de una metodología de diseño instruccional como lo es ASSURE incorpora los eventos de instrucción de Robert Gagne que permitan asegurar el uso efectivo de los medios en la instrucción, la teoría de aprendizaje en la que se basa es el constructivismo y que se fundamenta en tomar al estudiante de forma individual y particularizada, así como sus estilos de aprendizaje con una participación activa y comprometida del estudiante, a partir de una serie de etapas como son: el análisis de las características del estudiante, el establecimiento de objetivos de aprendizaje, la selección de estrategias, tecnologías, medios y materiales, la organización de escenarios de aprendizaje, la participación de los estudiantes y la evaluación y la revisión de la implementación y resultados del aprendizaje.

Por ende, el propósito de esta investigación es básicamente proponer una estrategia didáctica de aprendizaje que mejor se adapte al entorno del Laboratorio de física mecánica de la Universidad de Nariño, utilizando como herramienta TIC el uso de OVA.

Se parte de este principio para proponer una estrategia que enmarque los principales fundamentos del aprendizaje significativo por descubrimiento en conjunto con las TIC por medio del OVA en la temática de la segunda ley de Newton, en el laboratorio de Física Mecánica de la Universidad de Nariño. Es importante recalcar que dicha estrategia de aprendizaje no solo es la aplicación del OVA, también por otro lado está la propuesta de un mapa conceptual como técnica para apropiación de conocimientos teóricos y experimentales en situaciones reales y de esta forma evidenciar el aprendizaje significativo.

Para ejecutar el proceso de aplicación, desarrollo y apropiación del OVA en el laboratorio se dispone de una serie de prerrequisitos para obtener resultados claros y asertivos en el contexto educacional. Estos prerrequisitos abarcan desde el proceso de enseñanza, pasando por el proceso de aprendizaje y concluyendo con el OVA calificado de acuerdo con el objetivo general, como la herramienta digital principal para ejecutar este proceso. Simultáneamente, se tiene en cuenta las condiciones de infraestructura del laboratorio, los materiales, equipos e instrumentos con los que se cuenta, el personal y las herramientas educativas como libros, preinformes, bases teóricas, clases magistrales; y así contextualizar el proceso de enseñanzaaprendizaje y desarrollar una forma de enseñanza basada en la discusión, en la que los alumnos examinen críticamente los datos bajo la dirección del profesor.

Finalmente se esTable.ce como objetivo general proponer una estrategia didáctica a partir de herramientas OVA, que fortalezca el proceso de enseñanza– aprendizaje del Laboratorio de Física Mecánica de la Universidad de Nariño, a la par se generan unos objetivos específicos que van desde el análisis de estrategias predominantes de enseñanza por parte de los docentes , la caracterización de los proceso de aprendizaje en los estudiantes, el desarrollo de una experiencia en producción académica basada en OVA y por último la validación del OVA para el fortalecimiento en el aprendizaje significativo por descubrimiento.

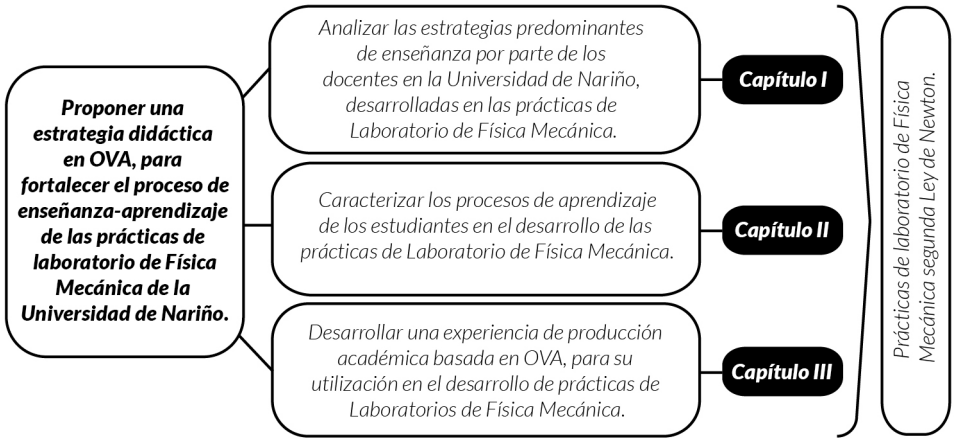


Fig 1. *Objetivos de investigación. Divididos en un objetivo general y tres objetivos específicos, que conllevan a una propuesta de investigación.*

Propuesta

En primer lugar, es importante mencionar que aquí no se trata de desligar al docente en el proceso de enseñanza- aprendizaje, más bien el docente es un medio, es la interfaz entre el estudiante y el OVA, es él quien dinamiza el proceso y se encarga de beneficiar al estudiante al explorar su independencia y autoaprendizaje que, de cierta manera, al contener temas teóricos previos, es más fácil acoplarlos a un modelo simulado y real. La ayuda del docente debe entenderse, al igual que la propia construcción que realiza el aprendiz, como un proceso, que permita la adaptación dinámica y contextual [8].

El docente como medio de ayuda al aprendizaje virtual, por tanto, no se trata de solo presentar información o de exponer actividades que pueda desarrollar el alumno. Es, fundamentalmente, seguir de modo continuo el proceso de aprendizaje que el educando desarrolla, y brindarle los apoyos y soportes necesarios en aquellos instantes en los que se requiera. Al momento de introducir el OVA al entorno educacional se abre un nuevo espacio donde no solo el estudiante y el profesor son actores principales de su proceso de enseñanzaaprendizaje. También se abre paso a ubicar a las TIC, como instrumentos digitales prácticos, cuya misión principalmente es la "realización de actividades en conjunto" entre educador y el educando: sólo comenzando con esa actividad conjunta se podrá efectuar una intervención sensible y eventual que facilite realmente al educando el ir más allá de lo que su interacción individual con el contenido le admitiría realizar. [8]

Sumando a lo anterior, las TIC mejoran la comunicación entre alumnos, favoreciendo el aprendizaje cooperativo al facilitar la organización de actividades grupales [4]. Para el caso de desarrollo de la estrategia de aprendizaje por medio del OVA, el trabajo que se visualiza en múltiples ocasiones en los estudiantes es el llamado trabajo colaborativo o aprendizaje colaborativo. En él los estudiantes conceptualizan un proceso de interacción donde se comparte, negocia y construye significados para solucionar un problema y se resalta la importancia de la comprensión simultánea.

En otras palabras, todos son un todo y cada uno corresponde a un engranaje que sin la disposición de otro no puede llegar a funcionar de manera adecuada. Es así como la construcción del conocimiento desde de la indagación de las interconexiones entre las diferentes ideas de los partícipes es fundamental a la hora de plantear un OVA como herramienta pedagógica donde sea posible comprometer a los estudiantes en soluciones colaborativas de los problemas de conocimiento.

III. RESULTADOS

Durante el periodo académico correspondiente al semestre B del 2017, acudieron a los Laboratorios de Física Mecánica I, 25 grupos de estudiantes de las diferentes carreras que se ofrecen en la Universidad incluyendo las sedes de Ipiales, Túquerres y Tumaco, pertenecientes a las Facultades de Ingeniería, Facultad de Ciencias Exactas y Naturales, Medicina y a la Facultad de Educación. Los 25 grupos representan una población de aproximadamente 500 estudiantes (20 estudiantes en promedio por cada grupo de trabajo). Sin embargo, para llevar a cabo el proceso de validación en esta investigación, se escogieron únicamente cuatro grupos deestudiantes, donde tres grupos corresponden al núcleo común de ingenierías, específicamente a los programas de Ingeniería Electrónica, Ingeniería de Sistemas e Ingeniería Civil; matriculados en distintos semestres de los diferentes programas. Y un cuarto grupo experimental estuvo conformado por estudiantes del primer semestre de la carrera de Física.

Table 1. *Clasificación de los estudiantes en grupos de trabajo*

Grupos	Subgrupos	Estudiantes	Total
1	1.1	4	16
	1.2	4	
	1.3	4	
	1.4	4	
2	2.1	4	15
	2.2	4	
	2.3	4	
	2.4	3	
3	3.1	3	8
	3.2	3	
	3.3	2	
4	4.1	4	12
	4.2	4	
	4.3	4	

Fuente: *Esta investigación.*

El primer grupo experimental con 17 estudiantes, de los cuales asistieron 16 estudiantes para realizar el preinforme; fueron organizados en cuatro grupos de trabajo estos conformados por cuatro estudiantes cada uno. Para el segundo grupo experimental se contó con 15 estudiantes, organizados en cuatro grupos de trabajo donde cada uno tenía cuatro estudiantes y un grupo de tres alumnos. Con el tercer grupo de tan solo ocho estudiantes que asistió a la tercera sesión del proceso de validación, se formó tres grupos de trabajo; dos grupos de tres estudiantes cada uno y un tercer grupo con dos de ellos. Por último, a la cuarta sesión asistieron 12 estudiantes, de un grupo de 13 estudiantes matriculados al primer semestre del programa de Física.

Por consiguiente, se conformó cuatro grupos de trabajo con tres estudiantes cada uno. En los dos grupos que faltaron dos estudiantes (uno en cada uno) se les realizó el preinforme individualmente. No se debe desconocer, entonces, la importancia que ejerce el trabajo colaborativo en los estudiantes, así como, el trabajo mediador del computador y móvil como herramientas en el aprendizaje, ya que influye positivamente en el conocimiento de experiencias reales y realza los conceptos teóricos aplicados a la simulación.

Se debe partir de la importancia por compartir lo que se va a desarrollar y también se debe asignar responsabilidades, dado esto no se contempla al aprendiz como persona solitaria sino en interacción con las demás personas, el aprendizaje colaborativo como elemento mediador de organización, favorece los procesos de interacción y de solución de problemáticas a manera conjunta [5]. Para llevar a cabo el OVA como estrategia de enseñanza- aprendizaje se escogió el tema de la Segunda Ley de Newton, este tema fue seleccionado por la importancia y dificultad que muestra el estudiante en su correcta comprensión. La propuesta metodológica que se implementó fue una combinación entre el uso de un OVA, la aplicación de un preinforme antes de la práctica de laboratorio, la explicación y el desarrollo de la práctica de laboratorio con los grupos conformados a través de una simulación en el laboratorio y, por último, se evalúa el conocimiento con una prueba escrita y la generación de un mapa conceptual. Con esta propuesta se busca eliminar la idea cada vez más creciente que hace que los alumnos,

cuando llegan a las universidades o al mercado de trabajo, lleven las deficiencias de aprendizaje adquiridas en los diversos niveles de enseñanza en cuanto a física se refiere.

La enseñanza de modo virtual presume del uso de material digital y por ende en Internet concurren varios recursos. Pero su uso en ocasiones es limitado por su formato de soporte, la obsolescencia de la tecnología con que fueron esTable.cidos, unificada al fallo de modularidad o la falta de conexión a internet dificultan su reutilización en un entorno puramente digital.

Continuando con la propuesta metodológica, se llevó a cabo el preinforme, éste fue realizado de forma virtual a través de la herramienta Formularios de Google *(Google Docs)* y aplicado a los 15 grupos y a los 2 estudiantes de manera individual (tal como se lo explica anteriormente). El Formulario de Google donde está estructurado el preinforme fue enviado vía enlace a uno de los integrantes del grupo para su posterior resolución, una vez respondidas las preguntas en el tiempo acordado, el formulario fue cerrado a la recepción de respuestas. El preinforme contó con ocho (8) preguntas de opción múltiple con única respuesta, todas las preguntas están relacionadas con la segunda ley de Newton y se reforzó con la herramienta OVA en la segunda Ley de Newton.

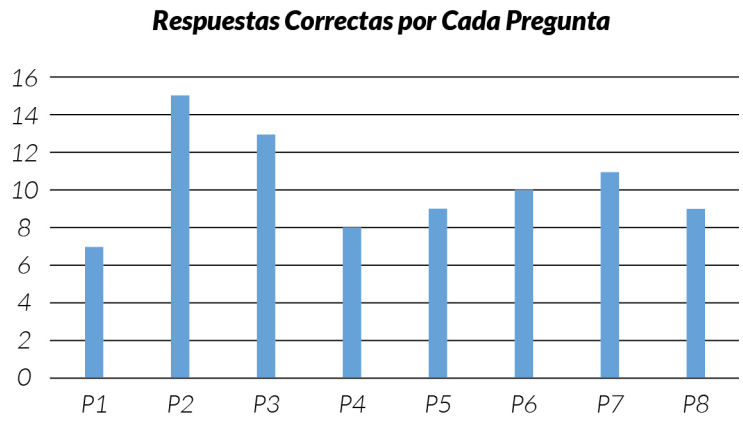


Fig 2. *Número de respuestas correctas en el preinforme. Son un total de 8 preguntas con múltiples opciones y única respuesta.*

El total de cuestionarios respondidos vía online, fueron 17. Se puede determinar mediante la gráfica que las respuestas fluctúan de pregunta a pregunta, así, para la primera pregunta (P1) de los 17 grupos tan solo 7 respondieron correctamente, siendo la pregunta con mayor dificultad al momento de responderla. Paralelamente la segunda pregunta cuenta con 15 grupos que respondieron acertadamente. Las preguntas número 4, 5, 6, 7, 8 (P4, P5, P6, P7 y P8) fueron respondidas correctamente en una proporción mayor al promedio general de grupos participantes, esto demuestra que el OVA influye positivamente en el conocimiento de los estudiantes, sin embargo, como se observa también en la misma figura los resultados de P4, P5, P6, P7 y P8 pueden mejorar si hay acompañamiento antes, durante y después de la implementación del OVA por parte del docente.

Uno de los planteamientos que se tiene en cuenta por parte del docente en el desarrollo de esta investigación, es la generación de una práctica de laboratorio en torno a la temática de la segunda ley de Newton; en dos horas de clase programadas según el

horario de actividades de los estudiantes. Esta práctica tuvo en cuenta la simulación llevada a cabo en el OVA Fuerzas y movimientos: Fundamentos, con la finalidad de reforzar lo aprendido durante la aplicación del instrumento digital. No obstante, en esta ocasión hubo acompañamiento del docente y lo más importante, se mostró un video con la simulación real en el laboratorio de los procesos a lograr durante la práctica de laboratorio; esto facilitó la apropiación del tema y se visualizó de manera más concreta el estado de aprendizaje en los alumnos. Es necesario reconocer que el reto que impone la sociedad actual del conocimiento como educadores, es el de lograr que los estudiantes aprendan a construir su propio destino, que ocupen el rol principal en el proceso de enseñanza-aprendizaje y que el docente asuma ser solo el orientador quien guie en el camino del autoaprendizaje, en el cual los alumnos son los principales responsables. (De Majo, 2005)

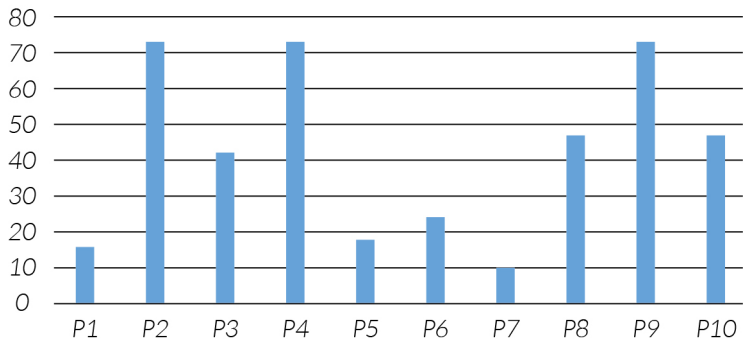


Fig 3. *Respuestas correctas de la preprueba (prueba diagnóstica).*

Aquí se observan los resultados correctos de la prueba diagnóstica con el objetivo de conocer el contexto de aprendizaje en los estudiantes que usan el laboratorio de Física.

Todo esto parece confirmar la idea de los estudiantes al referirse a la falta de instrumentos, equipos y fallas técnicas de los computadores en el laboratorio de Física Mecánica como la principal mejora a tener en cuenta, pues es necesaria la adquisición de nuevos materiales o bien la actualización de los mismos para mejorar la experiencia de simulación real de los ensayos y el reconocimiento de los aprendizajes en los alumnos.

Al mencionar a la prueba diagnóstica como una prueba previa, quiere decir que fue aplicada en instancias anteriores a la ejecución y desarrollo del OVA. Simultáneamente, estos resultados también proporcionan información comparativa entre un entorno libre de la propuesta metodológica y un entorno con la propuesta metodológica (OVA, Preinforme, práctica de laboratorio, simulación y evaluación- mapas conceptuales). No solo el preinforme muestra un avance en los estudiantes en la asimilación de conceptos de acuerdo con la segunda ley de Newton, posterior a todas las anteriores actividades, se ejecutó una evaluación del conocimiento mediante un informe de laboratorio grupal. En este se contempló la recolección de datos y su variación conforme a la alteración de variables.

De este modo los estudiantes eran conscientes en la ventaja del nuevo aprendizaje. Hasta este punto, el estudiante ya ha contado con diferentes estrategias que le ha permitido ahondar en el conocimiento de la Segunda Ley de Newton, estas estrategias son trabajadas involucrando al docente y al estudiante al proceso, además de un trabajo colaborativo.

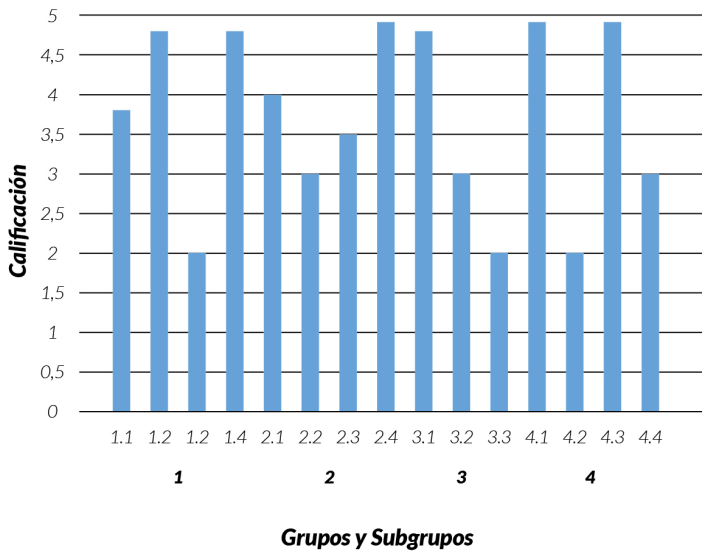


Fig. 4. *Variación de la calificación para cada grupo de trabajo.*

En la figura N°. 4 se observa la calificación para cada uno de los subgrupos que realizaron el informe de laboratorio grupal y la esquematización de los conceptos teóricos y experimentales en mapas conceptuales. En total son 17 los grupos que hicieron el preinforme, sin embargo, solo 15 de ellos completaron la propuesta metodológica. De los 15 grupos calificados, 9 grupos están por encima de la calificación de 3.0 y 6 grupos por debajo de la nota de 3.0. Esto comprueba que la propuesta metodológica por una parte afianzó los conocimientos de los estudiantes en cuanto a la temática trabajada.

Los mapas conceptuales son una técnica, "para exteriorizar la organización conceptual cognoscitiva que uno tiene para cierto conocimiento y puede ser utilizada como instrumento de acceso a la estructura cognoscitiva para lograr un objetivo". Los mapas conceptuales permiten que profesores y alumnos pueden dialogar y compartir significados diferentes para conceptos y relaciones entre conceptos [7]. Teniendo en cuenta lo anterior, el proceso de la propuesta metodológica se completó en este punto, cuando los estudiantes encontraron semejanzas entre los conceptos teóricos y los vincularon con los experimentos realizados en el laboratorio. Se completa, entonces un ciclo que permite generar un aprendizaje significativo en los alumnos y proveer estos conocimientos a largo plazo en el contexto real. Finalmente, se realizó una autoevaluación de esta propuesta metodológica con 12 preguntas.

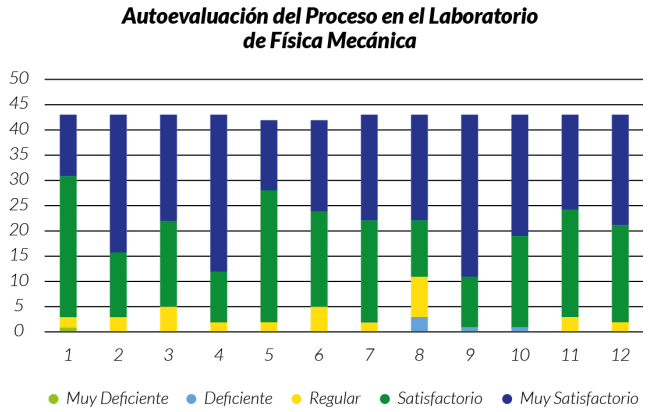


Fig.5. *Autoevaluación del proceso en el laboratorio de Física Mecánica.*

Se observa que los estudiantes en general calificaron la propuesta metodológica como muy satisfactoria y satisfactoria. En muy pocas preguntas califican como regular la temática expuesta y tan solo en una sola pregunta hay una calificación de muy deficiente.

IV. CONCLUSIONES

Los estudiantes son un actor principal en el proceso de enseñanza- aprendizaje, al ser parte fundamental de la educación son pioneros en contextualizar al docente en el desarrollo de las actividades y ser claves en la autonomía de su propio trabajo. Por su parte, el docente, como líder y ejecutor de las actividades, debe ser promotor de la autorregulación como cualidad esencial para cada grupo de trabajo, ya que a través de ésta se formulan metas concretas, se planifican actividades para el logro de esas metas, se monitorea el desempeño durante la ejecución de tales actividades y se evalúan a sí mismos de forma continua.

El docente no promueve el aprendizaje de tipo memorístico ya que debe ampliar el conocimiento de los estudiantes desarrollando y promoviendo la imaginación y usando las TIC a través del uso de OVA, como una herramienta en el proceso de enseñanza y aprendizaje en el laboratorio de física mecánica de la Universidad de Nariño.

El aprendizaje significativo por descubrimiento es una estrategia que mejora el rendimiento en el aprendizaje por parte de los estudiantes, los docentes ofrecen sus a sus estudiantes nuevas oportunidades de aprender por si mismos y construir su propio conocimiento [2], apoyado en el uso de una metodología como es ASSURE, de diseño instruccional, así como la implementación por parte del estudiante del uso de mapas conceptuales desarrollados por los estudiantes como una técnica que mide el logro del aprendizaje alcanzado por cada uno de ellos.

REFERENCIAS

[1] *Avila , P., & Bosco, M. D. (2001). Ambientes virtuales de aprendizaje, una nueva experiencia., (pág. 6). Düsseldorf, Germany.*

[2] *Bruner,J. (1966). Toward a Theory of Instruction. Cambridge, MA: Harvard University Press.*

[3] *Bruner, J. (1980). Investigación sobre el desarrollo cognitivo. España: Pablo del Río.*

[4] *Cenich, G. y Santos G. (2005). Propuesta de aprendizaje basado en proyectos y trabajo colaborativo: experiencia de un curso en línea. Revista Electrónica de Investigación Educativa.*

[5] *Gros, M. (2007). Experiencia de trabajo colaborativo: formación del personal docente universitario para la diversidad. Revista de Ciencias Sociales (Cr) [en linea] 2014, I [Fecha de consulta: 4 de octubre de 2018] Disponible en:<http://www.redalyc.org/articulo.oa?id=15333871003> ISSN 0482-5276 .*

[6] *Morales, A. O. (2003). Fundamentos de la Investigación Documental y la Monografía. En Manual para la elaboración y presentación de la monografía. Universidad de Los Andes, 20. Retrieved from http://www.saber.ula.ve/bitstream/123456789/16490/1/fundamentos_investigacion.pdf.*

[7] *Novak, J.D. and Gowin, D.B. (1984). Learning how to learn. Cambridge, Cambridge University Press.*

[8] *Onrubia, J. (2005). Aprender y enseñar en entornos virtuales: actividad conjunta, ayuda pedagógica y construcción del conocimiento.*

[9] *Oroño, S. U., & Cafferata, M. T. (2007). La enseñanza de procedimientos en el laboratorio de ciencias de formación docente. Cuadernos de Investigación Educativa, 2, 89–103.*

[10] *Ramírez, A., Escalante, S., & Salazar, L. (2008). La educación en tecnología: un reto para la educación básica venezolana. Educere, 12, 731–740.*

[11] *Salvarrieta, R. (2008). Lacoctelera. Recuperado el 2016, de http://lesalavarrieta.lacoctelera.net/post/2009/07/22/importancia-las-tics-la-educacion.*

[12] *Serrano, G. (19). (1990). La investigación acción: aplicaciones al campo social y educativo, Madrid.*

[13] *Unidas, C. N. (2016). Comisión Economica para América Latina y el Caribe. Obtenido de http://www.cepal.org/es.*

[14] *Wiley DA. Connecting learning objects to instructional design theory: A definition, a metaphor and a taxonomy. En: D. A. Wiley. TheInstructional Use of Learning Objects:Online Version [monografía en Internet]. 2000. Disponible en: http://reusability.org/read/chapters/wiley.doc.*

Biografía Autor(es)

a. Magíster en Educación. Docente tiempo completo, Universidad Cesmag. jmmunoz@unicesmag.edu.co. Pasto, Colombia.

b. Magíster en Docencia Universitaria. Docente, Universidad de Nariño. E-Mail: Fredysanta62@gmail.com. Pasto, Colombia.

c. Magíster en Docencia Universitaria. Docente tiempo completo, Universidad Cesmag. cfgonzalez@unicesmag.edu.co. Pasto, Colombia.

Aproximación metodológica para el desarrollo de sistemas interactivos basado en triadas de modelos

Alexandra Ruiz Gaona (a)
William Joseph Giraldo (b)
José Luis Arciniegas (c)

Resumen

Este artículo presenta una aproximación metodológica para el desarrollo de sistemas interactivos basado en tres modelos: modelo de dominio, de interacción y presentación. Este enfoque de ingeniería pretende unificar los modelos a partir de los cuales se generan sistemas interactivos. Igualmente, este enfoque se busca que el desarrollo de sistemas interactivos no parta desde cero, de tal forma que se promueva el prototipado rápido funcional y el producto mínimo viable. Una prueba de concepto en el contexto educativo es llevada a cabo como parte del proceso de validación de la aproximación metodológica propuesta.

Palabras clave: Diseño de interfaces, desarrollo dirigido por modelos, HCI, MBUID.

Methodological approach for systems development triad-based interactive of models

Abstract

This article presents a methodological approach for the development of interactive systems based on three models: domain, interaction and presentation. This engineering approach aims to unify the models from which are generated interactive systems. Likewise, this approach seeks that the development of interactive systems does not start from scratch, in such a way as to promote functional rapid prototyping and minimum viable product. A proof of concept in the educational context is carried out as part of the validation process of the proposed methodological approach.

Keywords: Interface Design, Model Driven Development, HCI, MBUID.

I. INTRODUCCIÓN

Los desarrolladores de sistemas interactivos se enfrentan a diferentes factores que deben resolver para satisfacer las necesidades de los usuarios. Dentro de estos factores se encuentran: (a) los diferentes perfiles de usuario. (b) diversidad de dispositivos de despliegue (smartphones, computadores, Table.ts, etc) lo que trae consigo diversidad de mecanismos de interacción. (c) las múltiples plataformas de desarrollo y de lenguajes. (d) los diferentes aspectos a promover en la interfaz final (usabilidad, accesibilidad, gamificación, etc.). De esta manera, vemos como el desarrollo de sistemas interactivos se ha convertido en una tarea cada vez más difícil, lo cual repercute en una mayor inversión y tiempo de desarrollo de los sistema interactivos.

Uno de los enfoques que puede dar solución a esta problemática es el diseño de la IU basada en modelos o MBUID. MBUID es un enfoque que tiene como objetivo hacer frente a los problemas de heterogeneidad relacionados a dispositivos, usuarios y contextos de uso, reduciendo el esfuerzo necesario para el desarrollo de la IU.

La literatura reporta gran variedad de aproximaciones basadas en MBUID, tal como se evidencia en el estudio realizado por los autores [1]. En dicho estudio se observó que no hay un consenso sobre la cantidad y los tipos de modelos que se requieren para generar sistemas interactivos. Así pues, este artículo propone un enfoque de ingeniería para el desarrollo de sistemas interactivos basado en tres modelos: modelo de dominio, modelo de interacción y modelo de presentación. A partir de estos tres modelos, la aproximación metodológica permite generar sistemas interactivos, en los cuales es promovida la usabilidad en la interfaz final.

En consecuencia, los aportes de este artículo se centran en proponer una aproximación metodológica que facilita al desarrollador generar sistemas interactivos a partir de la triada de modelos que se ajustan al modelo mental que el usuario tiene de los datos y de las tareas. La aproximación metodológica se construye a partir de propuestas existentes que recogen los principios y principales aportes de diferentes propuestas de MBUID y del HCI. Una primer versión de esta aproximación se presenta en [2]. El presente artículo se enfoca en exponer una actualización de la aproximación metodológica, la cual tiene diferencias sustanciales respecto de su versión inicial.

Este artículo se encuentra estructurado de la siguiente manera: la sección 2 expone el método utilizado para esTable.cer la aproximación metodológica propuesta, la sección 3 detalla el ciclo de vida y el flujo de actividades de la aproximación, la sección 4 presenta los resultados luego de haber aplicado la aproximación metodológica a un contexto específico y, finalmente la sección 5 presenta las conclusiones y trabajo futuro.

II. MÉTODO

Éste trabajo de investigación utilizó el método de integración de notaciones y procesos creado por Giraldo [3] y detallado por Villegas [4], como una guía para conformar el recorrido metodológico para el desarrollo de la interfaz de usuario de sistemas interactivos basado en la triada de modelos.

El método está compuesto por 3 etapas:

Análisis: todos los pasos necesarios para realizar la adaptación de un proceso a partir del estudio de los distintos componentes metodológicos existentes y de interés. Este abarca la identificación de los beneficios y las deficiencias asociadas a las propuestas metodológicas que están depositadas en el estado del arte.

Diseño: describe la solución metodológica propuesta mediante un lenguaje de especificación de procesos de desarrollo. El objetivo es definir una configuración de proceso (que cumple un propósito específico) que describe un proceso de desarrollo en forma de un producto. Las actividades de esta fase son:

- *Formalizar las propuestas seleccionadas:* llevar las propuestas a un mismo nivel de formalización que haga posible la identificación de puntos de integración entre ellas.
- *Identificar los puntos de integración:* evaluar los distintos puntos donde es posible explotar la complementariedad de los procesos.
- *Aplicar fundamentos de integración:* llevar a cabo una definición de los contenidos finales que resultan de la unión de las propuestas y aplicar los fundamentos de integración.

Implementación: describe o representa una configuración de proceso sobre una herramienta de implementación de procesos de desarrollo en forma de producto. Por ejemplo: para el caso de un libro, sería definir las plantillas, las guías de estilo y poblar dichas plantillas utilizando dichas guías de estilo. Para el caso de una página Web, sería utilizar una herramienta de composición de procesos (ej. RMC6 o EPFC) que pueda generar la página Web que contiene el proceso.

El recorrido metodológico que se presenta en este artículo se obtuvo a partir de la ejecución de las etapas de análisis y diseño del método de integración de notaciones y procesos. La etapa de implementación no hace parte de las contribuciones de este artículo.

A continuación se detalla la aproximación metodológica propuesta para el desarrollo de sistemas interactivos basado en la triada de modelos.

III. APROXIMACIÓN METODOLÓGICA PROPUESTA

La aproximación metodológica propuesta se fundamenta en un nuevo enfoque de ingeniería basado en una triada de modelos. La triada de modelos se construye a partir del análisis de las tareas de alto nivel del negocio y los datos del dominio que las soportan. Los datos del dominio se conciben a través de la exploración de patrones de datos en un dominio específico. Los patrones de datos son estructuras valoradas por expertos que reconocen que tienen esa estructura, esos elementos y esas relaciones (modelo mental del experto). Sin embargo, poco se ha discutido en cómo se deben presentar en una interfaz de usuario.

Para ello, se realizan estudios con usuarios representativos, en los cuales se captura el modelo mental que el usuario tiene sobre los datos del dominio (modelo de dominio) y la forma como desea interactuar con esos datos (modelo de interacción) y visualizarlos (modelo de presentación) en el contexto de una tarea.

A partir del análisis de la información recolectada y, a través de un proceso iterativo, los diseñadores generan la interfaz de usuario de negocio (Business User Interface - BUI), la cual representa un diseño inicial de la interfaz de usuario que podría contener la forma y la semántica de los datos. La BUI se caracteriza por ser una interfaz de usuario independiente de la tecnología (informática para este caso), sin botones o widgets concretos. Por tanto, tomando como punto inicial una BUI, el desarrollo puede finalizar en una aplicación móvil, una aplicación de escritorio o en un formulario en papel.

Así pues, la triada es un triángulo en donde sus vértices son: los datos del dominio como los percibe el usuario (modelo de dominio), la forma como desea interactuar con esos datos (modelo de interacción) y la forma como desea visualizarlos (BUI o modelo de presentación).

A partir de la definición de la triada a nivel de negocio, el proceso de desarrollo se basa en detallar la información encapsulada en la triada pero ya a nivel de sistema. Este proceso se realiza para diferentes tareas y escenarios de uso, por lo que el resultado es un catálogo de triadas que puede ser usado para dar solución a problemas de diseño para un dominio específico. Finalmente, las triadas son implementadas con el apoyo de herramientas de modelado y generadores de código que generan componentes (con funcionalidad y variabilidad en su presentación) que renderizan particiones de la interfaz usuario.

De esta manera, la implementación de cada triada que conforma el catalogo es una aplicación (componente) que instrumenta una actividad o labor que puede realizar el usuario. Las actividades que realiza el usuario se ejecutan de acuerdo con un flujo u orden que puede variar dependiendo del dominio especifico que se aborde. El modelo que describe este flujo se denomina modelo de orquestación de actividades. El diagrama que representa este modelo puede variar dependiendo del dominio de aplicación. La aproximación metodológica contempla la creación del diagrama que representará dicho modelo.

Por medio de este enfoque, se busca que el desarrollo de la interfaz de usuario no parta desde cero, de tal manera que se promueva el prototipado rápido funcional y el producto mínimo viable.

En las siguientes secciones se presenta el ciclo de vida y el contenido detallado del recorrido metodológico basado en la triada de modelos.

3.1 Ciclo de vida

El ciclo de vida del recorrido propuesto (Fig. 1) esta particionado en las etapas de modelado de Negocio (Business Modeling), Especificación (Specification), Análisis y Diseño (Analysis and Design) e Implementación (Implemantation), y está diseñado para que sea fácilmente acoplable con las metodologías tradicionales de la ingeniería de software. Igualmente, se observa en la Fig. 1, cómo el usuario participa activamente desde etapas tempranas y durante todo el proceso, el cual es iterativo e incremental y se apoya en técnicas de prototipado y evaluación como uno de los mecanismos para promover la usabilidad en la interfaz final.

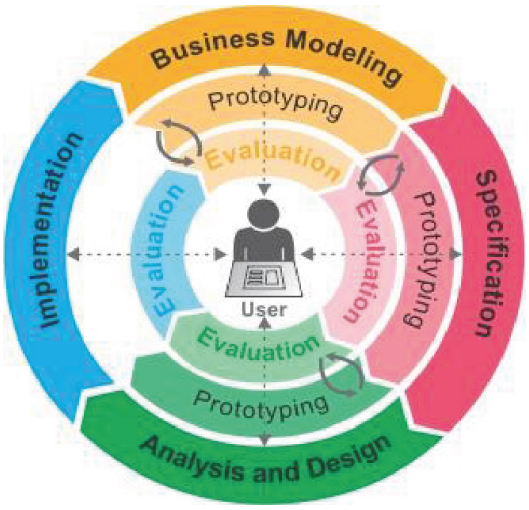


Fig. 1. *Ciclo de vida de la aproximación metodológica propuesta*

En la siguiente sección se expone el contenido metodológico detallado del recorrido propuesto.

3.2 Contenido detallado

El recorrido propuesto tiene dos flujos principales: el flujo para crear componentes (Fig. 2a) y el flujo para crear herramientas (Fig. 2b). El flujo para crear componentes produce componentes que renderizan particiones de la interfaz de usuario que pueden ser usados posteriormente en el desarrollo de sistemas interactivos en un dominio específico. El flujo para crear herramientas produce editores DSL y generadores de código para un dominio específico, los cuales apoyan el desarrollo de sistemas interactivos en dicho dominio.

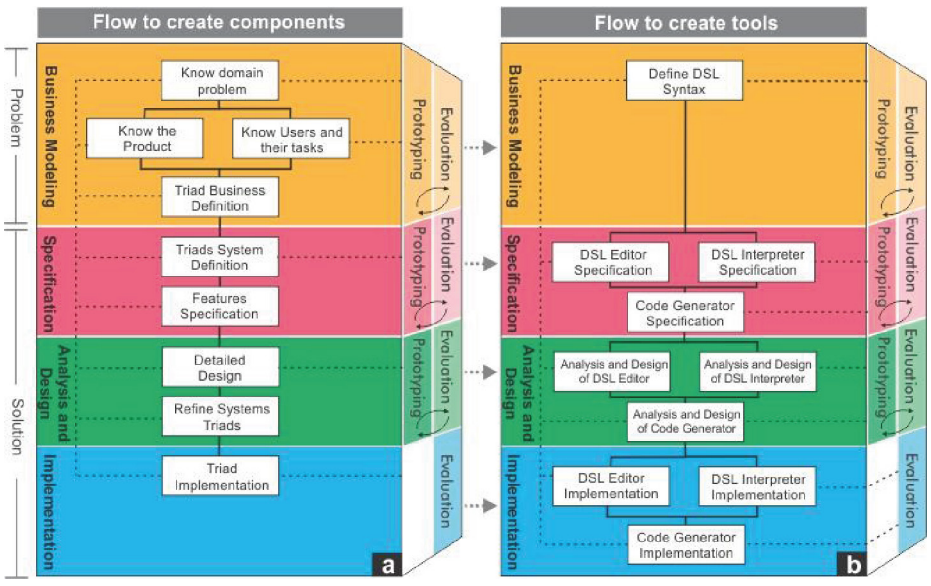


Fig. 2. *Contenido detallado de la aproximación metodológica propuesta*

A continuación se detalla cada flujo de acuerdo a las etapas que conforman la aproximación metodológica.

Flujo para crear componentes. El propósito de este flujo es generar un conjunto de componentes a partir de la implementación del catálogo de triadas, el cual puede ser usado para el desarrollo de sistemas interactivos en un dominio específico. Cada componente (con funcionalidad y variabilidad en su presentación) renderiza una partición de la interfaz de usuario. Cabe aclarar que para este flujo, se habla de componentes de primer nivel y de segundo nivel. Los componentes de primer nivel son los que se generan luego de ejecutar todo el flujo, es decir las pantallas. Los componentes de segundo nivel son particiones de la pantalla. En conclusión, un componente de primer nivel está formado por varios componentes de segundo nivel. A continuación se exponen las actividades de este flujo de acuerdo a las etapas del recorrido.

Modelado de Negocio (Business modeling): Esta etapa tiene como objetivo crear las triadas a nivel de negocio en el contexto de las tareas para las cuales se está diseñando. Para ello, se realizan las siguientes actividades (Fig. 2a): a) conocer el dominio del problema, b) conocer el producto, c) conocer los usuarios y sus tareas, y d) definir la triada a nivel de negocio. La actividad conocer el dominio del problema está encaminada a definir el conjunto de conceptos que caracterizan el dominio, con el fin de crear un cuerpo unificado de criterios que sirvan para comprender y analizar dicho dominio.

A través de revisiones literarias y otras estrategias se obtienen los conceptos dominio y el modelo de dominio preliminar que, posteriormente, es revisado por expertos. El modelo de dominio, además de integrar la triada, es un insumo que permite definir el lenguaje del diagrama que representará el modelo de orquestación de actividades. Posteriormente, se realizan tareas para conocer el producto como análisis de la competencia o pruebas de usabilidad. El objetivo de este tipo de tareas es definir o identificar los widgets y componentes de segundo nivel que caracterizan a las aplicaciones dentro de un dominio específico, con el fin de capturar ideas sobre funcionalidad, estilos de interfaz de usuario, componentes de segundo nivel, *widgets* principales, estructura de tareas, terminología, entre otros. Finalmente, se realizan acciones para conocer los usuarios y sus tareas como por ejemplo: identificar perfiles de usuario, realizar encuestas, entrevistas, observación contextual, etc. A través del análisis de tareas se espera identificar cuáles son las tareas y cuáles son los criterios que hacen que la tarea varíe. Hasta este punto se ha recolectado la información inicial para formar la triada como son los datos del dominio y las tareas del dominio, así como información del usuario y sus necesidades. Sin embargo, tanto los datos del dominio como las tareas del dominio están evaluados desde la perspectiva del experto y no del usuario final. Se hace necesario, por tanto, capturar el modelo mental que el usuario tiene de los datos y como desea interactuar con dichos datos en el contexto de cada tarea. Para ello, se identifican usuarios potenciales de acuerdo con los perfiles definidos y se les invita al laboratorio para capturar esta información. Con esta información, el diseñador refina el modelo de dominio, crea el modelo de interacción y genera una partición de la interfaz de usuario de negocio (BUI o modelo de presentación) en el contexto de la tarea evaluada. Para esta etapa, los tres modelos resultantes (modelo de dominio, modelo de interacción y BUI) se caracterizan por ser no modales, estar en el nivel físico y usar widgets abstractos.

Especificación (Specification): Las actividades de esta etapa son: a) definición de la triada a nivel de sistemas y b) Especificación de características. Para la especifica del catálogo de triadas a nivel de Sistema se realizan nuevamente pruebas con usuarios representativos. Como resultado se obtienen triadas de modelos que ahora son modales, están en el nivel del sistema y usan widgets concretos. Posteriormente, se crea la especificación de características del catálogo de triadas, un documento que especifica cómo

está particionada la IU y cuáles son sus características (componentes de segundo nivel, funcionalidad, variabilidad en la presentación, propiedades, etc.).

Análisis y diseño (Analysis and design): Esta etapa tiene como objetivo refinar el catálogo de triadas de sistema a partir de prototipos de interfaz. Los prototipos de interfaz se crean a partir del refinamiento de los bocetos que han sido el resultado de las pruebas con usuarios potenciales en el laboratorio. Para ello, se aplican patrones de diseño, heurísticas y demás lineamientos para el diseño gráfico detallado. Los prototipos, además, constituyen un insumo para el desarrollo del generador de código que se construye en el flujo alterno. De acuerdo con los prototipos finales obtenidos en la actividad anterior, el diseñador refina el catálogo de triadas pero manteniendo su descripción en un nivel concreto (modelos que son independientes de la tecnología pero dependientes de la modalidad).

Implementación (Implementation): En esta etapa se implementa el catálogo de triadas de acuerdo con la especificación de características. Es decir, se implementan los componentes de primer nivel (pantallas) y de segundo nivel (componentes y widgets que hacen parte de la pantalla), de acuerdo a la especificación de características. El generador de código es construido teniendo en cuenta esta implementación.

Flujo para crear herramientas. El propósito de este flujo es crear las herramientas que conforman el ambiente de desarrollo que da soporte a la generación de sistemas interactivos a través del enfoque de triadas. Dichas herramientas son de dos tipos: a) editores de modelos con sus respectivos interpretadores y b) generadores código para una tipo de tecnología. El proceso de creación de estas herramientas es apoyado por técnicas y frameworks, y es bien conocido por la comunidad de desarrolladores de MDD. En consecuencia, el aporte de esta investigación en este flujo se centra en describir los artefactos requeridos para el desarrollo y cómo los artefactos resultantes del flujo alterno son utilizados para la creación de las herramientas. A continuación se describe este flujo de acuerdo con sus etapas.

Modelado de Negocio (Business modeling): En esta etapa se definen la sintaxis abstracta y concreta del DSL que representará el modelo de orquestación de actividades en un dominio específico. La sintaxis abstracta actúa como una guía para la comunicación de todos los artefactos y herramientas que intervienen en el desarrollo de una herramienta de soporte a MDE. Una técnica común para la especificación de lenguajes es definir primero la sintaxis del lenguaje y luego describir su semántica estática y dinámica [3]. La sintaxis define qué elementos de modelado existen en el lenguaje y cómo estos elementos de modelado se construyen en términos de otros. Cuando el lenguaje tiene una sintaxis gráfica, es importante definir la sintaxis en una notación de forma independiente (sintaxis abstracta del lenguaje). La sintaxis concreta se define entonces mediante un mapping entre la notación y la sintaxis abstracta [5]. De esta manera, la especificación de la sintaxis concreta tiene como punto de partida el análisis visual de cómo se presentarán al usuario los elementos de modelado del lenguaje, el cual define la estructura de componentes gráficos usados para representar los conceptos. Para la definición de la sintaxis abstracta y concreta se tiene en cuenta el modelo de dominio generado en el flujo alterno y la aplicación de principios de diseño de notaciones como por ejemplo los propuestos por Moody [6]. La sintaxis y la semántica del lenguaje se validan con usuarios representativos, con el fin de identificar si el lenguaje es fácil de aprender, de usar y si se ajusta al modelo de negocio que describe.

Especificación (Specification): En esta etapa se refina la sintaxis abstracta y concreta de acuerdo con los resultados de las evaluaciones de la etapa anterior. Posteriormente, se definen las características *(features)* del editor DSL, lo cual se traduce en especificar de manera formal la sintaxis abstracta y concreta del DSL, con el fin de volverlo computacional. De igual manera, se definen las características para el interpretador del DSL. Para

ello, se definen los diferentes módulos que lo conformarán y se especifican los requisitos funcionales y no funcionales de dichos módulos. Inicialmente, el interpretador está compuesto por el motor del DSL que es el encargado de ejecutar los modelos DSL. Sin embargo, puede existir la necesidad de agregar otros módulos como por ejemplo un módulo de interoperabilidad que permita la comunicación con plataformas externas.

Análisis y diseño (Analysis and design): En esta etapa se diseñan los elementos del editor DSL, es decir, el lienzo (diagrama o sintaxis concreta), el browser (sintaxis abstracta) y la paleta, así como la relación entre ellos. Para ello se utiliza el listado de componentes y los sketches o prototipos generados en el flujo alterno. Finalmente, se diseña el interpretador del DSL. El interpretador DSL debe tener al menos un módulo llamado motor DSL, el cual es el encargado de recorrer y ejecutar los modelos DSL en tiempo de ejecución.

Implementación (Implementation): En esta etapa se generan los editores DSL, teniendo en cuenta la información de las etapas anteriores y utilizando *frameworks* de desarrollo para tal fin. Posteriormente, se crea el generador de código de acuerdo con la implementación de los componentes tanto de primer nivel como de segundo nivel que se crearon en el flujo alterno. Finalmente, se implementa el interpretador, compuesto por el motor del DSL y demás módulos que permitan la interoperabilidad con plataformas externas.

IV. PRUEBA DE CONCEPTO DEL RECORRIDO PROPUESTO

Para aplicar el recorrido metodológico propuesto es necesario identificar un dominio específico. En esta ocasión, se seleccionarán los sistemas interactivos basados en distribución de contenido de video como dominio de aplicación. Este tipo de sistemas se caracterizan por la importancia que dan al contenido de video dentro del modelo de negocio. Algunas de las aplicaciones más comunes son Netflix y Youtube. Sin embargo, este tipo de sistemas tienen diferentes contextos de aplicación. Se ha seleccionado el contexto educativo por la importancia que tiene el contenido de video en el proceso de enseñanza aprendizaje, ya que estimula los sentidos y los diferentes estilos de aprendizaje de los estudiantes.

Una vez definido el dominio del problema y su contexto, se inicia la ejecución del recorrido propuesto. Como se describe en el contenido detallado, el recorrido está compuesto por dos flujos. En ambos flujos son variados los artefactos que intervienen. Sin embargo, en esta ocasión, se detallará solo los artefactos más relevantes. En el caso del flujo para crear componentes se detallará un ejemplo de triada de modelos pasando por todas las etapas. En el caso del flujo para crear herramientas, se presentarán las herramientas creadas en la ejecución del recorrido para el contexto educativo.

4.1 Flujo para crear componentes

Este flujo inicia con el modelado de negocio en donde se estudia el dominio del problema, los productos y los usuarios y sus tareas. Para conocer el dominio del problema se realiza un análisis de diferentes propuestas en la literatura relacionadas con el acto didáctico ([7], [8], [9],[10], [11], entre otros). Como resultado del análisis se obtienen los conceptos más relevantes de este dominio. A continuación se presentan dichos conceptos.

De acuerdo con Marquès [11], la enseñanza se concreta en actos didácticos mediante los cuales el profesor o formador propone múltiples actividades a los estudiantes para facilitar los aprendizajes deseados. Estas actividades, conocidas como actividades de aprendizaje *(LearningActivity)*, se pueden clasificar en varios tipos de acuerdo a su naturaleza (asimilativas, gestión de información, de aplicación, productivas, etc.) [12]. Adicionalmente, las actividades de aprendizaje son adaptadas teniendo en cuenta las características de los estudiantes, los medios didácticos *(DidacticsMedium)* y recursos *(Resources)* disponibles y los contenidos objeto de estudio *(LearningUnits)*. Marquès plantea que la eficacia de los medios didácticos dependerá en gran medida de la manera en la que el profesor oriente su uso en el marco de la estrategia didáctica *(DidacticStrategy)* que está utilizando. Una estrategia es un procedimiento organizado y orientado a la obtención de una meta claramente esTable.cida. De esta manera, la estrategia didáctica es el procedimiento con el que el profesor pretende facilitar los aprendizajes en los estudiantes a través de actividades que contemplan la interacción de los alumnos con determinados contenidos.

Una vez identificados los conceptos del dominio es necesario conocer los usuarios y sus tareas. Para ello, se realizan encuestas (estudiantes universitarios) y observación contextual (actividades académicas llevadas a cabo en una institución educativa). El objetivo de la observación contextual es identificar los tipos de actividades de aprendizaje que se realizan en el aula de acuerdo con: (a) la naturaleza de la actividad. (b) quién, dónde, cuándo y cómo se realiza la actividad. (c) los artefactos necesarios para llevarla a cabo.

El modelo de datos del dominio y el modelo de tareas del dominio son identificados a partir de la aplicación de las técnicas anteriormente mencionadas. Una vez definidos estos dos modelos de datos y tareas, se procede a capturar el modelo mental que tienen los usuarios respecto de los datos del dominio y respecto de cómo desean interactuar con dichos datos y cómo desean visualizarlos en el contexto de una tarea específica. Posteriormente, las triadas a nivel de negocio son creadas a partir de la información capturada con los usuarios en el laboratorio.

Una de las actividades de aprendizaje más utilizada por los docentes son las actividades asimilativas, específicamente las de tipo clase magistral *(Lecture)*. Es por esta razón que se ha seleccionado este tipo de actividad de aprendizaje para desarrollar el ejemplo de las triadas. La Fig. 3 expone la triada a nivel de negocio para la actividad de aprendizaje *"Attend a lecture class"*.

La triada en la etapa de negocio está representada por tres diagramas: diagrama de interacción de negocio (Fig. 3a), diagrama de entidades de negocio (Fig. 3b) y diagrama de Interfaz de Usuario de Negocio - BUI (Fig. 3c). En el diagrama de interacción de negocio "Attend a lecture class" se describen las interacciones de negocio (Fig. 3a1), parte estructural del proceso y las tareas interactivas de negocio que llevan a cabo el profesor y el estudiante para describir la dinámica de dicho proceso (Fig. 3a2). El análisis de tareas interactivas de negocio da lugar a una descomposición de tareas. Las entidades de negocio que soportan el proceso se representan mediante el Modelo de Entidades de Negocio (Fig. 3b). Estas entidades representan un modelo inicial del contexto para el diseño de la interfaz de usuario de negocio representada a través del diagrama BUI (Fig. 3c).

El diagrama BUI se crea a partir de los datos, entidades que son manipuladas dentro del contexto de los procesos de negocio. El objetivo es diseñar un número de diagramas BUI que tenga concordancia con el número de sub-tareas de negocio. En la Fig. 3., se resalta la sub-tarea *"Listen to Speech"* (Fig. 3a3), la cual asocia el diagrama de entidades de negocio *"Lecture Domain"* (Fig. 3b) y el diagrama de interfaz de usuario de negocio *"Lecture BUI"* (Fig. 3c). De esta manera, cada sub-tarea es una triada, que relaciona un

diagrama de interacción (descomposición de la sub-tarea en pasos), un diagrama de dominio y un diagrama BUI. Lo que se busca a partir de la triada es tener un justo equilibrio entre el diseño de la interfaz de usuario basada en datos y el diseño de la interfaz de usuario basada en tareas.

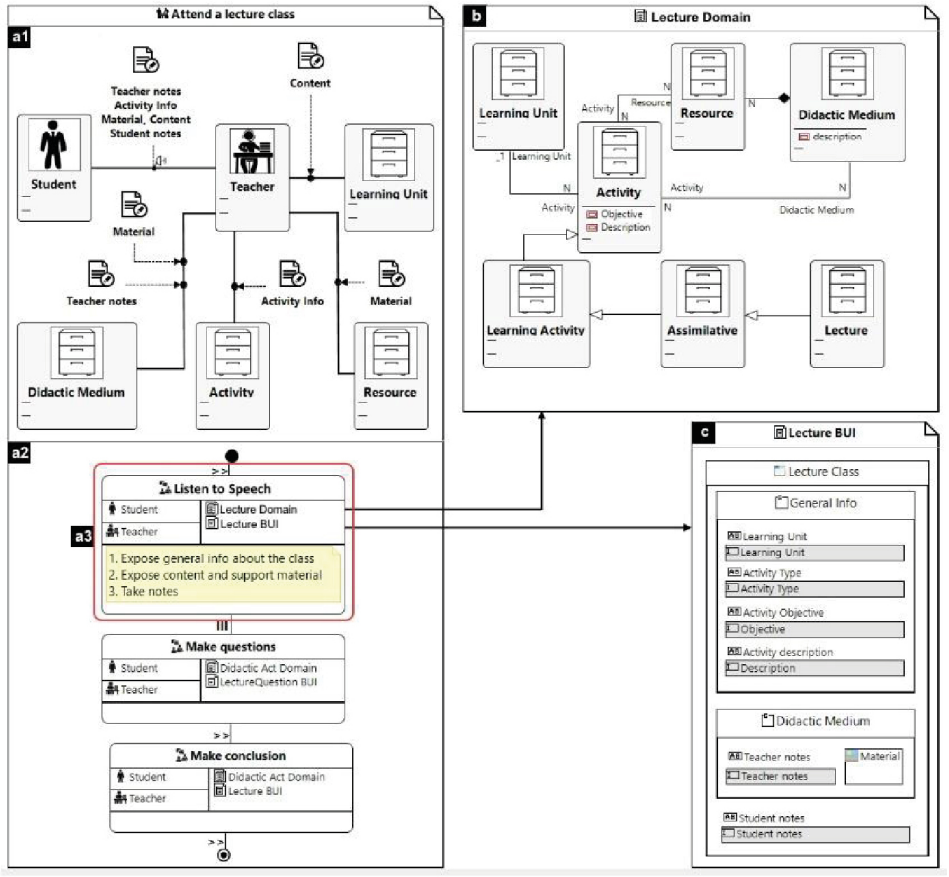


Fig. 3. *Triada a nivel de negocio para la actividad "Attend a lecture class"*

Así pues, se genera un catálogo de triadas a nivel de negocio, en el cual cada triada es creada en el contexto de una tarea específica. Una vez definido este catálogo, se procede con la etapa de especificación, en la cual se definen las triadas a nivel de sistema. Para ello, se llevan nuevamente los usuarios al laboratorio y se captura la información para tal fin. El ejemplo de triada para esta etapa no se presenta debido a restricciones de espacio. Sin embargo, la triada a nivel de sistema de la siguiente etapa reúne la información de ambas etapas.

Posteriormente, se procede con la etapa de análisis y diseño, en la cual se realizan prototipos finales para cada diagrama de presentación que conforma la triada. Para ello, se aplican patrones de diseño, heurísticas y demás lineamientos para el diseño gráfico detallado. Estos prototipos son el insumo para la construcción del generador de código que se desarrolla en el flujo alterno.

Finalmente, el catálogo de triadas de sistema es refinado a partir de los prototipos pero manteniendo su descripción en un nivel concreto (modelos que son independientes de la tecnología pero dependientes de la modalidad). Siguiendo el ejemplo que se viene desarrollando, la Fig. 4 expone la triada a nivel de sistema refinada.

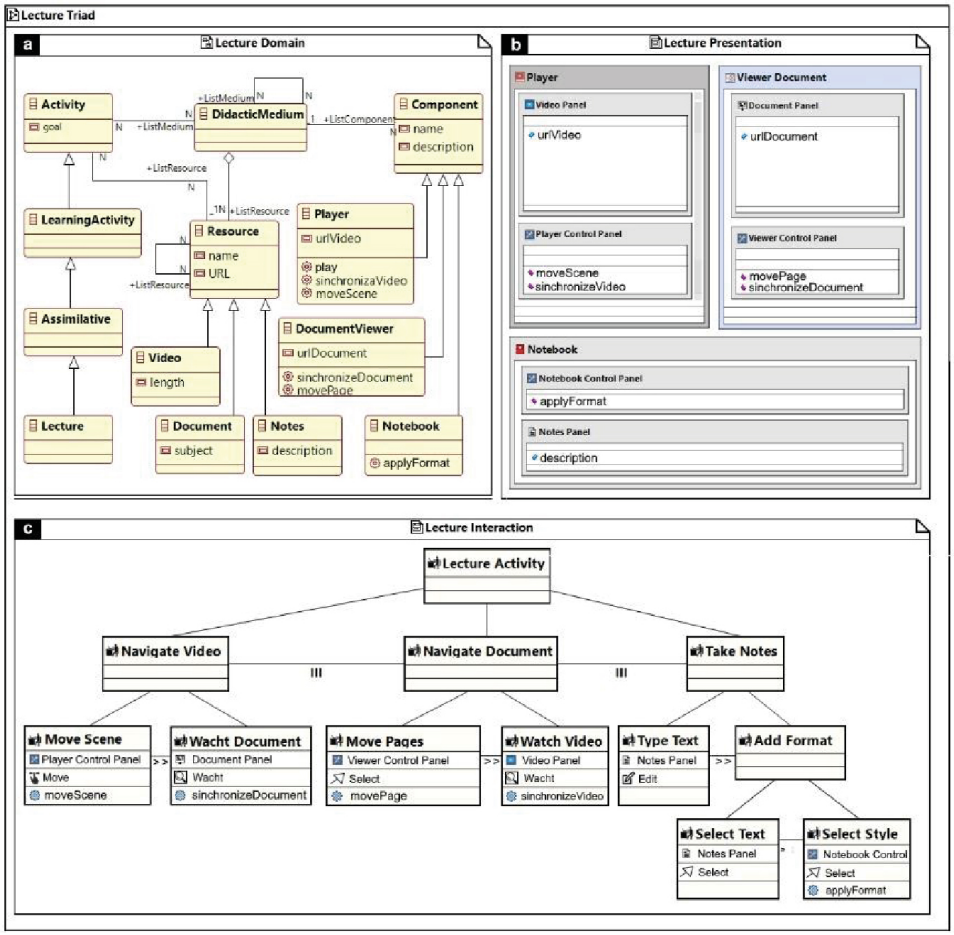


Fig. 4. *Triada a nivel de sistema refinada para la actividad Lecture*

Como se observa en la Fig. 4, el modelo de dominio que conforma la triada se describe con un diagrama de clases con la notación de UML (Fig. 4a); el modelo de presentación con un diagrama *DataForm* [13] (Fig. 4b) y el modelo de interacción está representado con un diagrama de dialogo [14], el cual describe las interacciones del usuario con los elementos de la pantalla (Fig. 4c). La diferencia entre la triada de la etapa anterior y esta etapa radica en que en esta etapa el diagrama de interacción define los controles que necesitará el usuario para llevar a cabo sus tareas, así como también, el modo de llevar a cabo las acciones sobre la interfaz (ej. mediante botones, menús, arrastrar y soltar, etc.) (Fig. 4c). Igualmente, los atributos de cada entidad y los aspectos de forma y apariencia (ej. Tamaños máximos y mínimos de widgets y/o componente) son refinados en el diagrama de clases y diagrama *dataform*, respectivamente.

Una vez refinado el catálogo de triadas a nivel de sistema, se procede a su implementación. Para ello, esta aproximación se apoya en metodologías existentes que permiten desarrollar la funcionalidad, como la que propone Giraldo [3]. Cada triada implementada

es un componente (con funcionalidad y variabilidad) que renderiza una partición de la interfaz de usuario y que puede ser utilizado para crear sistemas interactivos más robustos.

4.2 Flujo para crear herramientas

Para este flujo no se detallará el paso a paso de cada actividad que lo conforma. En vez de ello, se presentan las herramientas creadas como resultado de su ejecución en el contexto educativo.

Las herramientas que se crean en este flujo conforman el ambiente de desarrollo que hace posible que se puedan utilizar los componentes desarrollados a partir de las triadas para el desarrollo de sistemas interactivos. Tales herramientas son: i) editor de triadas y generador de código, b) editor DSML e c) interpretador DSML. A continuación se describen cada una de estas herramientas.

- *Editor de triadas:* el editor de la triada corresponde a la herramienta que da soporte al marco de desarrollo de CIAF, denominada Collaborative Interactive Application Tool – CIAT[3]. Dicha herramienta cuenta con diferentes editores de modelado a saber: editor del diagrama de clases (modelo de dominio), editor del diagrama de dialogo [14] (modelo de interacción) y editor del diagrama *DataForm* [13] (modelo de presentación). Para esta prueba de concepto en particular, las modificaciones a la herramienta CIAT se centraron en: (i) extender los elementos de modelado del editor del diagrama *Dataform* de tal forma que soportara los componentes y/o widgets que caracterizan la interfaz de usuario de los sistemas interactivos basados en distribución de contenido de video en el contexto educativo. (ii) crear un generador de código para HTML5. En el trabajo de Cruz [15], se expone el proceso de desarrollo y el resultado de dichas modificaciones.

- *Editor DSML:* permite crear modelos de estrategias didácticas utilizando la notación DSML [16]. DSML es un lenguaje visual orientado al flujo y centrado en la actividad que permite el diseño del aprendizaje a través de estrategias didácticas que responden a necesidades específicas de los estudiantes. El diagrama DSML es el que permite orquestar las actividades de aprendizaje que son instrumentadas a través de los componentes que se crean a partir de la triada, es decir es el diagrama que representa el modelo de orquestación de actividades para el dominio de la educación. A medida que se construye el diagrama de estrategia didáctica, cada elemento del diagrama se configura en un panel de propiedades. En el caso de las actividades de aprendizaje, además de configurar los diferentes atributos relacionados con la actividad, se configura la plantilla de presentación a utilizar. Las plantillas de presentación corresponden a los componentes que se desarrollaron con la triada y que instrumentan las diferentes actividades de aprendizaje. La selección de una variabilidad depende del tipo de actividad y los recursos que el profesor tenga disponibles para llevarla a cabo (videos, transcripción, documentos, etc.). La apariencia de cada plantilla puede ser personalizada a partir de la configuración de los parámetros de cada componente que la conforma. En el estudio de Ruiz [16], se presenta el proceso de diseño del lenguaje DSML y su respectivo editor.

- *Interpretador DSML:* El interpretador DSML es la herramienta encargada de la ejecución y gestión de los modelos de estrategias didácticas que se van creando. La herramienta está compuesta por dos módulos, un motor y un gestor de

modelos. El motor DSML es el encargado de desplegar (ejecutar) las actividades de aprendizaje en la vista del estudiante de acuerdo con el flujo esTable.cido en el modelo de estrategia didáctica. El gestor DSML es el encargado de administrar los modelos DSML (crear, modificar, borrar, asignar estudiantes, etc.) y reportar a través del *dashboard* información relacionada a la ejecución de las actividades de aprendizaje realizadas por los estudiantes.

En la Fig. 5 se detallan las tecnologías utilizadas para el desarrollo de los diferentes productos que se generan como resultado de la ejecución de ambos flujos.

A partir del despliegue de esta infraestructura es posible generar sistemas interactivos más robustos. Como por ejemplo, un curso online para una determinada asignatura. Para este caso, un desarrollador selecciona las triadas que se podrían utilizar en las diferentes actividades de aprendizaje del curso y genera los respectivos componentes. Para ello, utiliza el editor de la triada y el generador de código respectivamente. Posteriormente, el experto de dominio (en este caso un docente) realiza el o los diagramas de estrategia didáctica que guiarán el curso teniendo en cuenta los componentes disponibles generados previamente. Cada diagrama se crea de acuerdo con diferentes perfiles de los estudiantes. Finalmente, las actividades de aprendizaje son desplegadas a los estudiantes por medio del motor DSML, de acuerdo con el flujo esTable.cido en el diagrama DSML. El docente puede realizar el seguimiento de cada actividad a través del gestor DSML. Este tipo de aplicaciones pueden ser integradas con sistemas manejadores de aprendizaje (LMS) u otro tipo de aplicaciones, a partir de los cuales se pueda obtener la información de los estudiantes de una determinada cohorte

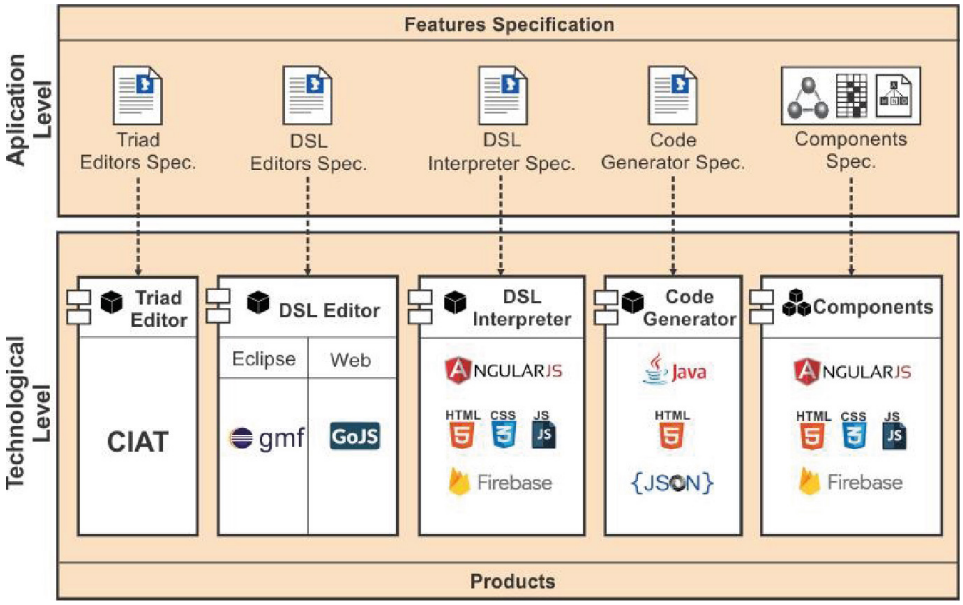


Fig. 5. *Tecnologías utilizadas para el desarrollo de productos generados en los flujos de la aproximación (flujo de componentes y flujo de herramientas)*

V. CONCLUSIONES Y TRABAJO FUTURO

En este artículo se ha presentado una aproximación metodológica para el desarrollo de sistemas interactivos basada en triadas de modelos. Este enfoque de ingeniería pretende minimizar la cantidad de modelos a partir de los cuales se pueden generar sistemas interactivos que promuevan la usabilidad en la interfaz final. Igualmente, por medio de este enfoque se busca que el desarrollo de los sistemas interactivos no parta desde cero, de tal manera que se promueva el prototipado rápido funcional y el producto mínimo viable. Una prueba de concepto en el dominio de la educación se llevó a cabo como parte del proceso de validación. La prueba de concepto evidenció que es posible generar sistemas interactivos a partir de las triadas y haciendo uso de las herramientas que conforman el ambiente de desarrollo. Como trabajo futuro se plantea configurar la aproximación metodológica sobre una herramienta de implementación de procesos de desarrollo en forma de producto, de tal forma que pueda ser consultada fácilmente por los desarrolladores. Igualmente, la ejecución de la aproximación en otros dominios de aplicación es planteada como foco de futuras investigaciones. Dichas investigaciones permitirían ratificar la efectividad del método para el desarrollo de cualquier sistema interactivo.

REFERENCIAS

[1] *W. J. Giraldo, "Marco de Desarrollo de Sistemas Groupware Interactivos Basado en la Integración de Procesos y Notaciones - CIAF," PhD Doctoral, Universidad de Castilla - La Mancha, 2010.*

[2] *M. L. Villegas, "Metodología para el Desarrollo de Sistemas Interactivos Usables CIA-F+HCI," Magíster en Ingeniería, Universidad EAFIT, 2012.*

[3] *OMG. (2008, 20-01-2019). OMG: Unified Modeling Language (UML) Version 2.0. Available: http://www.uml.org*

[4] *D. Moody, "The Physics of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering," IEEE Transactions on Software Engineering, vol. 35, pp. 756-779, 2009.*

[5] *S. Torre, "La comunicación didáctica: modelos y pautas para la acción," in Didáctica general para psicopedagogos / coord. por Núria Rajadell i Puiggròs, Félix Sepúlveda, , ed, 2001, pp. 103-154.*

[6] *J. L. Rodríguez, "Curriculum, acto didáctico y teoría del texto," Anaya, Ed., ed España, 1985.*

[7] *S. J. Gimeno, Teoría de la Enseñanza y desarrollo del curriculo. , Cuarta Edición ed. Buenos Aires. : Rei Argentina, 1992.*

[8] *G. Meneses, "NTIC, Interacción y Aprendizaje en la Universidad," Doctorado, Departament de Pedagogia, UNIVERSITAT ROVIRA I VIRGILI, 2007.*

[9] *P. Marqués. (2011). La Enseñanza Buenas Prácticas. La Motivación. Available: http://peremarques.net/*

[10] *G. Conole, "Describing learning activities: Tools and resources to guide practice," in Rethinking Pedagogy for a Digital Age: Designing and Delivering E-learning, H. Beetham and R. Sharpe, Eds., ed London and New York: Routledge, 2007.*

[11] *O. W. J. Giraldo, M. R. Arias, O. C. A. Collazos, A. I. Molina, C. M. Ortega, and M. A. Redondo, "Fast functional prototyping of user interfaces based on DataForm models, a tool (ToolDataForm)," in 2015 10th Computing Colombian Conference (10CCC), 2015, pp. 172-179.*

[12] *M. L. Villegas, "Marco conceptual, metodológico y tecnológico para el modelado y ejecución de la actividad que soporta el diseño de la interfaz de usuario de sistemas interactivos, dentro del contexto de TD_MBUID" Doctor en Ciencias de la Computación, Sistemas, Universidad del Cauca, 2017.*

Biografía Autor(es)

a. Doctora en Ingeniería Telemática. Docente, Uniquindio. E:mail: aruiz@uniquindio.edu.co. Quindio, Colombia.

b. Doctor en arquitectura y gestión de la información y del conocimiento en sistemas en red.

c. Doctorado en Ingeniería de Sistemas Telemáticos.

Comparación de controladores MPC y PID aplicados al proceso térmico de una máquina extrusora de poliestireno expandible

Javier Suarez Jaramillo (a)
Mateo Córdoba (b)
John Barco Jiménez (c)
Camilo Lagos (d)

Resumen

Este artículo muestra un estudio comparativo entre un Modelo de Control Predictivo (MPC) y un Control Proporcional Integral Derivado (PID), aplicado al control de temperatura en una máquina extrusora, su sistema térmico se basa en dos resistencias calefactoras (R1) y (R2), que se encargan de mantener los puntos de temperatura de referencia en una máquina extrusora de poliestireno expandible. Para analizar el comportamiento de aquellos sistemas de control, se implementan en la máquina, observando que ambos logran estabilizar la temperatura en el punto de referencia, pero el control MPC estabiliza la señal de temperatura y responde a perturbaciones en un tiempo menor que el control PID. Asimismo, el control MPC tiene la capacidad de prever las acciones futuras, permite minimizar su consumo de energía, reduciendo los costos monetarios en un 24%.

Palabras clave: Modelo de control predictivo MPC, controlador derivado integral proporcional PID, costos energéticos, consumo energético, temperatura.

Controller comparison MPC and PID applied to the process thermal from an extruder machine expandable polystyrene

Abstract

This paper shows a comparative study between a Model Predictive Control (MPC) and a Proportional Integral Derivative control (PID), applied to temperature control in an extruder machine, its thermic system is based in two heating resistors (R1) y (R2), which are responsible of maintain the reference temperature points in an Expandable Polystyrene extruder machine. For analyze the behavior of those control system, are implemented in the machine, observing that both achieve to stabilize the temperature in the reference point, but the MPC control stabilize the temperature signal and answer to disturbances in a few time than the PID control. Also, the MPC control has the capacity of foresee the future actions, it allows minimize its energy consume, reducing monetary costs in a 24%.

Keywords: Model Predictive Control MPC, Proportional Integral Derivative Controller PID, Energy Costs, Energy consumption, Temperature.

I. INTRODUCCIÓN

El poliestireno expandible EPS es un material polímero el que al ser compactado a base de vapor a altas presiones se convierte en un material comúnmente conocido como Icopor, que sirve como envase protector, laminado para construcción y arte. Este material una vez cumple su vida útil es desechado, causando problemas de contaminación en el medio ambiente.

La transformación de estos polímeros por medio de máquinas extrusoras de polímeros permiten reducir la contaminación generada por estos ya que disminuye su volumen, pero si este proceso no se lleva de forma adecuada el EPS puede quemarse conllevando a daños internos en la máquina, además genera un consumo de energía y costos energéticos mayores, por lo tanto un sistema de control apropiado para la máquina de extrusora es vital si se desea reutilizar los polímeros con un fin específico, como por ejemplo en la fabricación de postes en donde es necesario controlar la temperatura en el cañón de extrusión para que el producto sea resistente y homogéneo.

En este orden de ideas, en los procesos térmicos el análisis y diseño de controladores lineales ha sido práctica común durante mucho tiempo en procesos de temperatura, entre los cuales se pueden destacar los Proporcionales Integrales Derivativos (PID) y el Control Predictivo basado en el Modelo (MPC), en donde el PID ha mostrado buenos resultados en cuanto al tiempo de asentamiento, error y sobrepaso [1], además posee mejor tiempo de retardo y tiempo de subida que el control MPC [2]. El control PID tiene como ventaja sobre el control MPC, su bajo costo computacional, además, que solo es necesario resolver ecuaciones sencillas para su implementación [3]; en la práctica los ajustes de parámetros del control MPC son similares al PID, pero presentan mayor complejidad al realizar el ajuste [4].

Los sistemas de control PID son estudiados en diferentes procesos térmicos, donde, se optimiza y mejora el desempeño del lazo de control sin perder estabilidad, logrando demostrar que es posible mejorar la sintonización de un controlador PID realizando un procedimiento de optimización multivariable con las ecuaciones de la ganancia del proceso [5].

Varios autores han analizado el funcionamiento del control PID y MPC en diferentes procesos industriales, donde notaron la reducción de tiempo de esTable.cimiento y rapidez del sistema logrado por la implementación del control PID; además, observaron inestabilidad y oscilaciones de amplitud mínima en el régimen permanente [6]. En cuanto al consumo energético, investigaciones donde compararon el control PID y MPC no encontraron una diferencia significativa, siendo esta 0.04 % menor en el control MPC en comparación con el PID [7].

En el presente documento se desarrollaron dos estrategias de control capaces de optimizar el desempeño y mejorar la calidad de producción de una máquina extrusora tipo rosca de poliestireno expandible EPS, contribuyendo en la disminución de los desperdicios generados por este polímero, los cuales causan impactos ambientales negativos. Las estrategias de control diseñadas e implementadas son el control MPC y PID, que operan sobre las resistencias térmicas de la máquina extrusora; a partir de estos controladores se realiza el análisis del comportamiento de los principales parámetros del mismos, capacidad de respuesta ante perturbaciones y consumo de energía por parte de cada una de las resistencias térmicas.

El artículo está conformado por las siguientes secciones. En la sección 2. Materiales y métodos, se describe las características de la máquina extrusora, el modelo matemático del proceso térmico de extrusión y la descripción matemática del controlador MPC.

En la sección 3 Resultados experimentales, se presentan los escenarios de prueba del controlador operando sobre la máquina, las mediciones de desempeño de los controladores y el análisis del consumo energético. Finalmente, se presentan conclusiones acerca de la implementación y síntesis del consumo.

II. MATERIALES Y MÉTODO

2.1 Máquinas extrusoras

La extrusión es un proceso mecánico por medio del cual se transforma la materia prima mediante presión. Al utilizar esta técnica en polímeros se puede obtener diferentes productos, que pueden estar conformados de materia virgen y/o reciclada y que además pueden contener diferentes adherentes. El proceso se realiza forzando un polímero fundido a pasar a través de una boquilla, lo que requiere de una extrusora, donde su estructura básica varía según el tipo de material a trabajar; la extrusora tipo tornillo es la más conocida debido a su sencillez y costo, permitiendo que los gránulos ingresen por la tolva y por medio de temperatura y velocidad en el tornillo se puedan fundir [8].

El funcionamiento de las máquinas extrusoras se basa principalmente en forzar material polímero a pasar a través de un boquete o dado [9], capaz de fusionar el material a partir de los valores de temperatura y velocidad, que conllevan a homogenizar el polímero para la obtención de diferentes productos terminados.

Descripción de la máquina extrusora

La máquina extrusora que se desarrolló para este proyecto fue una extrusora tipo rosca, la cual cuenta con los siguientes elementos, como se observa en la tabla 1 y figura 1.

Table. 1. *Elementos de la máquina extrusora.*

Can	*Elemento*	*Características*	*Peso*
1	*Motoreductor*	*8 HP a 1500 RPM con relación 5:1*	*116 Kg*
2	*2 Resistencias*	*719.28 Watts*	*7 Kg*
1	*1 Cañón*	*Acero Ø= 4" x 1/4. 120 cm*	*42 Kg*
1	*1 Husillo*	*Acero Ø= 3.75". 130 cm*	*50 Kg*
1	*1 Dado extrusión*	*Acero Ø= 3.25"*	*9 Kg*
1	*1 Tolva dosificadora*	*Lamina de acero en 1/8"*	*35 Kg*
1	*1 Estructura base*	*Acero 1 1/2" x 3/16*	*65 Kg*

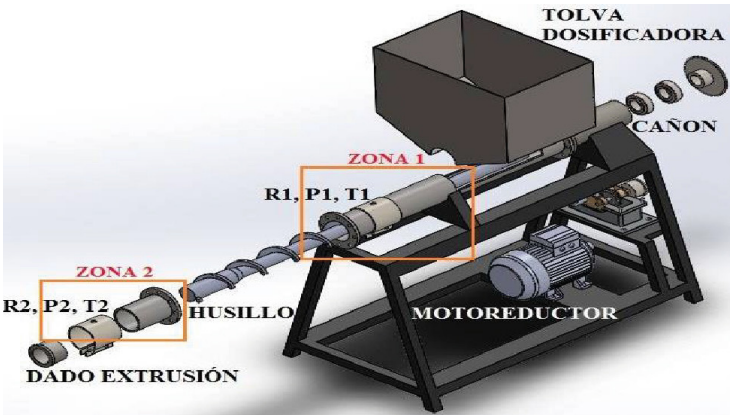


Fig. 1. *Ensamble de la maquina extrusora.*

2.2 Modelamiento del proceso térmico de extrusión

Un modelo matemático de un sistema dinámico se puede definir como un conjunto de ecuaciones que representa una aproximación dinámica del sistema. Dado que los sistemas se pueden representar de varios modos diferentes, un modelo matemático no es único para un sistema dado; su representación depende de los objetivos y perspectivas individuales del diseñador del sistema de control [10]. Para modelar el sistema se debe definir el punto de referencia, ya que realizando la diferencia de este con el valor de salida del sistema se encuentra el error presente de temperatura que se encuentra en la máquina.

Definición del punto de referencia

El punto de referencia de temperatura se define de forma experimental, variando a diferentes temperaturas las dos resistencias con las cuales cuenta la máquina extrusora hasta obtener la homogeneidad deseada del poliestireno expandible extruido; el análisis se hizo por medio de los pirómetros EBCHQ 58006 conectados a dos termopares tipo K, variando la temperatura desde 18°C hasta 260°C, el pirómetro informa la temperatura de las resistencias calefactoras en resoluciones de 1°C; definiendo así los puntos de referencia correctos en temperatura para cada una de las resistencias que permiten homogenizar el EPS, ver ecuación 1.

$$R1=205^{o}C,\ R2=210^{o}C \quad (1)$$

Funciones de transferencia de la máquina

Para esTable.cer cada una de las funciones de transferencia, se utilizó un sistema embebido mediante el cual se obtuvo los valores de temperatura y potencia para luego ser procesados en Matlab. En primera instancia se adquirió los datos de potencia P1 que afecta directamente a la temperatura de T1 (°C) de la resistencia (R1) ubicada en la zona 1 e indirectamente afecta a la temperatura T2 (°C) de la resistencia (R2) situada en la zona 2. Posteriormente se encontró los datos de potencia P2 que afecta directamente a la temperatura T2 (°C) de la resistencia (R2) en la zona 2 e indirectamente a la temperatura T1 (°C) de la resistencia (R1) ubicada en la zona 1, (ver figura 2).

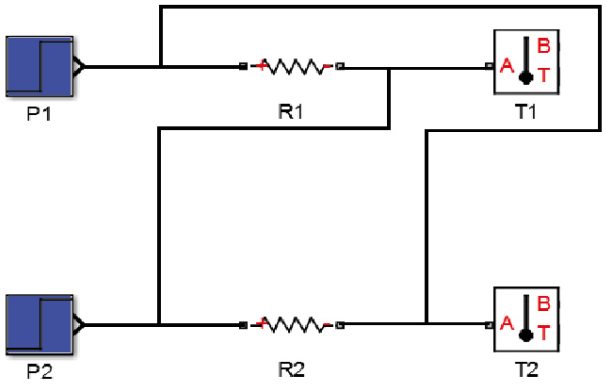


Fig. 2. *Esquema de la potencia, resistencias y temperatura*

Una vez esTable.cidos los datos se utilizó el Toolbox IDENT de Matlab, encontrando los modelos que representan la máquina. En la ecuación 2, 3, 4 y 5; se indican las funciones de transferencia de P1 respecto a R1, P1 respecto a R2, P2 respecto a R2 y P2 respecto a R1.

$$G_{11}(S) = \frac{0.0005024}{s + 0.001471} \tag{2}$$

$$G_{12}(S) = \frac{5.582e - 5}{s + 7.126e - 5} \tag{3}$$

$$G_{21}(S) = \frac{0.0004505}{s + 0.001} \tag{4}$$

$$G_{22}(S) = \frac{6.132e - 5}{s + 9.206e - 5} \tag{5}$$

2.3 MPC

El control predictivo del modelo es un esquema de control en la que la acción que realiza el control actual se obtiene resolviendo, en cada muestreo instantáneo, un problema de control óptimo de ciclo abierto en el horizonte finito, usando el estado actual de la planta como estado inicial; la optimización produce una secuencia de control óptima y el primer control en esta secuencia se aplica a la planta. Una ventaja importante de este tipo de control es su capacidad para hacer frente a restricciones duras en controles y estados [11]. El modelo de predicción se calcula a partir de la siguiente expresión:

$$\hat{y}(t+1),\ \hat{y}(t+2|t),\ \ldots\ ,\hat{y}(t+N|t) \tag{6}$$

En la cual: t, Tiempo continuo, $\hat{y}$, Salida estimada, N, Horizonte de predicción.

El cálculo de la anterior secuencia se basa en la información de la evolución del proceso y que hace referencia a: valores del pasado en las entradas o actuaciones del proceso, valores presentes y pasados de las salidas del proceso.

Configuración MPC

El sistema de planificación del controlador MPC se puede representar a partir de $r(k)$ como el punto de referencia de temperatura de las resistencias, $y(k)$ es la salida (Temperatura °C), $e(k)$ es la diferencia entre $r(k)$ y $y(k)$, por otra parte, $d(k)$ es la perturbación que producen las resistencias en forma de temperatura °C y $u(k)$ es la acción de control generada por el controlador MPC, en este caso Potencia (Watts); donde k es el tiempo discreto.

La ecuación que representa la salida del sistema es:

$$y\,(k+1) = f\,(x(k),\ u(k),\ d(k)) \tag{7}$$

El error se define de la siguiente manera:

$$e(k+1) = r(k+1) - y(k+1) \tag{8}$$

El controlador MPC reduce el error de seguimiento en todo el tiempo, tanto como sea posible, de tal manera que la salida sea idéntica a la referencia; para lograr lo anterior se considera los posibles planes o señal de control:

$$u^i[k,N] = u^i(k,0),\ u^i(k,1),\ \ldots\ ,u^i(k,N-1) \tag{9}$$

Donde: u^i, es la acción de control respecto al tiempo k en la i-ésima acción de control. Dichos planes corresponden a acciones de control que se generan por un conjunto de controladores que se configuran como indica la siguiente sección.

Generación de acciones de control

Se generan mediante un conjunto de controladores PID, como se indica en la siguiente ecuación.

$$u(k) = k_p e(k) + k_i \Sigma_{j=0}^{k} e(j) + k_d(e(k) - e(k-1)) \tag{10}$$

Los controladores se generan utilizando todas las posibles combinaciones entre k_p, k_i y k_d, obteniendo 55 acciones de control; las cuales presentaron el mejor resultado posible en cuanto a la simulación.

Criterio para la selección del plan

Es necesario tener un criterio para encontrar el objetivo del controlador. La acción de control se escoge mediante un criterio de selección [1], como se indica en la siguiente ecuación; en este caso, se utiliza una función $J(u^i[k,N])$ que cuantifica la calidad de cada plan $(u^i[k,n])$ en función del error de seguimiento y el consumo de energía (Watts), ver ecuación 11.

$$J(u^i[k,N]) = w_1 \Sigma_{j=1}^{N} r(k+j) - y^i(k,j))^2 + w_2 \Sigma_{j=0}^{N-1} (u^i(k,j))^2 \tag{11}$$

Donde la primera sumatoria representa el error de seguimiento, mientras tanto la segunda sumatoria es la energía utilizada por la acción de control, que en este caso es la potencia consumida por las resistencias calefactoras; además, se tienen los coeficientes $w1 \in [0,1] y w2 \in [0,1]$ que acompañan a las sumatorias, los cuales permiten ponderar la importancia de cada término.

Para obtener el mejor controlador se utiliza la ecuación 12, que permite seleccionar el controlador que minimiza la función costo J. El controlador óptimo (i^*) y la acción de control se aplica al proceso real en el instante k, ver ecuación 13.

$$\frac{argmin}{i^*} J(u^i[k;N]) \quad (12)$$

$$u(k) = u^{i^*}(k,0) \quad (13)$$

La ecuación 12 corresponde a la acción de control del horizonte de proyección. De tal manera que el proceso se repite en cada instante de tiempo k para generar una acción de control que regule la potencia en el proceso de temperatura, minimizando el consumo de energía y cumpliendo el objetivo de seguimiento.

Diseño del controlador MPC y PID

El MPC es un controlador basado en los parámetros del control PID, donde se tiene en cuenta k_p, k_i, k_d, pero la salida $y(t)$ se obtiene muestreando los instantes futuros. La figura 3 se presenta el controlador MPC y PID para R1 y R2 de la máquina extrusora, en diagrama de bloques.

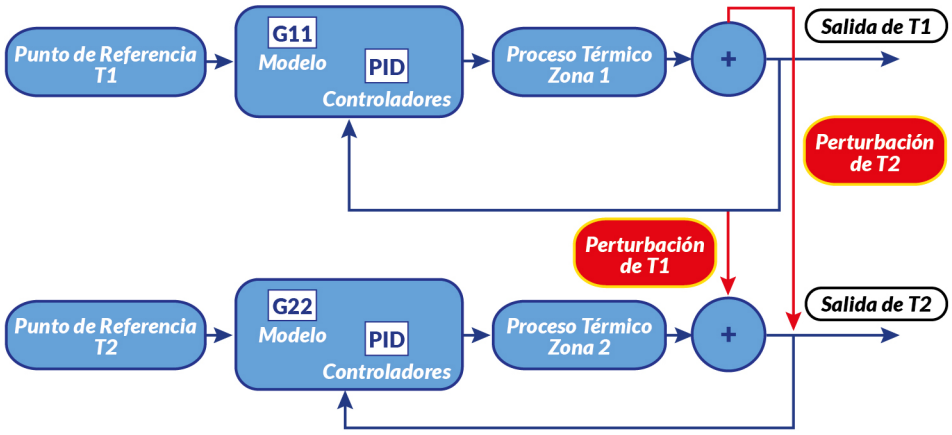


Fig. 3. *Diagrama de bloques para el diseño del control MPC y PID.*

Para la simulación del controlador MPC y PID, se tiene en cuenta cómo afecta cada potencia a su respectiva resistencia y como perturba a la resistencia contraria. La simulación de este controlador se llevó a cabo en un Script de Matlab, permitiendo esTable. cer el rango de los parámetros del controlador MPC y los valores de los parámetros del control PID para R1 y R2, (ver tabla 2 y 3).

Table. 2. *Rango de los parámetros del controlador MPC y PID para P1 respecto T1 de R1, considerando perturbación de P2*

Parámetro del Controlador	Rango del Parámetro (MPC)	Parámetro (PID)
Kp	*[1.5:0.1:2.5]*	*2*
Ki	*[0.004:0.001:0.008]*	*0.004*
Kd	*[300]*	*0*

Table. 3. *Rango de los parámetros del controlador MPC y PID para P2 respecto T2 de R2, considerando perturbación de P1*

Parámetro del Controlador	Rango del Parámetro (MPC)	Parámetro (PID)
Kp	*[2.5:0.1:3.5]*	*3*
Ki	*[0.001:0.001:0.004]*	*0.0035*
Kd	*[300]*	*0*

Hardware controlador MPC y PID

Para realizar la implementación del controlador PID se utiliza los componentes presentes en la tabla 4, omitiendo el uso de la tarjeta LattePanda con capacidad de procesamiento de alto rendimiento, esta tarjeta es utilizada para la implementación del control MPC, debido a la complejidad del Algoritmo desarrollado en Matlab; las características y componentes usados en la implementación de ambos controladores se presenta en la tabla 3.

Table. 4. *Características de los componentes para la implementación del control PID y MPC*

Componente	Características
LattePanda	*Capacidad de procesamiento a 64 Bit por 1.98 GHz con memoria de 2 G*
Arduino AT MEGA 2560	*Capacidad de procesamiento a 8 Bit por 16 MHz con memoria de 2 K*
Controladores de Potencia con Triac	*Relés de estado sólido (corriente máxima de switcheo de 5 AMP)*
Drivers MAX 6675	*Acondiciona la señal del termopar para el Arduino*
Termopares tipo K	*Capacidad industrial de -75 °C a 400 °C*

III. RESULTADOS EXPERIMENTALES

Se construyeron dos escenarios, en los cuales se aplica el control MPC y PID en la máquina extrusora de EPS, para todos los escenarios la máquina cuenta con material en el cañón, ver tabla 5.

Table. 5. *Características de los escenarios de implementación del control MPC y PID*

	MEPS	MA	CA
Escenario 1			*X*
Escenario 2	*X*	*X*	*X*

Donde:

MEPS: Máquina sin material EPS en tolva (perturbación).
MA: Motor activo (perturbación).
CA: Cargas de la fábrica, industrias y hogares aledaños activas (perturbación).
X: representa las variables que afectan a los escenarios propuestos.

La temperatura ambiente para los escenarios propuestos oscila entre los 14 y 17 °C.

3.1 Comparación de desempeño de los controladores

Este estudio se realizó para evaluar el desempeño de los controladores sobre la máquina extrusora, para ello se consideró los parámetros de tiempo de levantamiento, tiempo de asentamiento, sobrepaso máximo y error, el consumo de energía y la capacidad de respuesta ante perturbaciones presentes en el proceso de extrusión. Para hacer el cálculo del consumo de energía se utilizó el método de integración trapezoidal, representado en la ecuación número 14 Y 15.

$$E = \frac{B}{N}\left[\frac{P1 + P2}{2}\right] \tag{14}$$

$$B = N \cdot Tm \cdot Tp \tag{15}$$

Dónde:

N: Longitud del vector de datos de potencia; *P1:* Dato de potencia cada tiempo de muestreo; *P2:* Suma de datos de potencia; *Tm:* Tiempo de muestreo; *Tp:* Tiempo de prueba

Análisis en base al primer escenario.

A partir de las figuras 4, 5, 6 y 7 en las cuales se identifica el comportamiento de la temperatura y la potencia obtenida a partir de la acción de control para el escenario uno de las resistencias R1 y R2 durante una hora de prueba, con la implementación del control PID y MPC respectivamente, se obtuvo la tabla de comparación 6, que permitió determinar el comportamiento de cada controlador en iguales condiciones de trabajo.

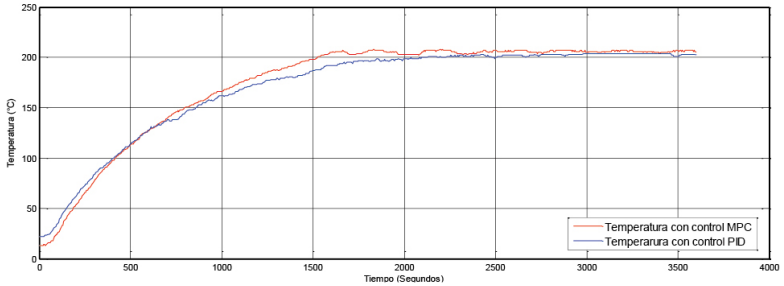


Fig. 4. *Comportamiento de la temperatura de R1 en el escenario uno con control PID y MPC*

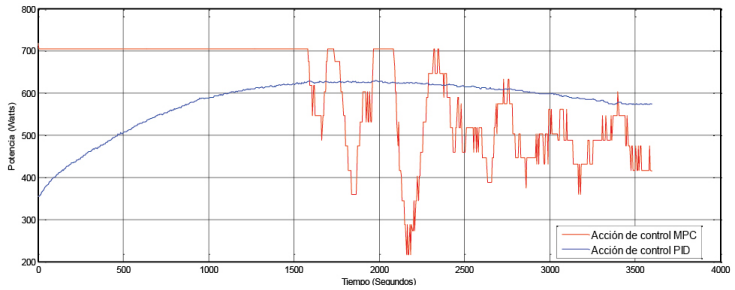


Fig. 5. *Potencia obtenida a partir de la acción de control P1 de R1 en el escenario uno con control PID y MPC*

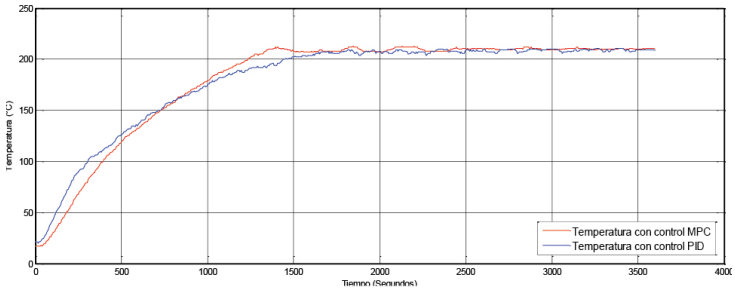


Fig. 6. *Comportamiento de la temperatura de R2 en el escenario uno con control PID y MPC*

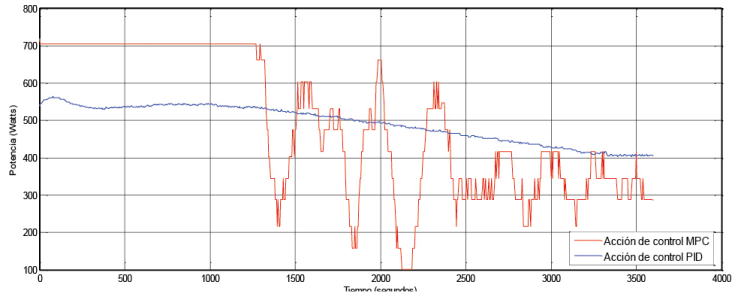


Fig. 7. *Potencia obtenida a partir de la acción de control P2 de R2 en el escenario uno con control PID y MPC*

Table. 6. *Comparación control PID y MPC en el escenario 1*

	Control PID		Control MPC	
	R1	R2	R1	R2
ts	2133 Seg	1617 Seg	1477 Seg	1266 Seg
tr	1477 Seg	1232 Seg	1205 Seg	1042 Seg
sm	0%	0%	0%	0%
e	(-1°C)	(-2°C)	1.6°C	1.1°C

Análisis en base al segundo escenario

Las figuras 8 y 10 representan el comportamiento de la temperatura en R1 y R2, durante una hora de prueba, las figuras 9 y 11 muestran la acción de control correspondiente a cada resistencia. El comportamiento descrito en estas gráficas se puede ver en la tabla 7.

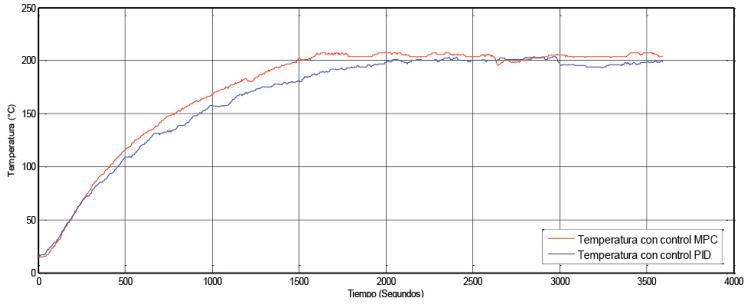


Fig. 8. *Comportamiento de la temperatura de R1 en el escenario dos con control PID y MPC*

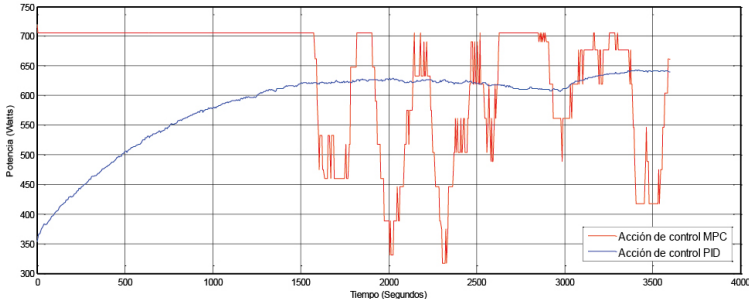


Fig. 9. *Potencia obtenida a partir de la acción de control P1 de R1 en el escenario dos con control PID y MPC*

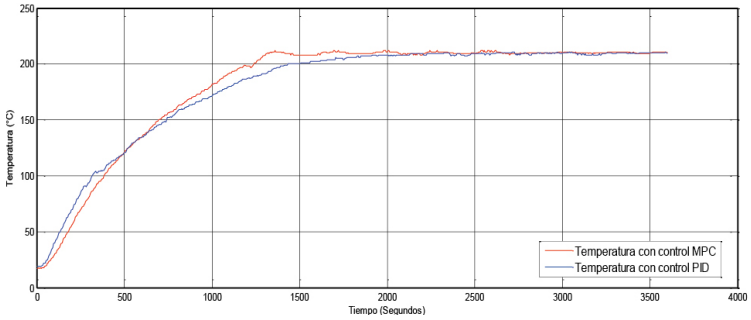


Fig. 10. *Comportamiento de la temperatura de R2 en el escenario dos con control PID y MPC*

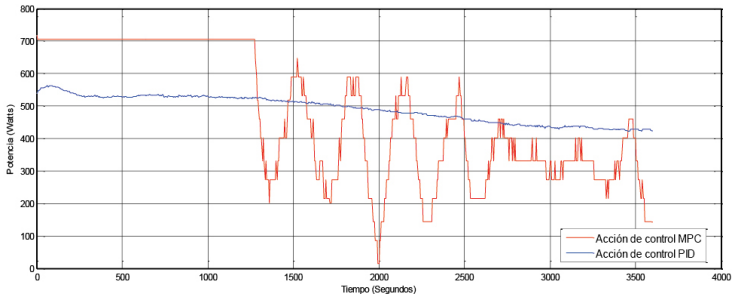


Fig. 11. *Potencia obtenida a partir de la acción de control P2 de R2 en el escenario dos con control PID y MPC*

Table. 7. *Comparación control PID y MPC en el escenario 2*

	Control PID		Control MPC	
	R1	R2	R1	R2
Ts	2187 Seg	1771 Seg	1495 Seg	1280 Seg
Tr	1483 Seg	1227 Seg	1260 Seg	1070 Seg
Sm	0%	0%	0.975%	0.975%
E	(-1°C)	(-1°C)	1°C	1°C

3.2 Análisis del consumo de energía de los controladores PID y MPC

Para este análisis se identificó dos fases principales, la primera hace alusión a la acción de control hasta el momento en el que llega al punto de referencia y la segunda fase representa a la acción de control en operación, es decir, cuando alcanza el punto de referencia.

Consumo de energía hasta alcanzar el punto de referencia El consumo de energía de las resistencias calefactoras R1 y R2 hasta lograr el punto de referencia con controlador PID y MPC se obtiene de las figuras 6, 10 para R1; 8, 12 para R2 en los dos escenarios con control PID y MPC. La tabla 8, muestra el consumo de energía de R1 y R2 con control PID y MPC a partir del método de integración trapezoidal, hasta alcanzar el punto de referencia, T1 = 205 °C de R1 y T2 = 210 °C de R2.

Table. 8. *Consumo de energía de R1 y R2 con control PID y MPC, hasta alcanzar punto de referencia*

	Escenario 1 y 2			
	Control PID		Control MPC	
	R1	R2	R1	R2
Promedio de energía consumida hasta alcanzar punto de referencia	568.2 W ts	543.9 W ts	704.6 W ts	705.1 W ts
ts (Tiempo de asentamiento promedio)	2153.7 Seg	1656.6 Seg	1488 Seg	1265.2 Seg

La potencia empleada para lograr el punto de referencia con la implementación del controlador MPC sobre R1 y R2 es mayor en promedio en un 24 % para todos los escenarios para R1 y 29.6 % para R2 en comparación a la potencia suministrada con la implementación del control PID, pero el tiempo empleado por este último es mayor, lo que hace que la energía total consumida también lo sea. A pesar que la potencia suministrada por el controlador MPC es mayor para todos los escenarios propuestos, el control PID tarda un tiempo promedio mayor de 44.7 % en R1 y 30.9 % en R2 respecto al tiempo que tarda el control MPC en lograr el punto de referencia.

Análisis del consumo de energía en el punto de referencia de R1 y R2 con control PID y MPC

Para calcular el consumo de energía en el punto de referencia de R1 y R2, se considera el tiempo restante hasta alcanzar una hora de prueba, después de llegar al tiempo de estabilización; esto se realiza para los dos escenarios con la implementación de los controladores PID y MPC en base a las gráficas 6, 10 para R1; 8, 12 para R2. A continuación se muestra el consumo de energía por parte de R1 y R2 después de alcanzar el tiempo de estabilización, trabajo esTable. (ver tabla 9).

Table. 9. *Consumo de energía por parte de R1 y R2 en trabajo esTable. con control PID y MPC*

	Escenario 1 y 2			
	Control PID		***Control MPC***	
	R1	***R2***	***R1***	***R2***
Energía consumida después de alcanzar el tiempo de estabilización	*616.9 W TR*	*487.5 W TR*	*552.2 W TR*	*369 W TR*
Tiempo restante de prueba (TR)	*1499.7 Seg*	*1968.2 Seg*	*2099.5 Seg*	*2334.7 Seg*

Se observa que R1 y R2, consumen mayor energía a partir de la implementación del controlador PID, durante el tiempo de operación o trabajo esTable., dado que el controlador PID debe generar una acción de control mayor a la del controlador MPC para mantener esTable. la temperatura T1 = 205 °C de R1 y T2 = 210 °C de R2, por otra parte, el controlador MPC requiere de menor energía que el controlador PID, 10.5% menos en R1 y 21.3% menos en R2 para mantener el punto de referencia de las resistencias calefactoras, debido que cuando llega la temperatura al punto de referencia genera una acción de control máxima para luego mantener la temperatura en la referencia definida con menor potencia.

IV. CONCLUSIONES

El controlador MPC, llega más rápido al punto de referencia con un tiempo de asentamiento menor en 44.7 % y 30. 9 % en R1 y R2 respecto al control PID, debido a que suministra la potencia máxima posible (719.28 Watts), mientras que el controlador PID genera una acción de control menor, debido a que no puede prever el estado futuro, absteniéndose de suministrar energía que genere sobrepaso.

Analizando la curva de energía total consumida de los controladores PID y MPC para todos los escenarios propuestos en el periodo de operación de la máquina, se puede observar que a pesar de que el controlador MPC suministra mayor potencia hasta lograr el punto de referencia la energía consumida es menor ya que se logra en un menor tiempo. De igual forma el consumo en operación normal es menor que el control PID, donde el consumo de energía en el control PID es mayor en 11.6 % (R1) y 27.8 % en (R2).

El controlador MPC presenta mejor desempeño que el controlador PID, debido a que presenta un tiempo de levantamiento promedio menor del 18.15 % en R1 y 13.5 % en R2, respecto al control PID; el tiempo de asentamiento promedio es 30.8% en R1 y 23.5 % en R2 en comparación al control PID; el error y el sobrepaso máximo no tiene variaciones considerables entre los controladores, por lo cual no se los considera, debido a que el material EPS extruido no es afectado por estas dos variables.

V. AGRADECIMIENTOS

Los autores agradecen a la Universidad Cesmag por brindar el apoyo suficiente en cuanto a conocimiento transmitido por sus docentes, herramientas y demás que permitieron llevar a cabo el desarrollo de los sistemas de control; además agradecen a la empresa ICOTALLA S.A.S por permitir realizar pruebas de implementación de cada uno de los controladores sobre la máquina extrusora tipo rosca de EPS.

REFERENCIAS

[1] *H. Arroyo, Comparación de diferencias técnicas de control aplicadas al control de temperatura en tiempo real, Revista Colombiana de Tecnologías de Avanzada, 18, 94 - 101, 2011.*

[2] *H. Arroyo, Estudio del comportamiento de un control MPC [Control predictivo basado en el modelo] comparado con un control PID en una planta de temperatura, Facultad de Ingeniería, 37, 45 - 54, 2014.*

[3] *M. Murillo, Control longitudinal para vuelo autónomo de aviones basado en el algoritmo Proporcional - Integral - Derivativo, Asociación Argentina de Mecánica Computacional, 32, 2655 - 2664, 2013.*

[4] *E. Lopera, Implementación de un control predictivo basado en modelo aplicado a un sistema de control de caudal de agua didáctico, Tecno Lógicas, 25, 109 - 130, 2010.*

[5] *J. Duarte, Optimización de sitinización de controladores PID bajo el criterio IAE aplicado a procesos térmicos, inge@UAN, 10, 35 - 45, 2015.*

[6] *J. Carlos, Diseño e implementación de un controlador PID y un MPC aplicado al posicionamiento lineal de un sistema neumático en bucle cerrado, Politécnica Salesiana , Cuenca - Ecuador, 2015.*

[7] *N. Mohammad, Evaluation on energy consumption in compact hydro distillation process between MPC and PID control, IEEE Control and system graduate research colloquium, 8, 116 - 121, 2017.*

[8] *C. Garces, Diseño del sistema de control para una máquina extrusora de polietileno de alta y baja densidad para fabricar peliculas de polietileno, para la empresa INCOLPA LTDA, Universidad Autonoma de Occidente, Cali, 2015.*

[9] *S. Romero, Proyecto de máquina extrusora para creación de hilos de polietileno teraflalato, Instituto Politecnico Nacional, Mexico DF, 2015.*

[10] *L. Ljung, System Identification Theory for the user, New Jersey: Prentice Hall, 1987.*

[11] *D. Mayne, Constrained model predictive control: Stability and optimality, Elsevier Automatica, 36, 789 - 814, 2000.*

[12] *K. Passino, Biomimicry for optimization, control, and Automation, London: Springer , 2005.*

Biografía Autor(es)

a. Ingeniero Electrónico. Docente, Instituto Técnico y Colegio de Bachillerato COLSUP. E-mail: sujajaal@gmail.com. Pasto, Colombia.

b. Magíster (c) en Diseño Mecánico con Énfasis en Producción Industrial. Jefe de mantenimiento de la Empresa ICOTALLA S.A.S. E-mail: matcr06@gmail.com.

c. Magister en ingeniería. Docente tiempo completo Facultad de Ingeniería, Universidad Cesmag. E-mail: jebarco@unicesmag.edu.co. Pasto, Colombia.

d. Magister en ingeniería. Docente tiempo completo Facultad de Ingeniería, Universidad Cesmag. E-mail: calagos@unicesmag.edu.co. Pasto, Colombia.

Creación de un algoritmo de cifrado simétrico denominado SARCH

Arturo Eraso Torres (a)
Christian Camilo Angulo Mirama (b)
Santiago Alvarez Alvarez (c)

Resumen

Este proyecto pretende fortalecer de manera significativa los avances tecnológicos para la protección lógica de información digital, también se puede afirmar que el estudio sirve como material base para la creación de futuras investigaciones relacionadas con el tema de algoritmos para el cifrado.

Palabras clave: algoritmo de cifrado simétrico, criptograma, criptoanálisis, cifrado, descifrado, texto plano, clave, llave, criptografía

Creation of an algorithm symmetric encryption named SARCH

Abstract

This project aims to significantly strengthen technological advances for the logical protection of digital information, it can also be said that the study serves as a base material for the creation of future research related to the topic of encryption algorithms.

Key Words: symmetric encryption algorithm, cryptogram, cryptanalysis, encryption, decryption, plain text, key, key, cryptography

I. INTRODUCCIÓN

En la actualidad la información digital que manejan las empresas y las personas es el recurso más importante puesto que es el activo con mayor valor que permite a la empresa tener un alto nivel de competitividad, parte fundamental para el desarrollo, también es necesario manifestar que la forma de tratarse es donde se debe tener mucho cuidado ya que puede ser robada o alterada por un usuario no autorizado o ciber-delincuentes consideradas personas con alto grado de conocimiento informático que utilizan estos conocimientos para realizar delitos.

Considerando lo anterior, en el presente proyecto se diseñó y creo un algoritmo de cifrado simétrico llamado SARCH, cuya denominación deviene de las iniciales de los nombres de los autores: Santiago, Arturo y Christian, el cual busca hacer ilegible la información a terceros, por medio del uso de una llave privada para el acceso, donde solo la persona que posee la llave puede conocerla.

De hecho, la creación del algoritmo esta apoyada en conceptos matemáticos que permiten diseñar el modelo matemático, puesto que la creación de un buen algoritmo está basada en conceptos como: La teoría de números, estadística y teoría de la información. La utilización de estos conocimientos es importante para que el algoritmo de cifrado simétrico alcance un alto nivel de robustez.

Si bien es cierto que alcanzar un 100% en seguridad informática es imposible, con el desarrollo de esta propuesta se buscó que el algoritmo sea capaz de brindar un alto nivel de seguridad y confiabilidad. Para ello, en este trabajo se realizaron pruebas para evidenciar que el algoritmo es resistente a los diferentes ataques como la técnica de criptoanálisis lineal, técnica que permite llegar a deducir una clave.

II. METODOLOGÍA

El estudio se abordó desde un paradigma positivista, un enfoque cuantitativo, un método científico, y un tipo de investigación de experimentos puros o verdades, Al tratarse de una investigación no convencional, la población está conformada por técnicas de criptoanálisis lineal que permiten evaluar el comportamiento del algoritmo frente a los diferentes ataques que se efectúan dentro un laboratorio de pruebas, la muestra para esta investigación fue determinista la cual se caracteriza porque el tamaño de la muestra y la selección de los elementos que la integran, están sujetos al juicio de los investigadores; de acuerdo con lo anterior se llegó a determinar qué el grupo estará conformado por algunos ataques contenidos dentro de las técnicas de criptoanálisis lineal que permiten medir la fiabilidad del algoritmo.

A. Problema a resolver

En la actualidad el manejo de información a través de internet se ha visto comprometido por fallos de seguridad que afectan directamente la privacidad de los usuarios y empresas, poniendo en riesgo en muchas ocasiones información de alta confidencialidad.

Principalmente el problema surge cuando la comunicación se realiza sin tener en cuenta la posibilidad de que puede ser interceptada por terceros. Por tal razón, si un mensaje fue capturado en el camino este podrá ser leído, tatrarrestar este problema se debe utilizar una herramienta que permita cifrar el mensaje antes de ser enviado.

B. Perspectiva del producto

La creación de algoritmo de cifrado simétrico SARCH contribuye a mejorar la protección de la información digital de la comunidad en general, todo esto en un aplicativo fácil de usar en donde se brinda una alternativa para la protección de la información digital de los usuarios.

El usuario que haga uso de este aplicativo para realizar el proceso cifrado de un texto plano, será el algoritmo SARCH el encargado de realizar los diferentes procesos internos siguiendo el modelo matemático determinado para realizar de manera óptima y segura las diferentes operaciones que le permitan al usuario final tener un texto cifrado ilegible a terceros que no conozcan la clave con la que fue realizado el cifrado.

De manera similar para descifrar un texto que ha sido cifrado previamente utilizando el algoritmo de cifrado SARCH es necesario que agregue el texto en el campo correspondiente, del mismo modo debe llenar el campo de contraseña con la clave que fue utilizada para realizar el cifrado anteriormente, una vez realizado este procedimiento se debe pulsar sobre la opción descifrar. A continuación, nuevamente el encargado de realizar el descifrado es el algoritmo SARCH mediante sus diferentes operaciones internas y siguiendo el modelo matemático que hace posible construir un texto plano claro original entendible por el usuario.

Table. 1. *Funciones del producto*

Cifrar.	*Generar un texto cifrado.*
Generador de claves.	*Crear una clave automática.*
Descifrar.	*Descifrar la información.*
Descargar archivo.	*Descargar en formato de texto el archivo cifrado o descifrado.*

Fuente: *Esta investigación*

C. Identificación de características del algoritmo de cifrado

Para poder llevar a cabo el diseño del algoritmo de cifrado simétrico, es necesario conocer de manera más cercana algunos algoritmos que en la actualidad son usados con mayor frecuencia, para lo cual es necesario realizar un análisis previo de sus características más relevantes, tales como: longitud de clave, tipo de cifrado, tamaño de bloque,

desarrollo, resistencia criptoanálisis, posibles claves, ronda, Velocidad de cifrado de algoritmos en un PC, clave. Para tener mayor claridad opta por recolectar información la cual se expone en la siguiente tabla (Tabla 2)

Table. 2. *Principales características de los algoritmos simétricos de cifrado*

Características	AES	Pts	EDES	Pts	DES	Pts	TWOFISH	Pts	BLOWFISH	Pts
Longitud de clave	128,192 y 256 Bits	5	112 y 168 bits	3	56bits	2	128 puede llegar 256 bits	4	32 bits hasta 448 bits clave variable	5
Tipo de cifrado	Simétrica bloques de cifrado	5	Simétrica bloques de cifrado	5	Simétrica bloques de cifrado	5	Simétrica bloques de cifrado	5	Simétrica bloques de cifrado	5
Tamaño de bloque	128,192.0 256 Bits	5	64 bits	2	64 bits	2	128 bits	3	64 bits	2
Desarrollo	2006	5	1978	4	1977	1	1996	3	1993	2
Resistencia criptoanálisis	Diferencial en contra de fuente, truncado diferencial e interpolación lineal y plazas de ataques	4	Vulnerable al diferencial de fuerza bruta atacante podría ser analizada de texto plano con criptoanálisis diferencial	1	Vulnerable al diferencial y lineal criptoanálisis y tablas de sustitución	1	Vulnerable al diferencial	3	Vulnerable al diferencial	3
Posibles claves	2^{128} y 2^{192}	5	2^{112} y 2^{168}	4	2^{57}	3	2^{57}	3	2^{32}	2
Ronda	10 (128-Bist), 12 (192 Bits), 14 (256 Bits)	5	48	3	16	2	Variable	4	16	2
Velocidad de cifra de algoritmos en un PC	255 Kbytes/seg	5	12 Kbytes/seg	1	35 Kbytes/seg	3	110 Kbytes/seg	4	110 Kbytes/seg	4
Clave	Sola	5	Sola (Dividida en tres partes)	4	Sola	5	Sola	5	Sola	5
PUNTUACIÓN	**TOTAL**	**44**	**TOTAL**	**27**	**TOTAL**	**24**	**TOTAL**	**34**	**TOTAL**	**30**

Teniendo en cuenta la información descrita, se puede concluir que las características necesarias, y de acuerdo con la calificación recibida según los ítems estipulados, el mejor algoritmo de cifrado simétrico es AES, Del anterior análisis se concluye que características necesarias y más relevantes que un algoritmo de cifrado simétrico debe cumplir son las siguientes.

- Longitud de clave 128 bits
- Tamaño de bloque de 128 bits
- Numero de rondas de 9

Por lo tanto, para la creación del nuevo algoritmo de cifrado simétrico denominado SARCH se parte de las características contenidas en el modelo matemático del algoritmo AES, posteriormente se procede a realizar modificaciones que permitan llegar a la solución más óptima para mejorar las características de AES con el nuevo algoritmo SARCH.

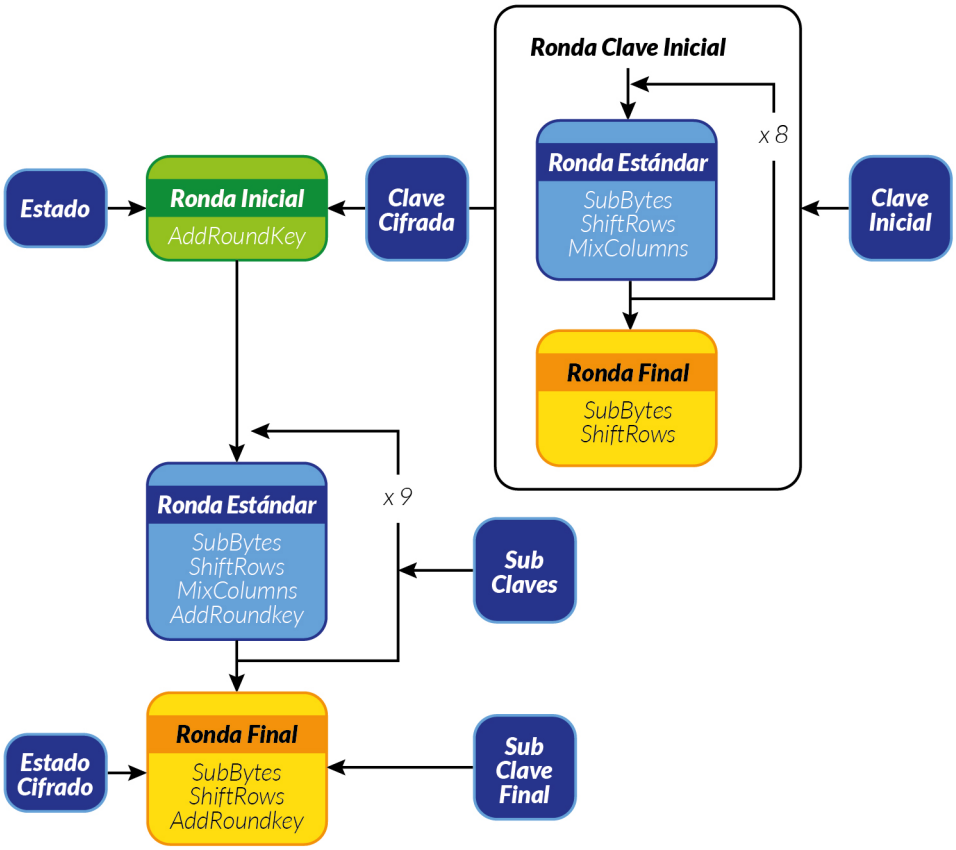


Fig. 1. *Diagrama de cifrado SARCH*

Modelo matemático del algoritmo de cifrado simétrico SARCH basado en AES.

Tamaño bloque del algoritmo de cifrado simétrico SARCH

Para realizar la construcción del algoritmo es necesario tener él cuenta el tamaño del bloque en este caso es de 128 bits por lo que los datos a ser encriptados se dividen en segmentos de 16 bytes (128 bits) lo cual se puede ver más claramente en la figura 2.

Fig. 2. *Bloque de cifrado de 16 bytes (128 bits)*

El gráfico muestra un bloque de 128 bits en el cual están los valores de A, B, C, D corresponderán a valores en hexadecimal. Este bloque también se lo puede expresar como una matriz 4x4 de la siguiente manera:

A1	B1	C1	D1
A2	B2	C2	D2
A3	B3	C3	D3
A4	B4	C4	D4

Fig. 3. *Matriz de estado*

Esta matriz es llamada como ***matriz de estado***

Clave

Como el algoritmo es tipo simétrico se utiliza la misma llave tanto para realizar el cifrado como para descifrar para nuestro caso la clave o llave será de una longitud de 128 bits.

a. Operación AddRoundKey

Partiendo de estos dos elementos anteriormente mencionados se procederá con la ronda inicial de modelo matemático la cual corresponde a AddRounKey donde se tendrá en cuenta el bloque a cifrar y la clave inicial como se muestra en la siguiente figura (Figura 4)

Fig. 4. *Ronda inicial (AddRounKey)*

Esta ronda inicial consiste en aplicar la operación XOR byte a byte entre el bloque a cifrar y la clave inicial como se observa en la siguiente figura (Figura 5)

Fig. 5. *Operación XOR*

+	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
1	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	10
2	2	3	4	5	6	7	8	9	A	B	C	D	E	F	10	11
3	3	4	5	6	7	8	9	A	B	C	D	E	F	10	11	12
4	4	5	6	7	8	9	A	B	C	D	E	F	10	11	12	13
5	5	6	7	8	9	A	B	C	D	E	F	10	11	12	13	14
6	6	7	8	9	A	B	C	D	E	F	10	11	12	13	14	15
7	7	8	9	A	B	C	D	E	F	10	11	12	13	14	15	16
8	8	9	A	B	C	D	E	F	10	11	12	13	14	15	16	17
9	9	A	B	C	D	E	F	10	11	12	13	14	15	16	17	18
A	A	B	C	D	E	F	10	11	12	13	14	15	16	17	18	19
B	B	C	D	E	F	10	11	12	13	14	15	16	17	18	19	IA
C	C	D	E	F	10	11	12	13	14	15	16	17	18	19	IA	IB
D	D	E	F	10	11	12	13	14	15	16	17	18	19	IA	IB	IC
E	E	F	10	11	12	13	14	15	16	17	18	19	IA	IB	IC	ID
F	F	10	11	12	13	14	15	16	17	18	19	IA	IB	IC	ID	IE

Fig. 6. *Tabla XOR*

Posteriormente el algoritmo sigue con la ronda estándar

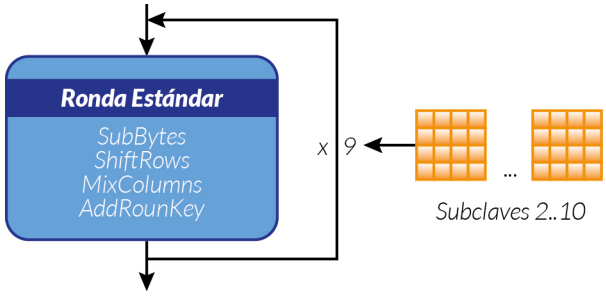


Fig. 7. *Ronda estándar*

b. Operación de sustitución (SubByte).

La transformación ByteSub es una sustitución de bytes no lineal, operando sobre cada uno de los bytes de estados independientemente. La transformación usa la misma tabla de sustitución (S-Box) para cada byte. La caja S-box ha sido seleccionada de tal manera que Rijndael sea resistente contra ataques lineales y diferenciales, así como ataques de interpolación. Esto asegura que los requerimientos tales como tener una pequeña correlación entre los bits de entrada y bits de salida y del hecho de que la salida no puede ser representada como una función matemática de la entrada [3].

La caja S-Box con la que se realiza la sustitución en el proceso de cifrado es la Figura 8.

hex	Y															
	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	63	7c	77	7b	f2	6b	6f	c5	30	01	67	2b	fe	d7	ab	76
1	ca	82	c9	7d	fa	59	47	f0	ad	d4	a2	af	9c	a4	72	c0
2	b7	fd	93	26	36	3f	f7	cc	34	a5	e5	f1	71	d8	31	15
3	04	c7	23	c3	18	96	05	9a	07	12	80	e2	eb	27	b2	75
4	09	83	2c	1a	1b	6e	5a	a0	52	3b	d6	b3	29	e3	2f	84
5	53	d1	00	ed	20	fc	b1	5b	6a	cb	be	39	4a	4c	58	cf
6	d0	ef	aa	fb	43	4d	33	85	45	f9	02	7f	50	3c	9f	a8
7	51	a3	40	8f	92	9d	38	f5	bc	b6	da	21	10	ff	f3	d2
8	cd	0c	13	ec	5f	97	44	17	c4	a7	7e	3d	64	5d	19	73
9	60	81	4f	dc	22	2a	90	88	46	ee	b8	14	de	5e	0b	db
a	e0	32	3a	0a	49	06	24	5c	c2	d3	ac	62	91	95	e4	79
b	e7	c8	37	6d	8d	d5	4e	a9	6c	56	f4	ea	65	7a	ae	08
c	ba	78	25	2e	1c	a6	b4	c6	e8	dd	74	1f	4b	bd	8b	8a
d	70	3e	b5	66	48	03	f6	0e	61	35	57	b9	86	c1	1d	9e
e	e1	f8	98	11	69	d9	8e	94	9b	1e	87	e9	ce	55	28	df
f	8c	a1	89	0d	bf	e6	42	68	41	99	2d	0f	b0	54	bb	16

X

Fig. 8. *S-box*

La caja S con la que se realiza la sustitución en el proceso de descifrado es la inversa por lo tanto es necesario calcular la inversa de la S-Box con la cual se realiza el cifrado el resultado es S-Box-Inversa. Como se puede apreciar Figura9.

hex		Y															
		0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	F
X	0	52	09	6a	d5	30	36	a5	38	bf	40	a3	9e	81	f3	d7	fb
	1	7c	e3	39	82	9b	2f	ff	87	34	8e	43	44	c4	de	e9	cb
	2	54	7b	94	32	a6	c2	23	3d	ee	4c	95	0b	42	fa	c3	4e
	3	08	2e	81	66	28	d	24	b2	76	5b	a2	49	6d	8b	d1	25
	4	72	f8	f6	64	86	68	98	16	d4	a4	5c	cc	5d	65	b6	92
	5	6c	70	48	50	fd	ed	b9	da	5e	15	46	57	a7	8d	9d	84
	6	90	d8	ab	00	8c	bc	d3	0a	f7	e4	58	05	b8	b3	45	06
	7	d0	2c	1e	8f	ca	3f	0f	02	c1	af	bd	03	01	13	8a	6b
	8	3a	91	11	41	4f	67	dc	ea	97	f2	cf	ce	f0	b4	e6	73
	9	96	ac	74	22	e7	ad	35	85	e2	f9	37	e8	1c	75	df	6e
	A	47	f1	1a	71	1d	29	c5	89	6f	b7	62	0e	aa	18	be	1b
	B	fc	56	3e	4b	c6	d2	79	20	9a	db	c0	fe	78	cd	5a	f4
	C	1f	dd	a8	33	88	07	c7	31	b1	12	10	59	27	80	ec	5f
	D	60	51	7f	a9	19	b5	4a	0d	2d	e6	7a	9f	93	c9	9c	ef
	E	a0	e0	3b	4d	ae	2a	f5	b0	c8	eb	bb	3c	83	53	99	61
	f	17	2b	04	7e	ba	77	d6	26	e1	69	14	63	55	21	0c	7d

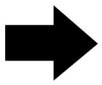
Fig. 9. *S-box inversa*

c. Operación para el desplazamiento de filas (ShitRow).

Esta operación se aplica a la matriz estado [ ], donde se realizan movimientos a manera de circulo por cada fila haciendo giros a la izquierda; como si cada fila fuera un anillo en donde la matriz estuviera conectada por las dos columnas extremas.

Donde la primera fila no toma ninguna rotación, la segunda hace un giro a la izquierda rotando una posición, la tercera por su parte rota dos posiciones y en la cuarta fila se rotan 3 posiciones. Estos movimientos se tienen en cuenta para al momento de cifrar la información.

0,0	0,1	0,2	0,3
1,0	1,1	1,2	1,3
2,0	2,1	2,2	2,3
3,0	3,1	3,2	3,3



0,0	0,1	0,2	0,3
1,1	1,2	1,3	1,0
2,2	2,3	2,0	2,1
3,3	3,0	3,1	3,2

Fig. 10. *Desplazamiento de filas shiftRow*

Fuente: Esta investigación

Para realizar el descifrado consiste en aplicar esta operación realizando los giros a la inversa (girar a la derecha) con el propósito de regresar a la posición inicial ver Figura 11.

1,1	1,2	1,3	1,0
2,2	2,3	2,0	2,1
3,3	3,0	3,1	3,2

1,0	1,1	1,2	1,3
2,0	2,1	2,2	2,3
3,0	3,1	3,2	3,3

Fig. 11. *Desplazamiento inverso de filas invertShiftRow*

d. Operación multiplicación de columnas (MixColums).

Se parte de la matriz generada en la operación anterior en donde se obtiene la matriz de estado posteriormente se realiza una operación de multiplicación de polinómica con una matriz 4x4 ya definida este proceso se realiza de esta manera [2].

Table. 3. *Matriz de estado*

d4	e0	b8	le
bf	b4	41	27
5d	52	11	98
30	ae	f1	35

Table. 4. *Matriz fija*

02	03	01	01
01	02	03	01
01	01	02	03
03	01	01	02

Se procede a escoger los valores de la primera columna de la matriz de estado

Table. 5. *Selección de valores de la primera columna de la matriz estado*

d4		02	03	01	01
bf		01	02	03	01
5d	.	01	01	02	03
30		03	01	01	02

La multiplicación se realiza con el primer elemento de la matriz de estado el cual se multiplica con el primer elemento de la fila de esta manera se realiza la suma de los respectivos resultados obteniendo así el nuevo valor del campo de la matriz de estado:

(d4 x 02) + (bf x 03) +(5d x 01) +(30 x 01) dando como resultado *04* ver Tabla 6

Table. 6. *Resultado la operación mixColums*

04	e0	48	28
66	cb	f8	06
81	19	d3	26
e5	9a	7a	4c

Este proceso se repite con todas las columnas de la matriz de estado este sería el resultado final de la nueva matriz de estado.

e. Operación para generar subclaves (key Schdule).

Para la generación de subclaves se parte de la clave inicial con ayuda de diferentes operaciones se obtiene las subclaves para diferentes rondas, estas operaciones se realizan con una matriz fija denominada Rcon

Table. 7. *Matriz fija Rcon*

01	02	03	08	10	20	40	80	1b	36
00	00	00	00	00	00	00	00	00	00
00	00	00	00	00	00	00	00	00	00
00	00	00	00	00	00	00	00	00	00

Table. 8. *Ejemplo de subclaves*

2b	28	4b	09	a0	88	23	2a
7e	ae	f7	cf	fa	54	a3	6c
15	d2	15	4f	fe	2c	39	76
15	a6	88	3c	17	B1	39	05

Primero se escoge la última columna de la clave inicial a la cual se le realiza la primera operación RotWord la cual consiste en mover el primer bit hacia el último elemento del a matriz entonces se tiene que:

Table. 9. *Operación RotWord*

09		cf
cf		4f
4f	RotWord	3c
3c		09

Posteriormente se aplica la operación de subBytes anteriormente ya mencionada

8a
84
eb
01

Finalmente se realiza una operación de XOR entre la primera columna de la clave inicial el resultado de la anterior operación y la primera columna de la matriz Rcon quedando de esta manera

Table. 10. *Operación XOR y entre la clave y la matriz Rcon*

2b		8a		01		a0
7e	XOR	84	XOR	00	=	fa
15		eb		00		fe
16		01		00		17

De esta manera se obtiene la primera columna de la subclave para la generación de las tres columnas faltantes se realiza una operación XOR entre el segundo elemento de la clave inicial y la primera columna de la subclave.

Table. 11. *Operación XOR entre el segundo elemento de la clave inicial y la primera columna de la subclave.*

28		a0		88
ae	XOR	fa	=	54
d2		fe		2c
a6		17		B1

Esta operación se repite con cada elemento de la clave inicial y la subclaves igualmente para la generación de claves posteriores

f. Ronda inicial aplicada a la clave

Para mejorar la seguridad en a la relación que existe entre la clave, el texto claro y el texto cifrado. Se vio la necesidad de agregar las operaciones que se observan en la siguiente figura:

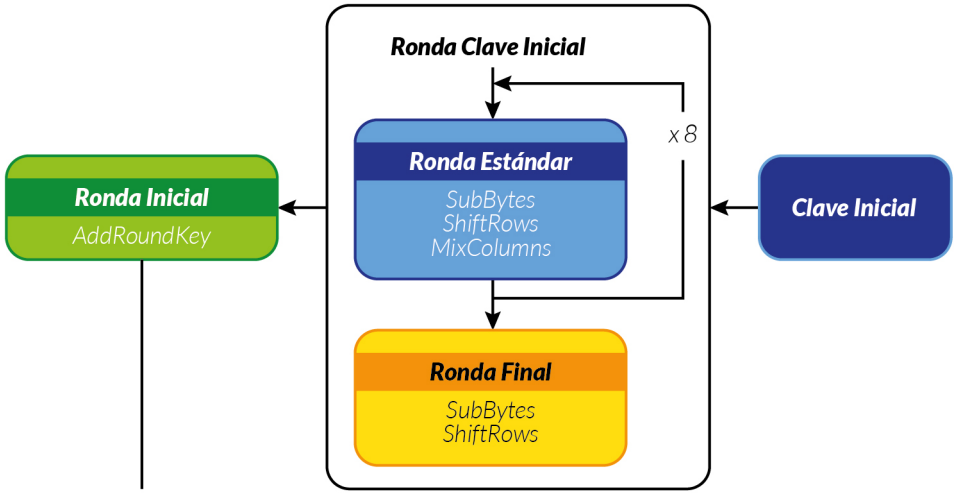


Fig. 12. *Ronda clave inicial*

El bloque de las operaciones agregadas se denominó como ronda clave inicial, la cual contiene una ronda estándar y una final esto con el objetivo de lograr una mayor ilegibilidad de la clave inicial antes de ser introducida en la ronda inicial del modelo matemático.

Al realizar este proceso se garantiza una mayor seguridad en el manejo de la clave, teniendo en cuenta que en un principio la clave utilizada para el cifrado sufre un cambio antes de iniciar el proceso de cifrado donde se combina la clave con el texto a cifrar conocido como ronda inicial, este cambio garantiza que la clave ingresada por el usuario y la utilizada para el cifrado no conserven ninguna relación.

III. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

Para alcanzar los objetivos descritos en el presente trabajo de investigación se analizó tres algoritmos de cifrado simétrico entre los cuales se encuentran el algoritmo TWOFISH, AES y SARCH este último fue desarrollado en la presente investigación. Con los resultados obtenidos, se verifico el nivel de seguridad alcanzado y así comprobar las hipótesis que se plantearon.

A. Criptoanálisis lineal

Técnica criptoanalítica de tipo estadístico, consistente en operar o-exclusivo dos bits del texto en claro, hacer lo mismo con otros dos del texto cifrado y volver a operar oexclusivo los dos bits obtenidos. Se obtiene un bit que es el resultado de componer con la misma operación dos bits de la clave. Si se usan textos en claro recopilados y los correspondientes textos cifrados, se pueden conjeturar los bits de la clave.

Para la prueba de criptoanálisis lineal es necesario primeramente conocer la tabla de verdad de la compuerta XOR.

Table. 12. *Tabla de verdad XOR*

XOR		
Entrada A	***Entrada B***	***Salida***
0	*0*	***0***
0	*1*	***1***
1	*0*	***1***
1	*1*	***0***

Para este proceso se realizó la operación XOR con los valores binarios de cada elemento a analizar, donde se tuvo en cuenta los diferentes octetos correspondientes a cada uno de ellos, primeramente, se realizó esta operación a la clave para posteriormente poder compararlos con los valores obtenidos más adelante.

Para esta prueba el texto y clave a utilizar serán los siguientes:

Clave:

Text (ASCII) = C1*4567D=?

Hexadecimal = 63 31 2A 34 35 36 37 64 3D 3F

Binario = 01100011 00110001 00101010 00110100 00110101 00110110 00110111 01100100 00111101 00111111

Texto:

Text (ASCII) = hola

Hexadecimal = 68 6F 6C 61

Binario = 01101000 01101111 01101100 01100001

a. Análisis de la clave

Una vez obtenidos los datos de la clave en formato binario se procedió a separar por octetos dichos valores, se realizó la operación de XOR por cada par de bits de cada octeto.

Los resultados obtenidos de la operación XOR entre los bits de la clave por octetos son los siguientes:

Primer octeto = 1100

Segundo octeto = 0001

Tercer octeto = 0111

Cuarto octeto = 0010

Quinto octeto = 0011

Sexto octeto = 0011

Séptimo octeto = 0010

Octavo = 1110

Noveno octeto = 0001

Decimo octeto = 0000

Estos resultados fueron de vital importancia para el análisis de los algoritmos de cifrado simétrico.

b. Análisis del algoritmo Twofish

Con la ayuda de la herramienta Online encrypt tool se realizó el cifrado con el texto y la clave propuestas anteriormente.

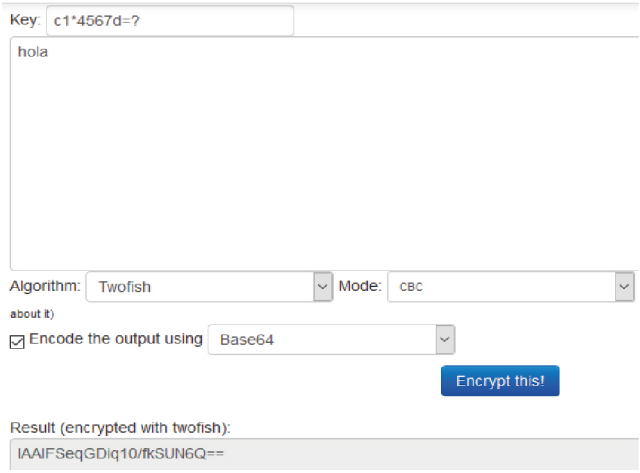


Fig. 13. *Cifrado Twofish*

Al igual que el texto y la clave el resultado de cifrado también fue convertido en formato binario

Clave:

Text (ASCII) = C1*4567D=?

Hexadecimal = 63 31 2A 34 35 36 37 64 3D 3F

Binario = 01100011 00110001 00101010 00110100 00110101 00110110 00110111 01100100 00111101 00111111

Texto:

Text (ASCII) = hola

Hexadecimal =68 6F 6C 61

Binario = 01101000 01101111 01101100 01100001

Cifrado:

Base64 = lAAIFSeqGDiq10/fkSUN6Q==

Hexadecimal = 94 00 08 15 27 AA 18 38 AA D7 4F DF 91

25 0D E9

Binario = 10010100 00000000 00001000 00010101 00100111 10101010 00011000 00111000 10101010 11010111 01001111 11011111 10010001 00100101 00001101 11101001

Lo siguiente fue realizar el análisis entre el texto plano o texto claro y el texto cifrado o criptograma, para lo cual se eligió los primeros cuatro octetos de cada uno de ellos a los cuales les aplico la operación XOR así:

Los resultados de la operación XOR entre texto claro y texto cifrado de los 4 primeros octetos son los siguientes:

Primer octeto = 0000

Segundo octeto = 1100

Tercer octeto = 1110

Cuarto octeto = 1010

Para poder observar mejor los datos se muestran en la siguiente tabla (Tabla 107):

Table. 13. *Resultados algoritmo Twofish*

Octetos	*Algoritmo Twofish*		
	Clave	*Coincidencias*	*Encontrado en*
1	*1100*		*Segundo octeto*
2	*0001*		
3	*0111*		
4	*0010*		
5	*0011*		
6	*0011*		
7	*0010*		
8	*1110*		*Tercer octeto*
9	*0001*		
10	*0000*		*Primer octeto*
Total, coincidencias	***3***		

En la anterior tabla se ha comparado la clave y el resultado de los 4 octetos con el fin de observar cuantas coincidencias se pudo encontrar, en este caso las coincidencias encontradas fueron 3 resultado que refleja que existe una probabilidad de 30% de encontrar la clave con la cual se realizó el proceso de cifrado.

c. *Análisis del Algoritmo AES*

Con la herramienta AES encryption se realizó el cifrado con el texto y la clave propuestas anteriormente

Fig. 14. *Cifrado AES*

Al igual que el texto y la clave el resultado de cifrado también fue convertido en formato binario.

Clave:

Text (ASCII) = C1*4567D=?

Hexadecimal = 63 31 2A 34 35 36 37 64 3D 3F

Binario = 01100011 00110001 00101010 00110100 00110101 00110110 00110111 01100100 00111101 00111111

Texto:

Text (ASCII) = hola

Hexadecimal =68 6F 6C 61

Binario = 01101000 01101111 01101100 01100001

Cifrado:

Base64 = e/CcZgTVJ+K5m23+BXhkMQ==

Hexadecimal = 7b f0 9c 66 04 d5 27 e2 b9 9b 6d fe 05 78 64 31

Binario = 01111011 11110000 10011100 01100110 00000100 11010101 00100111 11100010 10111001 10011011 01101101 11111110 00000101 01111000 01100100 00110001

Lo siguiente fue realizar el análisis entre el texto plano y el texto criptograma, para lo cual se ha elegido los primeros cuatro octetos de cada uno de ellos a los cuales les se les aplico la operación XOR así:

Los resultados de la operación XOR entre texto claro y texto cifrado de los 4 primeros octetos

Primer octeto = 0100

Segundo octeto = 1100

Tercer octeto = 0000

Cuarto octeto = 0010

Para poder observar mejor los datos se observan en la siguiente tabla (Tabla 14)

Table. 14. *Resultados algoritmo AES*

Octetos	Algoritmo AES		
	Clave	Coincidencias	Encontrado en
1	1100		Segundo octeto
2	0001		
3	0111		
4	0010		Cuarto octeto
5	0011		
6	0011		
7	0010		
8	1110		
9	0001		
10	0000		Tercer octeto
Total, coincidencias		**3**	

En la anterior tabla se ha comparado la clave y el resultado de los 4 octetos con el fin de observar cuantas coincidencias se pudo encontrar, en este caso las coincidencias encontradas fueron 3 resultado que refleja que existe una probabilidad de 30% de encontrar la clave con la cual se realizó el proceso de cifrado.

d. Análisis del Algoritmo SARCH

Este algoritmo fue el propuesto en la presente investigación

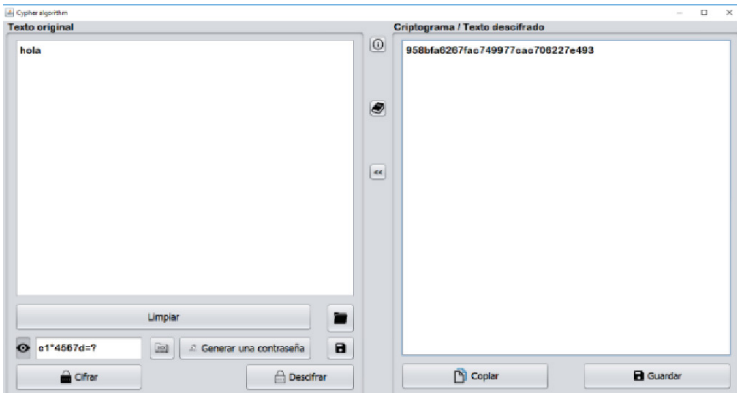


Fig. 15. *Cifrado SARCH*

Al igual que el texto y la clave el resultado de cifrado también fue convertido en formato binario

Clave:

Text (ASCII) = C1*4567D=?

Hexadecimal = 63 31 2A 34 35 36 37 64 3D 3F

Binario = 01100011 00110001 00101010 00110100 00110101 00110110 00110111 01100100 00111101 00111111

Texto:

Text (ASCII) = hola

Hexadecimal = 68 6F 6C 61

Binario = 01101000 01101111 01101100 01100001

Cifrado:

Base64 = lYv6Ymf6x0mXfKxwYifkkw==

Hexadecimal = 95 8B FA 62 67 FA C7 49 97 7C AC 70 62 27 E4 93

Binario = 10010101 10001011 11111010 01100010 01100111 11111010 11000111 01001001 10010111 01111100 10101100 01110000 01100010 00100111 11100100 10010011

Lo siguiente fue realizar el análisis entre el texto claro y el texto cifrado, para lo cual se ha elegido los primeros cuatro octetos de cada uno de ellos a los cuales se les aplico la operación XOR así:

Resultados de la operación XOR entre texto claro y texto cifrado de los 4 primeros octetos

Primer octeto = 0001

Segundo octeto = 0110

Tercer octeto = 1111

Cuarto octeto = 0000

Para poder observar mejor los datos se expresan en la siguiente tabla:

Table. 15. *Resultados algoritmo SARCH*

Octetos	Algoritmo SARCH		
	Clave	Coincidencias	Encontrado en
1	1100		
2	0001		Primer octeto
3	0111		
4	0010		
5	0011		
6	0011		
7	0010		
8	1110		
9	0001		
10	0000		Cuarto octeto
Total, coincidencias	**2**		

En la anterior tabla se ha comparado la llave y el resultado de los 4 octetos con el fin de observar cuantas coincidencias se pudo encontrar, en este caso las coincidencias encontradas fueron 2 resultado que refleja que existe una probabilidad de 20% de encontrar la cave con la cual se realizó el proceso de cifrado.

e. Resultado final del criptoanálisis lineal de los algoritmos TWOFISH, AES y SARCH

Se realizó el análisis de los algoritmos Twofish, AES Y SARCH con el fin observar su comportamiento en cuanto al criptoanálisis lineal con ello se logró encontrar coincidencias entre la llave, el texto plano y el texto cifrado o criptograma, los datos obtenidos se organizaron como se refleja en la siguiente tabla (Tabla 16):

Table. 16. *Comparación de Algoritmos de cifrado simétrico*

Octetos	Comparación			
	Clave	Algoritmo Twofish	Algoritmo AES	Algoritmo SARCH
1	1100			
2	0001			
3	0111			
4	0010			
5	0011			
6	0011			
7	0010			
8	1110			
9	0001			
10	0000			
Total, coincidencias		**3**	**3**	**2**

En la anterior tabla se puede observar todos los datos referentes a la relación que existe entre la clave y el respectivo análisis de cada uno de los algoritmos utilizados para esta prueba.

Para lograr una mejor interpretación y comprensión de los datos se optó por realizar un diagrama de barras que facilito el análisis de los datos.

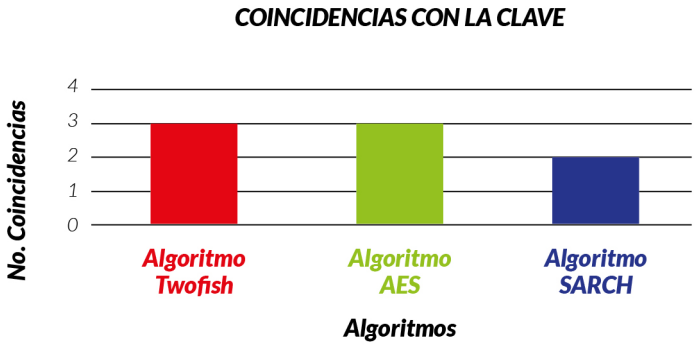


Fig. 16. *Diagrama de barras coincidencias con la clave*

De acuerdo con los datos observados en el anterior diagrama de barras, los algoritmos Twofish y AES tienen un número de coincidencias igual a 3 con respecto a los valores de la clave, mientras que el algoritmo SARCH cuenta con un valor de 2 coincidencias lo cual garantiza una mayor seguridad entre la clave, el texto claro y el texto cifrado en relación con los anteriores algoritmos.

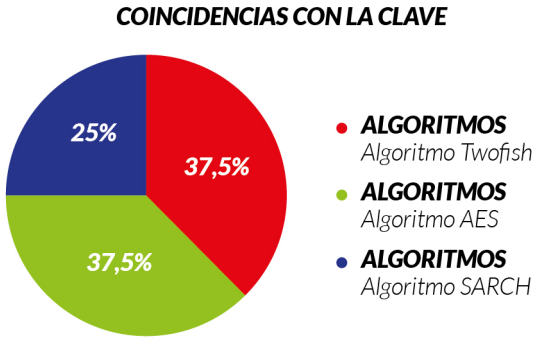


Fig. 17. *Diagrama circular coincidencias con la clave*

En la anterior grafica se observa que el algoritmo Twofish tiene un 37,5%, al igual que el algoritmo AES de coincidencias con respecto a la clave, por otro lado, el algoritmo SHARCH cuenta con un valor 25% de coincidencias lo cual indica un menor riesgo en el momento de encontrar debilidades en cuanto a la relación de la clave, el texto claro y el texto cifrado.

IV. CONCLUSIONES

La implementación del algoritmo de cifrado simétrico SARCH mejoró la protección de la información dado que el nivel de seguridad que ofrece es mayor a otros algoritmos.

Tras culminar la investigación se pudo determinar que la criptografía en sus diferentes tipos y variantes es un mecanismo de vital importancia al momento de proteger información digital, a su vez garantiza el cumplimiento de los tres pilares de la seguridad de la información (confidencialidad integridad y disponibilidad) [5].

Se formuló un modelo matemático que cuenta con un nuevo bloque de operaciones denominado ronda clave inicial, por medio del cual se logró la construcción del algoritmo de cifrado simétrico SARCH.

Se llegó a codificar el algoritmo en el lenguaje de programación java obedeciendo la estructura planteada en el modelo matemático, donde se construyó una aplicación de escritorio denominada encryptSarch la cual permite cifrar y descifrar textos planos.

Por medio de las pruebas realizadas al algoritmo se pudo determinar que este se encuentra en un menor riesgo de debilidades en relación con los anteriores comparados en esta investigación.

REFERENCIAS

[1] *ADAM, Berent. AES (Advanced Encryption Standard) Simplified. 2017. Revista ABIsoftware development. Disponible en <https://www.ime.usp.br/~rt/cranalysis/AESSimplified.pdf>*

[2] *Bertolín, J. A. (2008). Seguridad de la información. Redes, informática y sistemas de información. Editorial Paraninfo.*

[3] *Briones García J. A (2002), Implementación del Advanced Encryption Standard Rijndael en C, [en linea], 5 de agosto del 2002 Disponible en Internet: <http://delta.cs.cinvestav.mx/~francisco/arith/AESJbriones.pdf>*

[4] *CRIPTOHISTORIA. Introducción al criptoanálisis. España. Obtenido de <http://www.criptohistoria.es/files/Criptoanalisis.pdf>.*

[5] *ISO 27001, Sistema de Gestión de la Seguridad de la Información. Obtenido de <http://www.iso27000.es/download/doc_sgsi_all.pdf >.*

[6] *LUCENA LÓPEZ, M. J. (2015). Criptografía y seguridad en computadores. jaen, Andalucía, España: Universidad de jaen. Obtenido de <https://drive.google.com/file/d/0B9iIPOseFevtbGp2bEdtZzhi RUE/view?usp=sharing>*

[7] *MÉNDEZ APARICIO L. G. (2012). principios para el análisis de criptografía digital. Guatemala. Obtenido de <http://biblioteca.usac.edu.gt/tesis/08/08_0316_EO.pdf>*

[8] *ORJUELA NAVARRETE, J. D. (7 de Julio de 2008). desarrollo de un algoritmo de cifrado. Obtenido de <http://bibliotecadigital.usb.edu.co/bitstream/10819/2951/1/algoritmo_cifrado_simetrico_orjuela_2008.pdf>*

[9] *POUSA, Adrián. Algoritmo de cifrado simétrico aes. aceleración de tiempo de cómputo sobre arquitecturas multicore. Facultad de Informática - Universidad Nacional de La Plata. 2011. disponible en <https://postgrado.info.unlp.edu.ar/wpcontent/uploads/2014/07/Pousa_Adrian.pdf>*

[10] *RIOS ARRAÑAGA, Jaime David, et al. New S-box calculation for Rijndael-AES based on an artificial neural network ReCIBE. Revista electrónica de Computación, Informática, Biomédica y Electrónica. 2017. Universidad de Guadalajara Guadalajara, México. Disponible en <https://www.redalyc.org/jatsRepo/5122/512253718012/5122 53718012.pdf>.*

[11] *SIMANTEC. Informe de Symantec sobre las amenazas para la seguridad de los sitios web [en línea]. [Madrid, España] Symantec Spain S.L, feb. 2015. Disponible en internet: <https://www.certisur.com/sites/default/files/docs/Informeamenazas-2015.pdf>*

[12] *SOTO SANCHES J. J. (2010). Algoritmo de encriptación simettrico que procesa bloques de 192 bits, con llaves de cifrado de 192 bits, empleando los teoremas factorial y jv. Mexico D.F, Mexico. Obtenido de <http://tesis.ipn.mx/bitstream/handle/123456789/9407/165.pdf?sequence=1>*

[13] *Temas Selectos de Matemáticas. 3ra Unidad. Disponible en <http://prod11umtms.blogspot.com/2014/05/3raunidad.html>.*

[14] *Temas Selectos de Matemáticas. 3ra Unidad. Disponible en <http://prod11umtms.blogspot.com/2014/05/3raunidad.html>.*

[15] *VICENT FRANCÉS, J. F. (2008). Propuesta y análisis de criptosistemas de clave pública basados en matrices triangulares superiores por bloques. San Vicente del Raspeig, España: Biblioteca Virtual Miguel de Cervantes. Obtenido de <http://www.cervantesvirtual.com/nd/ark:/59851/bmcdn4k8>*

Correspondencia

a. Doctor en Ciencias de la Electrónica. Docente tiempo completo, Universidad Cesmag. E-mail: aeraso@unicesmag.edu.co. Pasto, Colombia.

Efectos Blockchain como estrategia de incorporación con criptomonedas en mercados de capitales

Rider Geovanny Benavides B. (a)
Elizabeth del Rosario Guerrero A. (b)

Resumen

La comercialización de criptomonedas y sus efectos generados en los procesos de oferta y demanda, establece toda una serie de actividades, ventajas y desventajas, que permiten el avance de la presente investigación, las tecnologías blockchain, y su influencia en los mercados de capitales, condiciona toda una serie de responsabilidades conjuntas, entre la sociedad y las nuevas tecnologías emergentes, las garantías de crecimiento, amenazas o posibles alternativas de segregación dispuestas en las negociaciones con criptodivisas, limitan el adecuado manejo de transacciones, minimizan los tiempos de ejecución y depuran de forma avanzada los estados de seguridad y accesibilidad en la cuantiosa e inagotable mina analógica del comercio electrónico.

Palabras clave: Blockchain, criptomoneda, banca, fluctuación, volatilidad, negocios, comercio electrónico.

Blockchain effects like onboarding strategy with cryptocurrencies in capital markets

Abstract

The commercialization of cryptocurrencies and its effects generated in the processes supply and demand, establishes a whole series of activities, advantages and disadvantages, that allow the advancement of this research, blockchain technologies, and their influence on capital markets, conditions a whole series of responsibilities joint, between society and new emerging technologies, growth guarantees, threats or possible alternatives of segregation provided in the negotiations with cryptocurrencies, limit the proper handling of transactions, minimize time execution and advanced debugging of security and accessibility states in the vast and inexhaustible analog mine of electronic commerce.

Keywords: Blockchain, cryptocurrency, banking, fluctuation, volatility, business, electronic commerce.

I. INTRODUCCIÓN

La era digital en la que nos encontramos inmersos, permite vislumbrar toda una serie de oportunidades y retos que dinamizan la economía y el desarrollo de la sociedad. Tal y como lo afirma [1], quien, en esta premisa permite indagar sobre el sinnúmero de posibles nuevas oportunidades de negocio dispuestos en el comercio electrónico.

El aumento en el consumo de tecnologías, a partir del internet de las cosas, crece de forma exponencial, las estrategias de mercado y venta para su distribución, se intensifican de acuerdo a las necesidades de los usuarios en línea, las transacciones de moneda electrónica, se dinamizan y diversifican de acuerdo al consumo mundial, los parámetros a tener en cuenta para la prestación de servicios, generan rentabilidad en los negocios y permiten cautivar, seducir e incorporar un mayor número de clientes dispuestos a invertir en alternativas renTable.s que incorporen tecnologías de la información y las comunicaciones, tal y como lo menciona, [2], en su artículo The Latin American Experience,: "hay 179 millones de personas usuarias de telefonía móvil en América Latina; el 82% de las cuales, acceden a Internet desde su teléfono móvil, el 73%, utilizan los servicios de mensajería y un 55%, transfieren datos a través de otros mecanismos. En países como Perú, solo un 0,01% de las economías domésticas rurales disponen de acceso a Internet, mientras que un 36,5% poseen un terminal móvil. En Chile, el porcentaje de penetración de la telefonía móvil es del 94,7.Con respecto a Colombia, según el Ministerio de las Tic, para octubre de 2017, el 98% de municipios, se encuentran conectados a Internet, a partir de 28 millones de conexiones [3].

Legalmente el dinero es una representación abstracta de un valor (independiente del propio valor inherente al papel o al metal) respaldada por una autoridad y generalmente admitida para la realización de intercambios [4]. La capitalización de moneda virtual a partir de criptomonedas, es considerada como un referente futurista que permite dilucidar posibles alternativas mucho más dinámicas, que ayuden a mejorar los actuales índices de comercialización en la moneda corriente.

II. METODOLOGÍA

El presente artículo plantea un estudio pragmático, basado en una investigación de tipo analítico, que contempla la recolección y adecuada organización de información, con el fin de elaborar un diagnóstico general, que permita conocer las condiciones propias del proceso investigativo. Por otra parte, se trata de una investigación de tipo descriptiva y documental, ya que representa el comportamiento de las criptomonedas en los negocios, a partir de un contexto tecnológico y financiero, que permita desarrollar lineamientos y parámetros ligados a un enfoque investigativo de tipo mixto, centrado en el análisis numérico de datos y la referenciación de fuentes primarias de información, concernientes al manejo y transacción de criptomonedas. Al respecto, [5] esTable.ce que; La investigación documental es un proceso basado en la búsqueda, recuperación, análisis, crítica e interpretación de datos secundarios, es decir, los obtenidos y registrados por otros investigadores en fuentes documentales: impresas, audiovisuales o electrónicas, que como en toda investigación, su propósito consiste en un análisis estratégico de la información, que además, permita suscitar, aportes significativos en la generación de nuevos conocimientos.

La metodología desarrollada, plantea un estudio que parte desde lo deductivo y exploratorio, ya que se indaga y analiza la información recopilada por diversas fuentes relacionadas con el tema de estudio, los resultados obtenidos, contemplan una metodología no experimental y de carácter transversal, que permite integrar factores relevantes, en el desarrollo de estrategias y mecanismos de funcionamiento, basados en la incorporación y difusión de criptomonedas, dentro de un periodo de tiempo previamente estipulado, con un soporte teórico general, en el que sea posible visualizar el sector estudiado y las variaciones que contemple, según una serie de políticas de funcionamiento y niveles de aceptabilidad requeridas. Tal y como lo describe [6], al mencionar que: El estudio puede comenzar siendo exploratorio para posteriormente proceder a la descripción, explicación, predicción y/o evaluación, de los contenidos programáticos.

La información ha sido recolectada, analizada y debidamente caracterizada, teniendo en cuenta diversas fuentes de información, dispuestas entre libros, revistas, periódicos, artículos, informes de trabajos y demás fuentes electrónicas o físicas propias del contexto investigativo; Luego de hacer una revisión detallada de los contenidos dispuestos, se consolida la información segregando las fuentes a partir de los referentes más representativos, que en definitiva, den lugar a la construcción consolidada del documento desarrollado. Para la recolección de la información, además de libros físicos, se consultan fuentes electrónicas, se analizan teorías y se consulta expertos, a partir de una serie de cadenas de búsqueda, compuestas por palabras y frases clave, referentes al tema de investigación, buscando encontrar documentos y escritos relevantes, que permitan generar un claro referente hacia las temáticas y contenidos más destacados en el ámbito de contexto estudiado.

III. RESULTADOS Y DISCUSIÓN

Luego de un análisis minucioso de la información recolectada, a partir de varias fuentes de información fidedignas y de diversa índole, ha sido posible generar una serie de resultados específicos, que demuestran las condiciones y efectos generados en la economía de consumo digital. Los diversos factores y variables que se tienen en cuenta para su regulación, esTable.cen toda una serie de posibilidades respecto a su financiación y tratamiento. La validación de características generales de las criptomonedas, en cuanto a ventajas, desventajas, fortalezas, oportunidades, debilidades o amenazas, son reflejadas a partir de una mirada clara, que permite visualizar, posibles futuras nuevas oportunidades de negociación, en cuanto al crecimiento o desvalorización de la economía digital emergente.

3.1 Matrices de eficiencia, respecto a ventajas o desventajas presentes en la distribución y comercialización de monedas sociales, en comparación con las criptomonedas.

Como punto de comparación, para la distribución del mercado físico y virtual de la moneda, se presenta a continuación, un comparativo de las posibles ventajas y desventajas que trae consigo el uso de monedas digitales, en relación a las monedas sociales, según un estudio hecho por parte del investigador Paredes, en su libro titulado: El uso de criptomonedas, como medio de pago de las Mymes exportadoras. Dicho estudio permite dilucidar las alternativas de mejora o posibles inconvenientes dispuestos en la incorporación y distribución de este tipo de divisas.

Table. 1. *Comparación general entre monedas digitales y moneda corriente. Fuente: [7].*

	Ventajas	Desventajas
Moneda Virtual	• Libertada de pago. • Seguridad y control. • Transparente y neutral. • Transacciones en tiempo real. • Es una moneda anónima.	• Enorme fluctuación. • Grado de aceptación. • Confianza. • Volatilidad.
Papel Moneda	• Eficiencia. • Confiabilidad. • Tamaño. • Durabilidad. • Fácil Transporte.	• Tienden a ser Falsificados. • Es muy frágil. • Costos de fabricación y emisión. • Costo de transporte.

3.2 Principales características, respecto a los tiempos de operación en las criptomonedas más representativas del mercado.

Si se hace alusión a los tiempos de espera y niveles de aceleración en las criptomonedas, en la siguiente tabla es posible visua-lizar, un comparativo de tiempos y niveles de transacciones efectuadas por segundo, teniendo en cuenta, factores de rendimiento en las operaciones generadas, así como, su porcentaje de energía consumida, respecto al tamaño de transacciones requeridas.

Table. 2. *Tiempos de ejecución según niveles de transacción en criptomonedas más representativas del mercado. Fuente: [8].*

Nombre	Acrónimo	Año	Blockchain	Consenso	Tiempo Bloque
Bitcoin	BTC	2009	Público	POW	10 minutos
Ether	ETH	2015	Público (propio)	POW	14 segundos
Bitcoin Cash	BCH	2017	Público (parte de bitcoin)	POW	Variables (18minutos media)
Ripple	XRC	2012	Privado (Riplenet)	Ripple	4 Segundos
Litecoin	LTC	2011	Publico (Basado en bitcoin)	POW	2,5 minutos
Dash	DASH	2014	Publico (Basado en bitcoin)	POW + POS (Service)	2,5 minutos
Bitconnect	BCC	2016	Público (propio)	POS + POS (Stake)	Minutos
Antshares NEO	NEO	2016	Público (propio)	Dbtf	Segundos
NEM	XEM	2015	Público (propio)	POI	1 minuto
Monero	XMR	2014	Público	CriptoNight POW	2 minutos

Para estudiar la capacidad de ejecución de las criptomonedas y tomando como punto de comparación al bitcoin, es posible determinar, que en relación con otras criptomonedas, su tiempo de bloque (10minutos), es relativamente alto a la media, dispuesto entre 2,5 y 3 minutos, haciéndola poco eficiente y demorada, en comparación a monedas como: Ether, Ripple, NEO, que ofrecen tiempos de espera, incluso menores a un minuto.

Table. 3. *Principales criptomonedas según tiempos de operaciones. Fuente: [8].*

Nombre	Confirmación	Privacidad	Mac tps[5]	Máximo monedas	Criptografía POW	Consumo Energía
Bitcoin	1 hora	Mediabaja	3-7	21 Millones	SHA256	Muy alto
Ether	3 minutos	Mediabaja	15	18 Millones por año	ethcash	Alto
Bitcoin Cash	2 horas	Mediabaja	28	21 Millones	SHA256	Alto
Ripple	4 segundos	Buena (privado)	1500	100 Millones	-	Muy bajo
Litecoin	15 minutos	Mediabaja	28	84 Millones	Script	Alto
Dash	minutos (opc instantáneo)	Alta (opcional)	1500-3500	19 Millones	Other	Muy alto
Bitconnect	minutos (opc instantáneo)	Mediabaja		28 Millones	?	Alto
Antshares NEO	Segundos	Mediabaja	1000	100 Millones	NeoQs	Alto
NEM	minutos (opc instantáneo)	Mediabaja	>100	9000 Millones	-	Bajo
Monero	20 minutos	Alta	>100	Sin limite	CriptoNote	Alto

Si se observa con detenimiento los datos generados en la tabla anterior, es posible determinar que en cuanto a niveles de inversión, es Dash, la moneda que menor índice presenta (19 millones), por su parte bitcoins, se encuentra en 21 millones como límite de adqusicion, dicho valor sigue aumentando, hasta llegar a un punto máximo de 9.000 millones, para el caso de NEM, sin embargo, es posible encontrar, que Monero, incluso no presenta límite, haciendo su valor de restricción indeterminado. Dadas estas circunstancias, es posible deducir, que el número de transacciones y capacidad de inversión, son determinantes en el desarrollo del mercado emergente para las criptomonedas, dado que presentan unos lineamientos que permiten divagar de forma conjunta, sobre los gustos o preferencias dispuestas por parte de los clientes interesados.

3.3 Matrices de análisis, respecto a las fortalezas, oportunidades, debilidades y amenazas presentes en la generación y utilización de criptomonedas.

Para tener una idea más clara sobre varias de las características más representativas, que dan lugar a la escogencia de monedas virtuales, a continuación se presenta un análisis efectuado por parte de una investigación realizada por Erbiti, M. en la que, basado en la matriz FODA, es posible identificar varias de las fortalezas, amenazas, oportunidades y debilidades, que presenta la incorporación de este tipo de divisas, en el mercado de la sociedad de consumo emergente.

Table. 4. *Matriz de fortalezas oportunidades, debilidades y amenazas concernientes al uso de criptomonedas. Fuente: [9].*

Fortalezas	***Oportunidades***
• *Para los adeptos de liberalismo económico: ninguna institución central controla la creación y emisión de bitcoins. La innovación se basa en un esquema anárquico de transmisión de datos, el mismo en la que se fundamenta el internet.* • *Su oferta monetaria esta predeterminada en el algoritmo que le da origen a la red, por ello es considerada por muchos como una moneda antiinflacionaria.* • *Usa la criptografía informática para garantizar altos estándares de seguridad, evitando la falsificación del circulante.* • *Provee al sistema financiero y al mercado de capitales en particular, de una nueva alternativa de inversión, como el atesoramiento de cualquier divisa con fines especulativos.*	• *Previsible obsolencia de papel moneda. Las divisas tradicionales están atadas a una vieja tecnología que se la imprenta Proliferación de comunidades virtuales que crean esquemas de monedas virtuales para intercambiar bienes y servicios, en la economía "virtual" o en la "real".* • *Alto índice mundial de personas no banca rizadas: Bitcoin podría incorporar al sistema financiero formal a personas que antes estaban excluida, ya que estas solo necesitan contar con acceso a internet.*
Debilidades	***Amenazas***
• *La fluctuación de su precio: como su oferta es fija, el precio lo determinan los drivers que mueven la curva de la demanda (la seguridad y confianza que inspire la red a sus usuarios, la robustez del ecosistema, la eficiencia como moneda para el intercambio, el grado de adopción en la sociedad, etc.).* • *Baja liquidez de bitcoins en el mercado.* • *Alto costo en inversión de infraestructura para minar bitcoins.* • *Su carácter abstracto plantea un desafío ya que son concretos los riesgos que presenta el anonimato de las monedas virtuales para los entes reguladores que centran su actividad en la prevención de los delitos de lavado de activos y financiación del terrorismo.* • *Si el usuario no toma las precauciones necesarias, puede eliminar o perder sus bitcoins sin poder recuperarlos: si el file digital se pierde, el dinero se pierde.*	• *Marco regulatorio difuso. No hay homogeneidad entre los países en cuanto a su regulación y tratamiento.* • *Altas probabilidades de que individuos que llevan adelante operaciones fraudulentas utilicen esta moneda como una alternativa para el financiamiento de actividades ilícitas, en buena parte porque las cuentas donde se atesora y las operaciones con la moneda no poseen información identificadora.* • *Capta mucho la atención de los entes de contralor, porque se encuentra fuera de la órbita de los Bancos Centrales.* • *Entes reguladores y gobiernos que prohíben Bitcoin lo hacen prohibiendo la actividad de los exchanges evitando cualquier operación cambiaria con Bitcoins (p.e. China).* • *Antecedentes negativos: la exposición mediática de los acontecimientos que atentan contra la subsistencia del ecosistema bitcoin, aumentan la volatilidad de la moneda.*

3.4 Índices de distribución, comercialización, capitalización y volatilidad, en el actual mercado de las criptomonedas, en comparación con otras monedas reguladas.

Para tener una idea más clara sobre los índices de volatilidad que ejercen las criptomonedas como bitcoin, en comparación a otros tipos de monedas regulares, como

el Yen, el Euro y otras monedas representativas del mercado de capitales, se presenta el siguiente gráfico que demuestra, los altos niveles de riesgo a considerar para su incorporación.

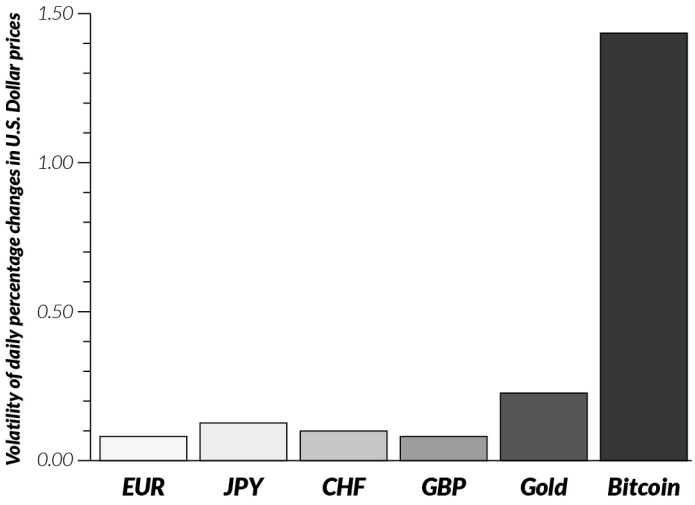
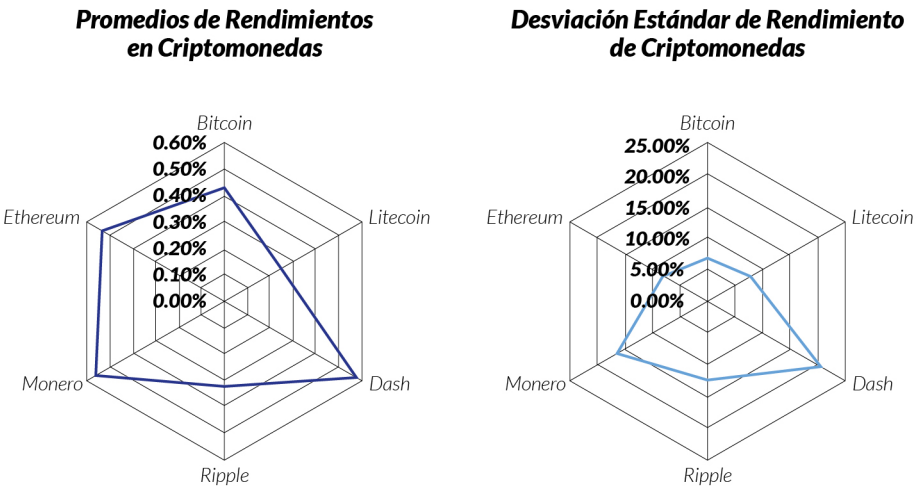


Fig. 1. *Volatilidad de las criptomonedas en comparación con monedas reguladas. Fuente: [10].*

Por medio de la gráfica anterior, dispuesta por parte de [10], es posible idealizar la enorme volatilidad que presenta el Bitcoin. Su valor está calculado, utilizando los datos diarios de 2013 anualizados, según el ratio de intercambio con el dólar americano. Y se compara con la volatilidad del euro, el yen japonés, el franco suizo, la libra británica y el oro. Como se puede ver, la volatilidad de Bitcoin es cuantiosa, 142%, lo que limita su uso, como moneda y demuestra la especulación subyacente en su distribución.

3.5 Rendimiento de las criptomonedas.



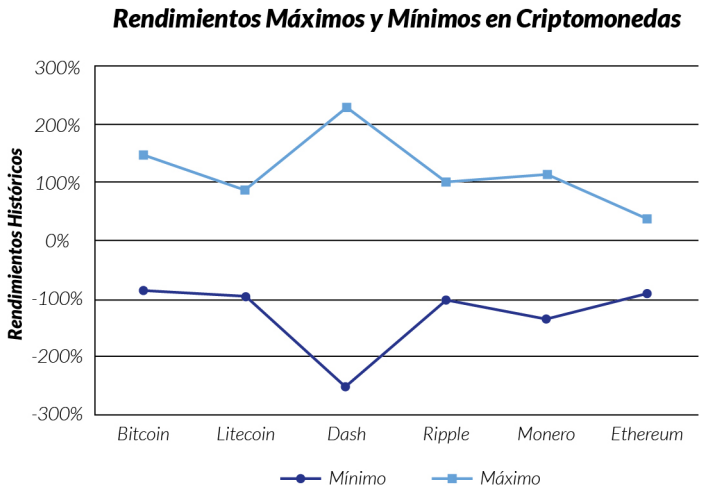


Fig. 2. *Promedios y desviación estándar de rendimientos en criptomonedas. Fuente: [11].*

Invertir capacidad computacional para procesar transacciones, garantizar la seguridad de la red, y conseguir que todos los participantes estén sincronizados. Es la clave de la sostenibilidad de todo el sistema [12]. Los máximos y mínimos de las rentabilidades diarias se muestran en el Gráfico, se aprecia que todas las criptomonedas con excepción de Ethereum, han tenido en su historia rendimientos iguales o superiores al 100%, en especial el Dash, quien en Febrero del 2014, mostró una rentabilidad diaria de 227%, esto debido a que en ese entonces se inició a "minar" solo con tarjeta gráfica, aunque tal hecho solo duró un día [11].

IV. DISCUSIÓN

Desarrollar nuevas alternativas de moneda en la que solo las personas intercambien de forma directa su dinero, evita intermediarios y la generación de obligaciones tributarias o financieras que controlen o limiten los niveles de negociación, con ello es posible, minimizar los índices de financiamiento y controlar los niveles de efectivo fluctuante; en el sector público por ejemplo, permite evitar riesgos relacionados con el desfalco o segregación de dineros, que enriquezcan la banca y otorguen beneficios particulares a la ya desigual y creciente sociedad diferencial de mercado, en el consumo y adquisición de divisas emergentes.

Los beneficios que trae consigo la incorporación de moneda virtual o criptomonedas a partir de tecnologías blockchain, se basa en la aparente seguridad de sus transacciones, si uno de los integrantes de la negociación intenta gastar más de lo que tiene en su fondo de dinero virtual, el sistema revisa los libros conTable.s e informa de forma inmediata sobre dicha situación al interesado, este suceso, genera el rechazo de la solicitud en trámite y por tanto el sistema criptográfico preserva su autenticidad y evita su falsificación [13].

EsTable.cer qué tipo de moneda virtual es la que mejores rendimientos trae para sus inversionistas, es un factor que está limitado a diversos lineamientos y estipulaciones,

los mercados de capital, están fuertemente regulados y sectorizados en secciones de rentabilidad ligadas a procesos productivos de cada nacionalidad, pensar en una moneda mundial que se asemeje de manera conjunta entre todas las fuerzas de mercado, es algo atrevido, ya que requiere de mecanismos bursátiles igualitarios, difíciles de parametrizar.

V. CONCLUSIONES

Las criptomonedas son un referente tecnológico que ha venido evolucionando al mismo nivel que el desarrollo de las nuevas tecnologías, su articulación como moneda legal y su regulación ante entes autorizados, está determinado por diversos factores, que en definitiva, limitan su funcionamiento y parametrizan su evolución en el tiempo, estos y otros agentes diferenciadores, hacen que se tomen como referente al comportamiento y evolución de las criptomonedas. Es po eso que Blockchain, puede considerarse como una base de datos distribuida que puede ser pública o privada con datos estampados e inmuTable.s relacionados con cada transacción realizada. Esta tecnología es el corazón de la criptomoneda Bitcoin [14].

La banca presenta toda una serie de restricciones, ante la disposición e incorporación de criptomonedas, los niveles de aceptabilidad, están ligados ante toda una serie de lineamientos y parámetros de seguridad, que condicionan y limitan la factibilidad o no conveniencia de nuevas alternativas de mercado, en el ya dominado y subyugado negocio de divisas emergentes; lo que sí está claro, es que las nuevas tecnologías basadas en blockchain, están incursionando de manera agresiva en los procesos internos de cualquier entidad, buscando encontrar mayores y mejores niveles de distribución electrónica, que permitan además, su fortalecimiento y disposición en grandes niveles de transacciones y altos niveles de seguridad en sus operaciones.

Los índices de eficiencia que presentan las criptodivisas, están limitadas de acuerdo a sus tiempos de ejecución, dichos lineamientos dependen de la velocidad en sus transacciones y de la capacidad de incursion en los mercados capitales, los niveles de seguridad y volumen de tiempo requerido, permiten validar de forma efectiva, cada uno de los procesos necesarios para su correcta ejecución.

Las instituciones gubernamentales a corto o largo plazo, podrán regular el registro y adaptación comercial de las llamadas criptomonedas, las tendencias en su utilización son altas y crecen de manera paralela a los avances propios de la tecnología; la estimación y ponderación de su valor adquisitivo en moneda de curso legal, conseguirá ser reglamentada de acuerdo a una serie de consideraciones y retos que permitan poner en marcha su normal distribución, teniendo en cuenta los niveles de inclusión y regulación de los mercados capitales.

Los altos niveles de fluctuación en la distribución y comercialización de criptomonedas, pueden llevar, de igual forma, a la completa desestabilización de este recurso digital, dejando en la quiebra a muchos inversionistas incautos, que en lugar de alcanzar el propósito de enriquecerse de la noche a la mañana, terminan buscando culpables o alimentando una vaga ansiedad, que solo conlleva a un mal desempeño de sus inversiones.

Nunca se sabe si es buen o mal momento para invertir en Bitcoin, o en otras criptomonedas, lo más seguro es que los niveles de fluctuación generados esTable.cen cambios drásticos en sus valores de adquisición, sobre todo, tras el anuncio de nuevos forks

(copiar un proyecto a partir de este); que en sus contexto, la gente que tiene bitcoins, pasa a tener más bitcoins, pero repartidos en otras redes. Las criptomonedas, están presentes en un sin número de nuevas monedas virtuales, que a diario compiten por alcanzar índices de aceptabilidad parecidos o iguales a bitcoin, invertir en Ethereum por ejemplo, es considerado como una buena idea de negociación, ya que bancos de todo el mundo lo están haciendo. Otro ejemplo práctico, se presenta en España, con Bankia, Santander, Abanca y otros, haciendo smart-contracts, a la par de bitcoin, con lo cual, es muy probable que Ethereum, siga subiendo de igual manera [15].

Parametrizar la consistencia y liquidez de las monedas virtuales, a la fecha no es posible, ya que presentan un alto grado de inestabilidad y en concordancia al dinero corriente, no esTable.cen lineamientos y normas claras que regulen su comportamiento y distribución. Así por ejemplo, el éxito del bitcoin está condicionado por tres grandes obstáculos: su escasa implementación, la falta de estabilidad de precios y la falta de confianza. El primer obstáculo es salvable, ya que es el resultado del éxito de la moneda. Sin embargo, los problemas de la estabilidad de precio y la confianza, son más difíciles de solventar [16]. Al respecto, [16], para descifrar este problema, especifica que: Según la mayoría de manuales de economía, toda moneda debe poseer tres atributos básicos: ser un medio de intercambio, una unidad de cuenta y un contenedor de valor.

REFERENCIAS

[1] *M. Castells, La era de la información. Economía, sociedad y cultura. La sociedad, México: red. volumen 1. Siglo xxi editores, 1999.*

[2] *C. L. M. Céspedes y M. F. J. Martin, The Latin American Experience, I4D. Vol. VII, núm. 9. Enero-marzo, 2010.*

[3] *Minitic., «MinTIC le cumple a Colombia: 98% de municipios conectados a Internet y 28 millones de conexiones,» 16 10 2017. [En línea]. Available: http://www.mintic.gov.co/portal/604/w3-article-61094.html.*

[4] *Á. Bernardo, « Bitcoin es legal: ¿verdadero o falso?.,» Hipertextual [en línea], 27 marzo 2013. [En línea]. Available: Disponible en: https://hipertextual.com/2013/03/bitcoin-es-legal. [Último acceso: 7 junio 2018].*

[5] *A. Fidias, El Proyecto de Investigación, Introducción a la metodología científica, Caracas - República Bolivariana de Venezuela: EDITORIAL EPISTEME, C.A, 2012.*

[6] *C. y. R. B. Marshall, Design qualitativeresearch, California. Sage. USA, 1989.*

[7] *K. D. Rosero Paredes, Uso de criptomonedas como medio de pago de las Mymes exportadoras, Guayaquil, 2017.*

[8] *M. Rodriguez Garagorri, Principales criptomonedas., 2017.*

[9] *M. Erbiti, Tendencias de evolución hacia un nuevo paradigma, 2015.*

[10] *D. Yermack, Is bitcoin a real currency?, 2014.*

[11] *C. A. T. R. Cavello Max, Las Criptomonedas: Una estrategia de inversión, Lima: Universidad Esan, 2018.*

[12] *M. Montoya, «Inversión en bitcoins: modelo para Costa Rica.,» Repositorio del Instituto Tecnológico de Costa Rica, 2016. [En línea]. Available: Recuperado de: https://repositoriotec.tec.ac.cr/bitstream/handle/2238/6548/Inversion_bitcoins.pdf?sequence=1&isAllowed=y. [Último acceso: 3 junio 2018].*

[13] *L. García, «telemundo.com.,» 2 5 2017. [En línea]. Available: Obtenido de http://www.telemundo.com/noticias/2017/05/02/bitcoin-la-moneda-demas-de-1400-dolares-que-esta-cambiando-al-mundo.*

[14] *R. Caetano, «Learning Bitcoin,» de Birmingham, UK: Packt publishing., 2015.*

[15] *L. Javier., «Bitcoin y Ethereum, ¿inversión o burbuja? Inversores que han puesto mucho dinero, inversores que jamás lo harán,» xataka, 24 enero 2018. [En línea]. Available: https://www.xataka.com/empresas-yeconomia/bitcoin-y-ethereum-inversion-o-burbuja-inversores-que-hanpuesto-mucho-dinero-inversores-que-jamas-lo-haran. [Último acceso: 7 06 2018].*

[16] *G. Cullerés, «El bitcoin y su posible impacto en los mercados.,» 10 2014. [En línea]. Available: https://www.iefweb.org/es/finanzas/visordocumentospdf/41.*

Biografía Autor(es)

a. Magíster en Gestión de la Tecnología Educativa. Docente, SENA Cauca. E-mail: rbenavidesb@misena.edu.co. Colombia.

b. Magister en Sistemas Integrados de Gestión. Centro Internacional de Producción Limpia Lope, SENA Regional Nariño. E-mail: eliguerrero@misena.edu.co. Colombia.

El portafolio digital como medio para promover la identidad digital académica y profesional en estudiantes universitarios

Berenice Castillejos López (a)
Carlos Arturo Torres Gastelú (b)

Resumen

Los estudiantes universitarios se enfrentan a una serie de desafíos. En el plano tecnológico, la participación en la red implica dejar una huella digital por todos los espacios de la web. Desde la investigación acción participante, este trabajo tuvo como propósito analizar el uso de los portafolios digitales como medio para gestionar la identidad digital académica y profesional en estudiantes universitarios. Los sujetos participantes, estudiantes del octavo semestre de la licenciatura en Administración Turística de la Universidad del Mar, ubicada en el estado de Oaxaca, México. Las técnicas de investigación que se emplearon fueron la observación participante, entrevista grupal y la entrevista individual semiestructurada. Dentro de los hallazgos se obtuvo que los estudiantes no estaban familiarizados con el blog y la cuenta de Twitter para fines de aprendizaje. Esta última herramienta resultó ser un recurso estratégico para promover la identidad digital profesional. Como conclusiones se obtuvo que los portafolios digitales resultan ser un elemento clave en el aula para gestionar la identidad digital académica y profesional de los universitarios, Aunque existen barreras y desafíos como las competencias digitales, la conexión a internet, la disponibilidad de tiempo, así como la autorregulación y la motivación para seguir desarrollando su perfil después de concluir la experiencia educativa.

Palabras clave: Identidad digital, competencias digitales, millennials.

The digital portfolio as a medium to promote identity academic and professional digital in college students

Abtract

College students face a number of challenges. At the technological level, participation in the network implies leaving a digital footprint in all spaces on the web. From the participant action research, this work aimed to analyze the use of digital portfolios as a means to manage academic and professional digital identity in university students. The participating subjects, students of the eighth semester of the Tourism Administration degree at the Universidad del Mar, located in the state of Oaxaca, Mexico. The research techniques used were participant observation, group interview, and semi-structured individual interview. Among the results it was obtained that the students were not familiar with the blog and the Twitter account for learning purposes. This last tool turned out to be a strategic resource for professional digital identity. As conclusions, it was obtained that digital portfolios turn out to be a key element in the classroom to manage the academic and professional digital identity of university students, although there are barriers and challenges such as digital skills, internet connection, time availability, as well as self-regulation and motivation to continue developing their profile after completing the educational experience.

Keywords: Digital identity, digital skills, millennials.

I. INTRODUCCIÓN

1.1 Antecedentes

Los estudiantes se enfrentan a desafíos económicos, sociales, culturales, ambientales y tecnológicos, reflejo de los efectos de la globalización, crisis financiera, demandas de la sociedad actual, la aguerrida competencia en el mercado laboral, el cambio climático, así como la falta de una cultura de buenas prácticas en el uso de internet. Es importante resaltar que, a pesar de considerarlos la cohorte demográfica que creció con el desarrollo de la web, tal segmento presenta una falta de homogeneidad en su nivel de competencias digitales [1, 2, 3, 4].

El uso de la web en actividades del mundo académico y laboral ha motivado el desarrollo de nuevas formas de gestionar información y mecanismos de comunicación. Cuando se analiza este aspecto, es importante considerar lo expuesto por Castells [5], quien identifica tres tipos de comunicación: la interpersonal, de masas y la autocomunicación de masas. Este último surge a partir de la transformación de la web, se caracteriza por ser interactiva, puede llegar a una audiencia global por el uso de las herramientas digitales, el mismo usuario genera el mensaje, así también define a la comunidad receptora y selecciona qué tipo de mensajes o contenidos digitales quiere recuperar.

Ahora bien, preparar al futuro profesionista no sólo involucra el desarrollo de sus competencias profesionales, sino también la gestión de una marca personal y especialmente la construcción de una identidad digital profesional [6]. La valoración de esta última, lleva consigo una autorreflexión sobre la huella digital. El estudiante debe estar consciente que la información que comparte en medios digitales tendrá un impacto en su marca personal. Por lo anterior, el estudio tuvo como finalidad analizar el uso de los portafolios digitales como medio para gestionar la identidad digital académica y profesional en estudiantes universitarios.

1.2 La identidad digital de los estudiantes universitarios

Desde un plano general, el diccionario de la Real Academia Española, define a la identidad como el "conjunto de rasgos propios de un individuo o de una colectividad que los caracterizan frente a los demás" [7]. En el caso específico de la identidad digital, se conceptualiza como el "conjunto de rasgos diferenciadores que define a personas, marcas comerciales o comunidades y les caracterizan" [6]. Ambos conceptos coinciden en la identificación de aspectos diferenciadores que describen a una persona o grupo.

En otro estudio [8] se señala que la identidad digital no fue diseñada para ser compartida y por tal hecho es única. En este ámbito se conceptualiza como un conjunto de características particulares que permiten al usuario autentificarse ante algún servicio digital. Desde el plano académico y profesional, gestionar la identidad digital en estudiantes universitarios invita a reflexionar sobre el impacto que produce trabajar con su huella digital antes de salir al campo laboral.

Para analizar a profundad este aspecto es importante considerar primero las diferencias entre la identidad digital y la marca personal. Este último concepto refiere a todas aquellas acciones que realiza un individuo para promover su carrera profesional, mientras

que el perfil digital se orienta a valorar el rastro que dejan todas las interacciones que se tienen en la red [9].

Cuando se reflexiona sobre el tema de la identidad digital, se le identifica tres aspectos: a) lo que no está disponible en internet no existe; b) una vez que algo aparece en internet, éste deja huellas digitales permanentes; c) el yo real recibe un adicional, la capa transparente digitalizada que acompaña a su vida real, la parte del yo, no puede evitar tomar decisiones con referencia a esa identidad digital [10].

Por su parte, otros autores [11] señalan que al trabajar con la identidad digital se deben considerar tres preguntas clave: ¿Qué enseño en la red sobre mí mismo(a)? ¿Quién influye en mí, en la forma de pensar y actuar en la red? ¿Quién se ve influenciado(a) por mí?

En otros planos, desde el *crowdsourcing* (colaboración masiva y abierta) se ha analizado la práctica existencial de crear identidad a través del uso de herramientas digitales [12]. El ser, convertirse en y pertenecer a son considerados tres puntos reflexivos en las prácticas sociales vitales para existir como alguien.

Ahora bien, la identidad de un individuo se compone de la parte analógica (entorno físico) y la digital. Estos dos elementos son el resultado de una identidad digital híbrida, aspecto que se traduce en dos realidades, el mundo presencial y la virtualidad. Diversos autores [13] señalan que la identidad digital es todo lo que se comunica y la forma en que se interactúa en el ciberespacio. Un perfil digital híbrido bien gestionado contribuye al desarrollo profesional del usuario.

En tanto, otros autores [14] sostienen que la cultura digital ha transformado la forma en que se estructura la concepción del yo. El internet rompe con los esquemas tradicionales de comunicación y permite que cualquier persona con acceso a la red tenga la posibilidad de ser vista y escuchada.

Es importante comentar que cuando se analiza el perfil digital del internauta, no sólo se identifica lo que produce el usuario sino también el de sus contactos, personas que pueden ir desde familia, amigos, conocidos, desconocidos que comparten alguna afinidad y en ciertos casos expertos y organizaciones que producen conocimiento de alguna disciplina de interés [15].

Ahora bien, dentro las competencias digitales, se ubica la gestión de la identidad digital, tal competencia demanda tener conocimiento sobre los beneficios de gestionar un perfil digital, así como las formas en que se crean interrelaciones en la red. Conlleva tener nociones sobre los agentes que pueden aportar o afectar la identidad. Implica desarrollar habilidades para proteger la imagen ante posibles amenazas. Asimismo, la capacidad de construir un perfil o perfiles que no sólo respondan a las necesidades personales sino también a la parte profesional. De igual forma poder generar rastreos de la huella digital y valorar las ventajas y riesgos de la información que se expone. Por último, asegurarse que la información que se comparte no pone en riesgo la integridad como usuario, además de tener la facilidad de aplicar diferentes formas de expresarse en la red [16, 17].

Se ha argumentado que una gran proporción de la sociedad tienen conocimiento que cuentan con una identidad digital pero muy pocos tienen nociones del impacto que tienen sus datos desde marco legal y comercial. El individuo, el gobierno y las empresas privadas juegan un determinado rol en la llamada ciudadanía digital [18].

Como medio de registro, los servicios digitales normalmente solicitan a los usuarios información personal como nombre, edad, género, profesión, estado civil, teléfono y correo electrónico. Ya en posición de usuario registrado, hay internautas que exponen

de manera pública relaciones con sus allegados (familiares y amigos) y dan acceso a toda la información que generan o comparten. Otros aplican criterios de privacidad, manteniendo de manera abierta información básica que no ponga en riesgo su seguridad. Los contenidos que se llegan a compartir se traducen en textos, imágenes, audios y videos [15].

Otros autores [19] señalan que los egresados universitarios en ocasiones no son conscientes del impacto que produce su huella digital. Una buena gestión de la identidad digital puede significar que un individuo obtenga el trabajo o no. Argumenta que una minoría emplea estas herramientas para proyectar su perfil profesional. Por lo anterior, es importante que las instituciones educativas promuevan estrategias y herramientas para la construcción de la identidad digital profesional.

Ahora bien, a partir de la propuesta de González [20], que consiste en la adaptación del modelo de negocios Canvas para gestionar la identidad digital, Borras [21] perfecciona esta metodología para desarrollar la identidad digital académica y profesional, en función de dos etapas. La primera orientada al diagnóstico actual del perfil del usuario en la red, está compuesta de cinco apartados: actividad, comunidad, enfoque, privacidad y comunicación. En tanto, la segunda se basa en la elaboración de un plan de identidad digital, la plantilla contiene nueve apartados: colaboradores, actividades clave, propuestas de valor, objetivos, público objetivo, recursos clave, canales, costos y contenidos.

Por lo antes expuesto, promover la gestión de la identidad digital dentro de la comunidad universitaria se convierte en un compromiso que busca visibilizar el conocimiento de una forma segura y eficaz. Los profesores investigadores y estudiantes deberían adoptar entornos personales de aprendizaje (PLE, de *personal learning environment)* en la red con fines de divulgación y colaboración [22].

II. METODOLOGÍA

El método empleado para la investigación fue de tipo cualitativo y centrado en la investigación acción participativa. Los sujetos participantes, universitarios de la licenciatura en Administración Turística de la Universidad del Mar. Las técnicas de investigación que se emplearon fueron la observación participante, entrevista grupal y la entrevista individual semiestructurada. Los 68 participantes fueron estudiantes del octavo semestre, quienes construyeron su portafolio digital durante el periodo marzo-junio del 2017.

Con el objeto de realizar la fase de diagnóstico y planeación, se utilizó la herramienta adaptada del modelo Canvas, que consistió en el empleo de la plantilla para el análisis del perfil del usuario y la del plan de identidad digital académico y profesional [21]. Esta práctica de observación se acompañó de dos entrevistas grupales, debido a que los sujetos participantes estaban divididos en dos grupos.

La plantilla Canvas para el análisis de la identidad digital fue empleada como guía de observación para la fase diagnóstica en esta etapa se analizaron cinco elementos: la actividad, la comunidad, el enfoque, la privacidad y la comunicación. En lo que refiere a la planeación, se empleó la segunda plantilla Canvas que involucró ocho aspectos: colaboradores, actividades clave, propuestas de valor, objetivos, público objetivo, recursos clave, canales, costos y contenidos.

Para la fase de implementación, cada estudiante trabajó con un blog en Blogger y una cuenta de Twitter, ambas herramientas constituyeron lo que se le denominó portafolio

digital. Estos recursos fueron empleados en dos experiencias educativas cursadas en el semestre. Para efectos de la construcción del espacio de aprendizaje, los estudiantes recibieron asesoría de los dos docentes y de un técnico en informática.

Durante el tiempo que cursaron las experiencias educativas en el semestre, los estudiantes desarrollaron sus entradas de blog con temas vinculadas a sus materias, de igual forma participaban y compartían en Twitter información relacionada con su formación profesional y temas de su interés.

Por último, en lo que refiere a la evaluación del portafolio digital, espacio donde trabajaron su identidad digital académica y profesional, se aplicaron 68 entrevistas individuales semiestructuradas. Esto se realizó cuando finalizó el semestre y concluyeron las experiencias educativas cursadas. Es importante mencionar que, para el análisis e interpretación de la información se aplicó análisis de contenido.

III. RESULTADOS

3.1 El diagnóstico – análisis de la identidad digital

A partir del modelo canvas adaptado para el análisis de la identidad digital académica y profesional (Borras, 2016) se obtuvieron cinco categorías: actividad (ACT), comunidad (COMU), enfoque (ENFO), privacidad (PRIV) y comunicación (COMN). Respecto al primer aspecto (ACT), se identificó que las actividades cotidianas estaban relacionadas con el uso de las redes sociales como Facebook e Instagram, la mensajería instantánea, la búsqueda de información, escuchar música, así como ver películas y videos en línea (Ver figura 1).

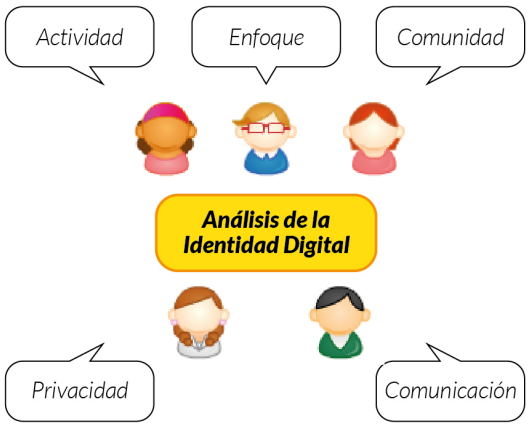


Fig. 1. *Análisis de la identidad digital. (Elaborada a partir de la información de Borras, 2016).*

En lo referente a la comunidad (COMU), las personas u organizaciones con las que se relacionaron los estudiantes, estuvo representado por su círculo de amigos, familia, amigos de sus contactos y organizaciones vinculadas a su área de formación. Estos vínculos fueron más de carácter social que de tipo académico o profesional.

En la categoría de enfoque (ENFO) se observó que la mayoría poseía una sola identidad digital, aunque se identificaron casos que argumentaron tener más de dos identidades,

justificando que era por cuestiones de seguridad. Para abordar el tema de enfoque, se les invitó a ingresar su nombre en el buscador de *Google*, una acción muy cotidiana en estos días, que se le ha denominado *googlear*. Los hallazgos de tal ejercicio los llevó a detectar los espacios más posicionados, tales como su cuenta de Facebook, la lista de ingreso a la universidad, entre otros aspectos. Los estudiantes reflexionaron sobre la huella que dejan en la red.

Del tema de la privacidad (PRIV) se identificó que hay estudiantes que aplican criterios de privacidad en sus redes sociales, dejando disponible sólo cierta información. Se observaron también casos que mantenían en abierto toda su información. En este punto se consideró el tema del etiquetado, situación que para algunos, en ocasiones resultaba ser desagradable, ya que sin permiso, terceras personas compartían imágenes en su espacio personal.

Por último, en el aspecto comunicación (COMN), se observó que el teléfono inteligente o *Smartphone* es uno de los dispositivos que los conecta con la red la mayor parte del tiempo. Algunos testimonios revelaron que lo mantenían encendido las 24 horas del día. Cabe señalar que se identificaron casos que no disponían de internet en el hogar y sólo se conectaban desde la sala de cómputo de la universidad a determinadas horas. Es relevante señalar que uno de los factores era el horario de clases, es decir, de tiempo completo (8:00-13:00 y de 16:00-19:00 horas) lo que los limitó a conectarse a determinadas horas. Otro aspecto a considerar fue el idioma en el que principalmente se comunicaban, en este caso el castellano, sólo una minoría externó que podían entablar comunicación en inglés y contaban con la facilidad de intercambiar experiencias con personas de otros países.

3.2 La planeación – el portafolio para gestionar la identidad digital

Los hallazgos del análisis de la identidad digital sentaron las bases para la planificación de la actividad. En esta fase se empleó la plantilla del plan digital académico y profesional [21]. En esta etapa se identificaron las siguientes categorías: colaboradores (COLS), actividades clave (ACTC), propuestas de valor (PROV), objetivos (OBJE), público objetivo (PUOB), recursos clave (RECL), canales (CANA), costos (COST) y contenidos (CONT) (ver figura 2).

Fig. 2. *Planeación de la identidad digital. (Elaborada a partir de la información de Borras, 2016).*

Respecto al primer factor, denominado colaboradores (COLS), los docentes de dos asignaturas y un técnico en informática fungieron como facilitadores para la construcción de la identidad digital académica y profesional. Los profesores otorgaron asesoramiento en la estructuración del portafolio digital y la conformación de las redes personales de aprendizaje. En el caso del personal de informática, su función se centró en dar soporte técnico en el diseño del blog y la configuración de ciertas herramientas de la web.

Las actividades clave (ACTC) estuvieron vinculadas con dos experiencias educativas. En tales asignaturas se inició con una charla de entornos personales de aprendizaje (PLE de *Personal Learning Environment*), así como del binomio identidad digitalmarca personal, seguida de la construcción del portafolio electrónico, que incluía la creación del blog y la cuenta de Twitter. Es importante comentar que cada docente contaba con su portafolio de aula de su respectiva asignatura.

La propuesta de valor (PROV) implicó la formación de redes personales de aprendizaje, es decir, compartir información y conocimiento con terceros, así como identificar eventos turísticos, ofertas de apoyos para proyectos, becas, estancias, cursos de formación, entre otros aspectos. En términos generales, un espacio de difusión para comenzar a esTable.cer vínculos con el medio laboral y académico relacionados con el turismo y otros tópicos de interés.

Los objetivos (OBJE) de la construcción del perfil estuvieron orientados a desarrollar una identidad digital académica profesional que permitiera a los estudiantes comenzar a trabajar con su huella digital. Tal actividad sentó las bases para que los estudiantes a posteriori tuvieran intenciones de seguir gestionando su identidad digital académica y profesional en la red.

El público objetivo (PUOB) estuvo integrado principalmente por pares de administración turística, profesores investigadores, organizaciones públicas y privadas dedicadas a la actividad turística, así como aquellos contactos relacionados con los temas de viajes, administración, mercadotecnia, cultura, medio ambiente, gastronomía, economía, música, entre otros.

Como recursos claves (RECL) se identificó en primer lugar al factor tiempo, debido a que los participantes estudian bajo un sistema de tiempo completo. Por tal hecho deberían considerar administrar bien sus horas de trabajo. Otro aspecto fueron las competencias digitales, la gran mayoría tuvo el desafío de comenzar a trabajar con un blog y usar la cuenta de Twitter. En la parte de recursos tecnológicos, contar con una buena conexión de internet, implicó valorar los diferentes espacios (escuela, hogar y los paquetes de datos del teléfono inteligente), asimismo considerar los equipos del que disponían los estudiantes.

Respecto a los canales de comunicación (CANA), la cuenta de Twitter desempeñaría un papel importante a los momentos de intercambiar experiencias con el exterior. Este medio de comunicación resultó útil, aunque en un inicio requirió que los estudiantes se familiarizaran con el menú de la herramienta. Este microblog no sólo funcionó como espacio de comunicación, sino también como fuente de gestión de información y producción de contenidos *(tweets)*.

Del factor costo (COST), es importante considerar que los recursos tecnológicos fueron factores determinantes para el desarrollo de la identidad digital. Al hablar de recursos, se identificó la necesidad contemplar costos para la conexión a internet, ya sea conexión *wifi* en el hogar y el paquete de datos en el dispositivo. Además de contemplar aquellos casos que requirieron hacer uso de un café internet para conectarse.

En lo que se refiere a contenidos (CONT), los blogs estuvieron orientados a compartir productos que se generaron en las dos experiencias educativas. Actividades académicas

que se desarrollaron bajo un esquema de aprendizaje basado en proyectos y casos. En la cuenta de Twitter se generó una discusión grupal, además que se compartieron las ligas de los contenidos que realizaban en el blog, así como artículos, noticias, eventos y demás recursos que enriquecían el tema que se abordaba en la materia.

Como resumen de la experiencia, se observó que los participantes poseían nociones básicas sobre el uso de las herramientas digitales con fines de aprendizaje. Una gran proporción utilizó por primera vez *Blogger y Twitter*. Se identificaron pocos casos que trabajaron más a detalle la construcción del perfil, realizando actividades que superaron lo agendado por los docentes. En lo que refiere al blog, se trabajó con material de dos experiencias educativas. Respecto al microblog, tal herramienta representó un desafío al momento de conectar con otras personas y compartir información relacionada con su área profesional. Al menos un 15% de los participantes no disponía de internet en el hogar y por tanto, las actividades las desarrollaron desde las salas de cómputo de la universidad, en casa de amigos o desde un café internet.

3.3 La retroalimentación – Identidad digital professional-portafolio digital

Dentro de los hallazgos de la retroalimentación se identificaron seis categorías: uso de herramientas digitales (UHED), autovaloración de lo que se comparte (AUTC), la influencia social (INSO), autoevaluación del desarrollo del portafolio (AVHE), intenciones futuras (INFU), barreras y desafíos del portafolio (BADE).

3.3.1 Uso de herramientas digitales (UHED)

Para efectos de la gestión de la información se identificaron herramientas digitales tales como el buscador genérico Google, así como sus temáticos como Google Académico, Google Books y Google Maps. También se observó el uso de Bing, YouTube, la página de la biblioteca de la universidad, el Centro de Documentación Turística de la escuela, bibliotecas virtuales, bases de datos electrónicas, e-books, Instagram, Scribd, información de blogs, Twitter, Facebook, Canva, Pinterest y páginas oficiales relacionadas con la formación. Para el almacenamiento y recuperación de archivos comentaron que usan Google drive, Box y Gmail principalmente.

Respecto a la comunicación y colaboración con terceros, ya sea en ambientes formales o informales de aprendizaje emplean: Facebook, Twitter, Blogger, Box, WhatsApp, Instagram, Youtube, Cacoo, Scribd, Google drive, correo electrónico como el caso de Gmail, Outlook, Skype, Google Maps y el blog en Blogger.

Por otra parte, crean y comparten contenidos empleando la paquetería de Microsoft Office, Blogger, Twitter, Instagram, YouTube, Facebook, Canva, Cacoo, Scribd, Picktochart, Googlemaps, Photoshop, Mindmap, Pinterest, Mindjet y Photo Studio. Los tipos de contenidos incluyeron texto, imágenes, videos, infografías, y representaciones gráficas.

3.3.2 Autovaloración de lo que se comparte (AUTC)

Los estudiantes argumentaron que en su portafolio digital muestran información útil y relevante, es material de interés para la comunidad estudiantil de la licenciatura

en Administración Turística. Se comparten además noticias y tópicos sobre Bahías de Huatulco. Tanto el blog como la cuenta de Twitter resultan ser espacios que les ofrece visibilidad, ya que se identifican como universitarios. También les permite mantenerse en comunicación con la comunidad vinculada con el turismo y los viajes. Además comparten libros favoritos, logros escolares, tópicos diversos relacionados con el medio ambiente, la economía, la política, la música, por mencionar algunos. Externaron también que evitan proporcionar datos personales sensibles (teléfonos, ubicación, entre otros).

3.3.3 La influencia social (INSO)

Las personas u organizaciones que influyen en la forma de pensar y actuar en la red están relacionadas con las cuentas de docentes, compañeros de clase, familiares, amigos, personajes políticos y organismos oficiales. De igual forma profesionales dedicados al mercado de viajes y servicios turísticos, *Youtubers* viajeros, así como cuentas relacionadas con la literatura.

En el caso de poder influenciar en otros se ubican los amigos, compañeros de clase, maestros, familiares, turistas potenciales, prestadores de servicios turísticos, especialistas en el ramo. Se observó en las cuentas de Twitter que, a partir del acto de comenzar a seguir, se van obteniendo algunos seguidores.

3.3.4 Autoevaluación del desarrollo del portafolio (AVHE)

Una gran proporción de los entrevistados externaron que trabajar con el portafolio digital fue satisfactorio, interesante, divertido, práctico y útil. Se identificaron casos extremos que iban desde aquellos estudiantes que justificaron que su blog se quedó en estados incipientes con algunas deficiencias en la producción del contenido, hasta aquellos que realizaron modificaciones de diseño y emplearon diversas herramientas digitales, más allá de lo exigido en clase. La mayoría identificó que a partir de las actividades agendadas en el blog, ciertas áreas de sus competencias digitales mejoraron (gestión de información, comunicación y colaboración, así como la creación de contenido digital).

Los estudiantes expresaron que las actividades desarrolladas en el portafolio digital permitieron comunicarse y compartir con terceros, de igual forma esto motivó a revisar la información que se comparte. Aplicar la estrategia de aprendizaje en el blog y Twitter permitió mostrar al exterior lo que se hace en el aula y de esta forma darse a conocer en el mundo laboral.

Trabajar con estas dos herramientas digitales permitió aplicar el pensamiento crítico y creativo, de igual forma promover la identidad digital académica y profesional. Se identificaron entrevistados que consideraron que era importante que sus espacios ofrecieran información confiable y relevante.

3.3.5 Intenciones futuras (INFU)

Una gran mayoría de estudiantes externó el interés de seguir usando las herramientas digitales, principalmente Twitter. En el caso del Blog, existen algunos

entrevistados que argumentan que ya no lo seguirán usando, otros expresaron que lo modificarán y publicarán diversos tópicos de interés y posiblemente éste se incluya en otras materias. Una estudiante señaló tener intenciones de ligar su cuenta de Facebook a su portafolio digital. Otro entrevistado identificó la importancia de fortalecer su identidad digital.

En lo que respecta a Twitter, se identificaron casos que tienen intenciones de recomendar el uso de esta herramienta a pares o conocidos. Otros, en cambio, buscarán incrementar su red personal de aprendizaje y compartir en red información sobre viajes y turismo.

3.3.6 Barreras y desafíos del portafolio (BADE)

Dentro de las barreras y desafíos para el uso del portafolio se ubicaron algunos aspectos tales como la administración del tiempo, el manejo de las herramientas digitales (blog y Twitter), la conexión a internet, la falta de equipo, el filtrado y evaluación de la información, la redacción y ortografía, el pensamiento creativo, el desconocimiento de la gestión de la identidad digital y seguridad en la red, así como la falta de interés. Cabe señalar que existieron casos que externaron no haber percibido problemas con el desarrollo de su portafolio digital.

IV. DISCUSIÓN Y CONCLUSIONES

Hoy en día, atender el tema de identidad digital en ambientes formales de aprendizaje se convierte en una necesidad de formación. Como se mencionaba en líneas anteriores, no basta con desarrollar competencias profesionales vinculadas a su área disciplinar. Se reitera que es necesario educar a la generación de los *millennials* como ciudadanos cultos y prepararlos para participar de forma responsable en el ciberespacio [1].

Estudios previos argumentan que todo lo que hay en la red sobre un individuo, comunica y forma parte de su huella digital [6, 21]. Responder a las tres preguntas claves sobre lo que enseñamos en la red sobre nosotros, los actores que influyen en la persona respecto a la forma de pensar y actuar en la red, así como quienes se ven influenciados por lo que hacemos nosotros [11] invita a reflexionar sobre la cantidad y tipo de información que se comparte en la red, además de los contactos que se generan al momento de seguir y viceversa. Por tanto lo que ocurre en la red se forma a partir de la autocomunicación de masas [5].

La huella que dejan los estudiantes en la red, tiene un impacto en su desarrollo profesional. La red es una ventana al mundo para que ellos puedan darse a conocer. Este aspecto es relevante para los jóvenes universitarios, dado que el imperativo actual demanda que lo que no está disponible en internet no existe [10].

En este sentido, la colaboración masiva y abierta en estudiantes universitarios conduce a identificar dos mundos: lo real y lo virtual [12, 14]. En este último espacio, los jóvenes tienen la posibilidad de que sus actividades de aprendizaje sean vistos por terceros. En este sentido, la red de contactos va más allá de familiares, amigos, profesores y compañeros de clase [15].

Al considerar a la identidad digital, una competencia digital, ésta debe ser promovida en el aula como una habilidad transversal. Al parecer muchos saben que están dejando una huella en la red pero, muy pocos están conscientes de los aspectos legales y comerciales que involucra la gestión de tales datos [18].

La metodología de González [6] y Borras [21] ofrece recursos útiles para todo aquel individuo que busque gestionar su identidad digital profesionalmente. Es relevante también aclarar que el uso del blog y Twitter resultan ser dos herramientas digitales básicas para el desarrollo de una identidad digital académica y profesional. Esta recomendación no exenta el uso de otras aplicaciones, sino tales elementos se vuelven necesarios al momento de gestionar el portafolio digital, con miras a traspasar las barreras de la educación formal.

Comenzar a promover interacciones entre el estudiante, el mercado laboral y el ámbito profesional invita a considerar que el futuro egresado no tiene que esperarse a terminar la universidad para comenzar a compartir información y por ende, conocimiento en su campo de acción.

Para finalizar, los portafolios digitales resultan ser un elemento clave en el aula al momento de gestionar la identidad digital académica y profesional de los universitarios. Aunque también, cabe señalar que las competencias digitales, la conexión a internet, la disponibilidad de tiempo, así como la autorregulación y la motivación determinan el interés de seguir desarrollando su entorno de aprendizaje, después de concluir la experiencia educativa.

REFERENCIAS

[1] *Area, M.M., Borrás, M., J. F. y San Nicolás, S. B. Educar a la generación de los Millennials como ciudadanos cultos del ciberespacio. Apuntes para la alfabetización digital. Revista Estudios de Juventud, 109, 13-32. (2015).*

[2] *Howe, N. y Strauss, W. Millennials rising: The next great generation. Nueva York: Vintage. (2000)*

[3] *Pedró, F. The new millennium learners: challenging our views on ICT and learning. Paris: OECD-CERI. (2006).*

[4] *Romo, G. J. y Tarango, J. Factores sociodemográficos, educativos y tecnológicos en estadios iniciales de cibercultura en comunidades universitarias. Apertura, Revista de Innovación Educativa, 7(2), 1-15. (2015).*

[5] *Castells, M. Comunicación y Poder. Madrid: Alianza Editorial. (2009).*

[6] *González, D. Educando la identidad digital en el aula. https://www.slideshare.net/diana.gonzalez/educando-la-identidad-digital-en-el-aula (2016)*

[7] *Real Academia Española. Diccionario de la Lengua Española. (2019).*

[8] *Birch, D. Digital Identity Management: Technological, Business and Social Implications. New York: Routledge. (2016).*

[9] *Muñoz, M. Por qué trabajar tu marca personal e identidad digital. Neurona digital. http://neuronadigital.es/por-que-trabajar-tu-marca-personal-e-identidad-digital/ (2015).*

[10] *Feher, K. (2015). Digital Identity: The Transparency of the Self. En: Jane Montague, Tan Lee Ming (eds.) Applied Psychology. 2015 Asian Congress of Applied Psychology. Singapore: World Scientific, Paper 46. (2015).*

[11] *Castañeda, L. y Camacho, M. (2012). Develando nuestra identidad digital. El profesional de la información, 21 (4), 354-360. (2012).*

[12] *Hällgren, C. Crowdsourcing identities: On identity as an existential practice mediated by contemporary digital technology. First Monday, 24(1): 8112 doi: 10.5210/fm.v24i1.8112 (2019).*

[13] *Aparici, R. y Osuna, A.S. La cultura de la participación. Revista Mediterránea de Comunicación, 4 (2) 137-148. doi: 10.14198/MEDCOM2013.4.2.07 (2013).*

[14] *Hildebrandt, K., y Couros, A. Digital selves, digital scholars: Theorising academic identity in online spaces. Journal of Applied Social Theory, 1(1). (2016).*

[15] *Laurent, M. y Bouzefrane, S. Digital Identify Management. ISTE Press-Elsevier. (2015).*

[16] *Ferrari, A. DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe. http://ipts.jrc.ec.europa.eu/publications /pub.cfm?id=6359 (2013).*

[17] *INTEF. Marco común de competencia digital docente. http://educalab.es/documents/10180/12809/MarcoComunCompeDigiDoceV2.pdf (2017).*

[18] *Sullivan, C. Digital identity – From emergent legal concept to new reality. Computer Law & Security Review. 34(4), 723-731. https://doi.org/10.1016/j.clsr.2018.05.015 (2018).*

[19] *Valerio, U.G., Castro, L. H. y Herrera, M. D.J. El rol de la identidad digital de los universitarios en el ingreso a la vida laboral. Revista Espacios, 7(16). (2016).*

[20] *González, D. Canvas de identidad digital. https://www.slideshare.net/diana.gonzalez/canvas-identidad-digital (2015).*

[21] *Borrás, G. O. Plan de identidad digital (académico y profesional). https://es.slideshare.net/oriolupm/plan-de-identidad-digital-acadmico-y-profesional-60283071 (2016).*

[22] *Tena, M.A., García-Holgado, A., Merlo, J.-A., García-Peñalvo, F.J. (2017). Diseño de un plan de visibilidad científica e identidad digital para los investigadores de la Universidad de Guadalajara (México). Ibersid.11 (1), 83-92.*

Biografía Autor(es)

a. Doctora en Sistemas y Ambientes Educativos, Universidad del Mar Campus Huatulco. E-mail: berenicecas@huatulco.umar.mx. México.

b. Doctor en Ciencias de la Administración. Universidad Veracruzana. E-mail: ctorres@uv.mx. México.

El software como una herramienta didáctica de las matemáticas universitarias

Hector Alirio Guerrero (a)
Ricardo Javier Hernandez Revelo (b)

Resumen

A partir de esta investigación se pretende determinar la metodología de la enseñanza de la matemática numérica en los estudiantes de educación superior; como en el campo de la ingeniería, observando los errores conceptuales, metodológicos y la actitud que hacia la matemática presentan tanto estudiantes como profesores, teniendo en cuenta el apoyo sustancial de la disciplina. Buscando plantear la manera de cómo los estudiantes adoptan los fundamentos matemáticos y algebraicos para someterse a la solución de problemas que conllevan la aplicación de los métodos numéricos como un caso particular de las matemáticas aplicadas en la búsqueda de soluciones aproximadas del modelamiento matemático planteado para estos problemas.

Palabras clave: Matemática numérica, métodos numéricos, metodología, conocimientos matemáticos, teoría del error, modelamiento, solución de problemas.

Software as a teaching tool of mathematics university

Abstract

From this research is to determine the methodology of teaching numerical mathematics among students in higher education and in the field of engineering, observing the conceptual, methodological errors and attitude toward mathematics have both students and teachers as substantial support from the discipline. Looking to raise the way how students adopt the mathematical foundations and algebraic to undergo solving problems involving the application of numerical methods in finding approximate solutions of mathematical modeling posed for these problems.

Keywords: Numerical mathematics, numerical methods, methodology, mathematical knowledge, error theory, modeling, troubleshooting.

I. INTRODUCCIÓN

En los últimos años, cada vez que se pretende abordar un problema de la vida cotidiana, tanto los docentes como los estudiantes se ven envueltos en una problemática, de cómo aplicar e impartir los conocimientos de las matemáticas para dar la correcta solución a estos problemas, donde es casi imposible llegar a una solución exacta a través de los métodos tradicionales de los cálculos matemáticos, por esto es necesario llegar a la aplicación de los métodos numéricos, campo de la matemática numérica, los cuales ayudan a tener una o varias soluciones aproximadas para obtener una solución alternativa.

Con base a lo anterior, se debe tener en cuenta que la solución encontrada por medio de la aplicación de los métodos numéricos es una aproximación, pero el éxito de esta práctica es llegar a esta solución con un error muy pequeño, de tal manera que, al reemplazar en el modelo matemático del problema o fenómeno, sea muy cercano a la solución real. Para esto los docentes deben tener muy presente que la base de todo, son los fundamentos tanto teóricos de la matemática como los algebraicos; que sin los cuales es imposible tener éxito en el desarrollo de la solución a este tipo de problemas. Esto nos permite realizar pruebas diagnósticas para esTable.cer el nivel de conocimientos previos.

En consecuencia, cobra gran importancia la metodología del cómo enseñar y aprender a aplicar los métodos numéricos en la solución de problemas y en otras áreas afines en relación con las matemáticas universitarias. En este sentido se debe abordar o pensar en una forma didáctica, metodológica para llevar a los estudiantes al aprendizaje de los métodos numéricos en busca de la solución de problemas.

Por lo tanto, uno de los productos del resultado de la investigación fue la creación de un software didáctico denominado Thomas Bayes que permite aplicar los algoritmos de los métodos numéricos como el de bisección, punto fijo y Newton – Raphson, el cual nos permite ingresar la ecuación matemática, las condiciones iniciales para resolver la ecuación, analizar el comportamiento de la ecuación en forma gráfica, y de esta manera esTable.cer los requerimientos de cada algoritmo de los métodos mencionados, que como resultado de su ejecución se presentan las iteraciones necesarias para alcanzar la solución alternativa aproximada con un error de estimación acorde a las exigencias del tema de estudio.

II. METODOLOGÍA

2.1 Enseñanza de la matemática numérica

Pensar en la educación, prácticas educativas, metodologías de enseñanza-aprendizaje dentro del aula de clases y fuera de ella y aún más dentro de un marco investigativo de como impartir y apropiarse de la mejor forma las matemáticas numéricas y dentro de ellas los métodos numéricos como un apoyo en la resolución de problemas que surgen bajo una situación de contexto o el producto de un modelamiento matemático de una situación real o ficticia por el momento, como práctica de clases de la asignatura como tal o dentro de otro campo que no necesariamente sea matemáticas pero como fruto del análisis del problema surge una ecuación compleja de tratar.

En la actualidad la apropiación y construcción del conocimiento formal o científico tiene diferentes exigencias metodológicas y epistemológicas que han proporcionado un nuevo sentido e importancia de las matemáticas en el mundo actual y especialmente en el campo ingenieril, permitiendo así una reorientación de las mismas para que dejen de ser meras ilustraciones de los conocimientos transmitidos y pasen a soportar procesos algebraicos y aritméticos en el desarrollo o transformación del conocimiento mediante actividades investigativas, promoviendo estrategias didácticas que aproximen las actividades del aula a campos accionares de la disciplina.

Por otra parte, el uso del software como recurso didáctico en la búsqueda de nuevos métodos de enseñanza - aprendizaje, permite modificar las formas de aprendizaje especialmente en el campo numérico [1]. Haciéndolo más dinámico e interactivo y a la vez permite la rápida visualización de la solución de los problemas que se resuelven a través de los métodos numéricos. Siendo la visualización uno de los aspectos del proceso de enseñanza – aprendizaje, que según [2]: "describe la visualización como la habilidad para representar, transformar, generar, comunicar, documentar, reflejar información visual en el pensamiento y el lenguaje del que aprende". De modo que al realizar actividades de visualización se requiere de la utilización de nociones matemáticas asociadas a los ámbitos numéricos, gráficos, algebraicos o verbales para la comprensión de situaciones vivenciales dentro y fuera del aula de clases.

Teniendo en cuenta que dentro de los ámbitos numéricos se desenvuelven los métodos numéricos que bajo el estudio didáctico [3] se destacan los métodos de bisección, punto fijo y de Newton - Raphson, hablando de ellos desde su parte algebraica, algorítmica, analítica para el margen del error de cada uno de ellos y en lo posible estimar cuál de los métodos es más pertinente para la resolución de una ecuación no lineal de una sola variable real que puede resultar de un ejercicio teórico o del modelado de una situación problémica, pero también apoyándose con herramientas didácticas para su mejor análisis de la ecuación de interés.

Desde el punto de vista conceptual se procede a definir los métodos mencionados desde un punto más didáctico y acorde para los estudiantes de ingeniería o de otra disciplina tal como se describe en [3].

2.2 Importancia del Software Thomas Bayes

Tomando la definición de Software, se trata de la parte lógica de un máquina que ha sido modificada cada día para mejorar el procesamiento de la información para obtener una nueva; para ello se debe seguir unos pasos lógicos para llegar a una solución de problemas planteados y que pueden tener una aproximación de la solución, para ello se en marca en procesos iterativos los cuales exigen de tiempo en la manipulación de datos que conlleva a la solución aproximada de la ecuación, donde el software como herramienta didáctica toma gran importancia en el proceso de enseñanza aprendizaje de la matemática numérica.

Es aquí donde los diferentes algoritmos de los métodos numéricos como caso particular de las matemáticas aplicadas, en los que se destaca, como mecanismo de solución el uso elevado de cálculos repetidos y aproximaciones, conocidas como iteraciones. Por tal motivo, es útil usar el computador, ya que este permite solucionar ecuaciones complejas de manera rápida, y es por eso que al buscar herramientas computacionales que ayuden a obtener con gran rapidez y de forma gráfica la solución a las ecuaciones se encontró varias, pero entre ellas se destacan las siguientes: GEOGEBRA [4], que es un software matemático que integra la parte gráfica, algebraica, de las ecuaciones

utilizado como un procesador matemático y algebraico para representar y manipular las diferentes ecuaciones, como se muestra en Fig.1.

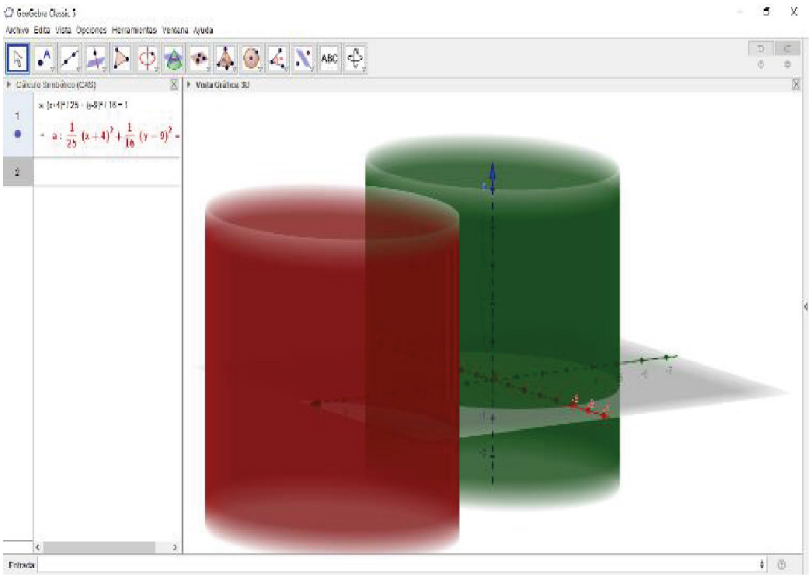


Fig. 1. *en 3D realizadas en Geogebra*

permitiendo el trabajo del algebra, cálculo y hasta inclusive un manejo de estimaciones para las decisiones empresariales, lo cual hace que sea posible implementar y desarrollar proyectos de física, estadística y procesos empresariales de estimación para la toma de decisiones, entre otros. También en el arte de descubrir nuevo software educativo para el tratamiento del proceso matemático y manipulación del cálculo se encuentra en línea una herramienta muy útil como es WOLFRAMALPHA, no es simplemente un buscador web, es mucho más y ofrece búsquedas en línea, puesto que hace procesamientos de información para entregar respuestas detalladas de matemáticas, tecnología científica, sociedad y cultura e incluso preguntas de la vida cotidiana, como se muestra en Fig.2., tomada de la página de wólfram Alpha [5].

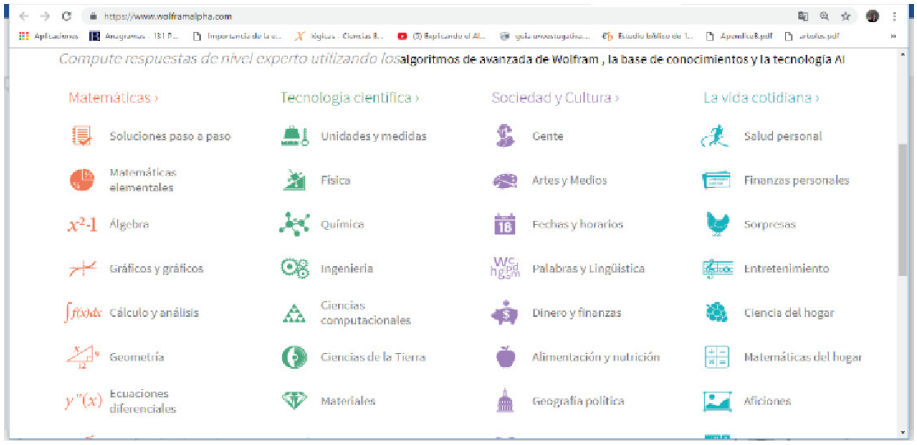


Fig. 2. *Servicios de WOLFRAMALPHA*

A diferencia del anterior hay software que permite de manera fácil ingresar las ecuaciones o incluso ya funciones con un grado de complejidad y de mucho tratamiento matemático para resolver. El software se encarga de resolver con gran rapidez, es por esto que dentro de la investigación se desarrolló un Software didáctico que involucra la parte gráfica y la parte analítica para los métodos de bisección, punto fijo y Newton - Raphson denominado software Thomas Bayes [3]. Donde toma gran importancia el proceso de la enseñanza aprendizaje de la interpretación y el modelamiento matemático de los fenómenos, ya sea de tipo natural o de eventos empresariales, los cuales demandan una cantidad de tiempo en procesos repetitivos. Además, esta herramienta muestra de manera intuitiva el trabajo con los diferentes métodos de solución numérica de ecuaciones no lineales, realizando un proceso integrado de graficación y tratamiento numérico para encontrar la solución aproximada a los modelos planteados, y por eso se la ha tomado también como una forma didáctica de explicar los métodos numéricos mencionados anteriormente.

La importancia del software desde la didáctica de las matemáticas le permite al docente hacer que los estudiantes visualicen con gran detalle las diferentes funciones, que de manera manual sea muy complicado realizarlas y permite que ellos realicen las gráficas de las funciones, observar su comportamiento por sectores de la curva, sobre todo cuando la curva se expande al infinito, además en que intervalo está la raíz de la ecuación de interés o función de estudio, donde para algunos métodos es necesario esTable.cer un intervalo apropiado para la convergencia exitosa del algoritmo. De igual manera se observa en la interfaz los requerimientos necesarios de cada método numérico con sus respectivas iteraciones y la raíz aproximada de acuerdo a su tolerancia.

Según lo anterior, se presenta la interfaz del menú principal del software Thomas Bayes que consta de los tres métodos de estudio como es el método de bisección, el método del punto fijo y el método de Newton – Raphson, más la graficación de funciones como se aprecia en Fig.3.

Fig. 3. *Interfaz del Software Thomas Bayes*

Con este software se pretende apropiar el concepto de los algoritmos matemáticos de los métodos numéricos para la resolución de ecuaciones no lineales de una variable real, con el fin de agilizar los procesos iterativos en búsqueda de una raíz aproximada.

III. RESULTADOS Y DISCUSIÓN DE LA INVESTIGACIÓN

3.1 Estrategia didáctica en la apropiación del software

Desde el estudio comparativo de los métodos de bisección, punto fijo y Newton - Raphson, donde se involucraron las variables de estudio entre las cuales se destaca la eficiencia y la exactitud, lo cual hace parte de la matemática aplicada y que encuentra su apoyo en la ciencia de la computación, la ingeniería, entre otras disciplinas. Por consiguiente, el método a trabajar dentro de la investigación es el método científico [6] que permitirá la enseñanza didáctica de los diferentes métodos numéricos en la apropiación y la aplicabilidad de los algoritmos para la resolución de un modelo matemático de la forma $f(x)=0$. Dentro del proceso investigativo se logró definir las variables para poder cuantificarlas y compararlas para el estudio en mención, definiendo eficiencia como el número de iteraciones de cada proceso y la exactitud como la evaluación de la función en la raíz calculada y el valor más cercano a cero será el valor más exacto.

La investigación del estudio comparativo de los métodos se enmarca en el enfoque cuantitativo porque se midió las variables: eficiencia y exactitud de los métodos, determinando el método más pertinente para encontrar la solución aproximada de un modelo matemático; a partir de ello se propone una didáctica para la enseñanza y aprendizaje de la aplicabilidad de los métodos con ayuda de la herramienta del software para dinamizar el proceso de enseñanza aprendizaje del docente y estudiante o viceversa con el fin de interiorizar la teoría matemática que muchas situaciones es confusa o abstracta para muchos estudiantes, donde se procede a identificar el comportamiento de las variables de estudio. En este orden de ideas se inicia con la graficación de la función utilizando el software Thomas Bayes para visualizar el intervalo que contiene la solución y poder esTable.cer las condiciones iniciales requeridas para el proceso. Seguidamente se aplica el algoritmo en cada método para solucionar manualmente la ecuación, los resultados se corroboran solucionando la ecuación con el software. Los datos obtenidos se analizan para verificar el proceso realizado manualmente el cual fortalece la apropiación de los algoritmos de los métodos y al final se identificaron ventajas y desventajas en cada uno.

Por otra parte, cabe recalcar que el énfasis del estudio radicó en cómo aprovechar las herramientas tecnológicas, entre ellas el software para interactuar mejor con la teoría y al mismo tiempo con la reacción de los estudiantes frente a diversos softwares que se implementaron para el estudio del comportamiento de las curvas matemáticas, identificando los requerimientos necesarios para el arranque del algoritmo y el posterior análisis de convergencia o no a la raíz deseada.

Por lo tanto, se resuelve una ecuación matemática de forma manual para que el estudiante apropie todos los conceptos algebraicos para encontrar la raíz aproximada por los métodos mencionados anteriormente y posteriormente se la resuelve apoyándose con la herramienta del software Thomas Bayes, para evitar todos los procesos y cálculos que exige la parte manual del ejercicio y de esta manera optimizar el tiempo de la resolución de la ecuación y concentrarse mejor en el análisis e interpretación de la respuesta, su algoritmo, sus iteraciones, entre otros procesos, que permitan entender cómo opera el método numérico e interiorizar procesos matemáticos que involucra el algoritmo.

De acuerdo a lo anterior se plantea resolver la siguiente ecuación que involucre funciones trigonométricas, exponenciales y polinómicas, como es el caso:

Encontrar una de las raíces de la función $f(x)=e^x cosx + 5x^2 + 3x - 3$, usando los métodos de bisección, punto fijo y el de Newton – Raphson con un error $\leq 10^{-4}$.

Solución:

Proceso Manual

En primera instancia, se debe realizar la gráfica de la función $f(x)$ para tener una idea aproximada de cómo es su comportamiento con el apoyo de la matemática básica y de esta manera identificar los posibles intervalos donde se encuentra las raíces o ceros de la ecuación $f(x) = 0$.

Para realizar su gráfica se utiliza uno de los métodos de graficación como es la tabulación como se ve en la tabla 1, que consiste en dar valores a la variable x y obtener los valores de la función $f(x)$, para ello es importante analizar el dominio de la función, donde el dominio para este caso son todos los números reales, ya que ningún valor de x presenta alguna inconsistencia.

Tabulación de la función: $f(x)=e^x cosx + 5x^2 + 3x - 3$ *(1)*

Table. 1. *Proceso de la tabulación de (1)*

x	-3.5	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	-0.5	1.0	1.5	2.0	2.5	3.0	3.5
$f(x)$	47.7	33	20.7	10.9	3.8	-0.8	-2.7	**-2**	1.2	6.5	13.1	19.9	26	31.1	37.7

En segundo lugar, se procede a graficar las coordenadas de cada punto o pareja ordenada de la forma (x, y) donde $y = f(x)$ en el plano cartesiano con los ejes x (eje horizontal) y eje y (eje vertical), como se muestra en la Fig. 4.

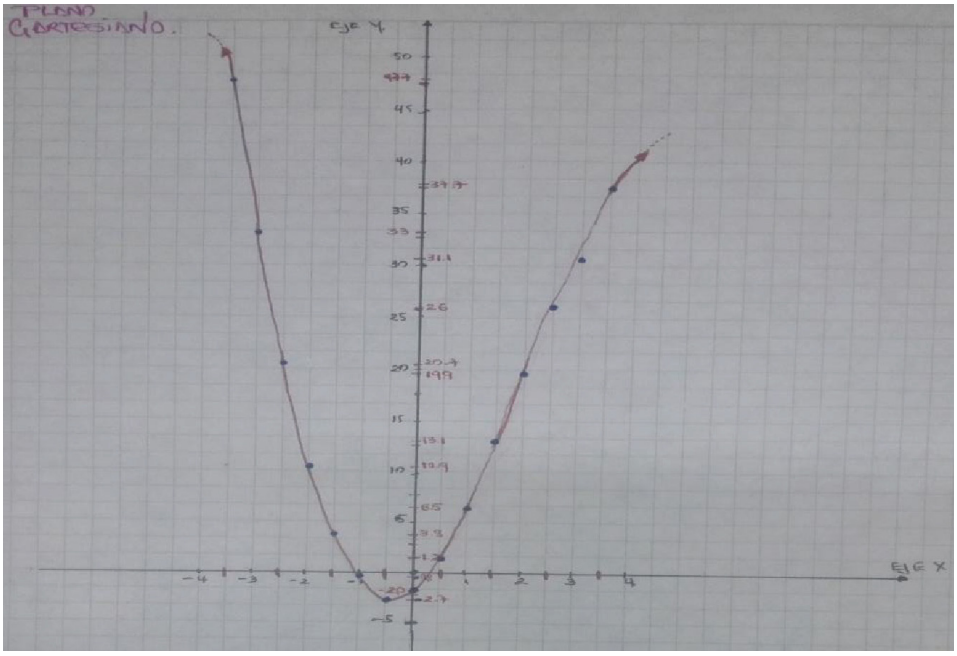


Fig. 4. *Gráfica de la función (1)*

Se observa una porción de la gráfica de la función y se puede apreciar dos posibles intervalos como son ***(-1.5, -0.5)*** y ***(0,1)*** donde se encuentra una de las raíces de la ecuación. Para efectos de la aplicación de los métodos numéricos se trabaja en el intervalo ***(0,1)***.

En tercer lugar, se aplica los algoritmos de los métodos numéricos.

3.2 Método de bisección

$$e^x cosx + 5x^2 + 3x - 3 = 0 \qquad (2)$$

Para aplicar el algoritmo del método es necesario seguir los siguientes pasos:

- **Primero:** Encontrar el intervalo de partición $[a , b]$

Probando dos números reales del dominio de la función, de tal manera que las imágenes de los números involucrados sean de signos contrarios, se tiene que:

$$f(0) = -2 \text{ y } f(1) = 6.4686$$

De esta manera se identifica que el intervalo de partición es [0,1], donde el valor de sus imágenes es de signo contrario, el cual es requisito primordial para el método de bisección.

- **Segundo:** Determinar el número de particiones $\boldsymbol{n}$

Como se pide encontrar una raíz aproximada con un error de estimación $\leq 10^{-4}$ y aplicando la fórmula de $\boldsymbol{n}$ se obtiene:

$$n \geq \frac{log(10^4|1 - 0|)}{log(2)} = 13.28 \approx 13 \text{ (por redondeo)}$$

Así que en este caso se requiere de 13 particiones del intervalo inicial para encontrar la raíz aproximada.

- **Tercero:** Proceso de particiones.

p_i: Son los puntos medios de cada subintervalo, $\forall\, i = 0,1,2, \ldots ,13$.

$$0 \underset{p_0}{\text{———}} 1 \qquad p_0 = \frac{0 + 1}{2} = 0.5 \qquad f(p_0) = 1.1968 > 0$$

Significa que se toma la parte izquierda de la partición, es decir: ***(0.0, 0.5)***, para la siguiente iteración:

$$0.0 \underset{p_1}{\text{———}} 0.5 \qquad p_1 = \frac{0.0 + 0.5}{2} = 0.25 \qquad f(p_1) = -0.6933 < 0$$

Significa que se toma la parte derecha de la partición, es decir: ***(0.25, 0.5)***, para la siguiente iteración:

$$0.25 \quad \underset{p_2}{\text{———}} \quad 0.5 \qquad p_2 = \frac{0.25 + 0.5}{2} = 0.375 \qquad f(p_2) = 0.1820 > 0$$

Significa que se toma la parte izquierda de la partición, es decir: ***(0.25, 0.375)***, para la siguiente iteración:

$$0.25 \quad \underset{p_3}{\text{———}} \quad 0.375 \qquad p_3 = \frac{0.25 + 0.375}{2} = 0.37125 \qquad f(p_3) = -0.2735 < 0$$

Significa que se toma la parte derecha de la partición, es decir: ***(0.3125, 0.375)***, para la siguiente iteración:

$$0.3125 \quad \underset{p_4}{\text{———}} \quad 0.375 \qquad p_4 = \frac{0.3125 + 0.375}{2} = 0.3437 \qquad f(p_4) = -0.0505 < 0$$

Significa que se toma la parte derecha de la partición, es decir: ***(0.3437, 0.375)***, para la siguiente iteración llamada p_5 y de ahí en adelante se resumen las siguientes iteraciones en la siguiente tabla 2.

Table. 2. *Resultados de las iteraciones*

Intervalo	***Punto Medio*** (p_i)	$f(p_i)$
$0.3437 - 0.375$	$p_5 = 0.3593$	$0.0642 > 0$
$0.3437 - 0.3593$	$p_6 = 0.3515$	$0.0065 > 0$
$0.3437 - 0.3515$	$p_7 = 0.3476$	$-0.0220 < 0$
$0.3476 - 0.3515$	$p_8 = 0.3495$	$-0.0081 < 0$
$0.3495 - 0.3515$	$p_9 = 0.3505$	$-0.0007 < 0$
$0.3505 - 0.3515$	$p_{10} = 0.3510$	$0.0028 > 0$
$0.3505 - 0.3510$	$p_{11} = 0.3507$	$0.0006 > 0$
$0.3505 - 0.3507$	$p_{12} = 0.3506$	$-0.00005 < 0$
$0.3506 - 0.3507$	$p_{13} = 0.35065$	$0.0003 < 0$

Se acaban de hacer 13 iteraciones, tal como se ve en el algoritmo anterior y en la última iteración se obtuvo: $f(p_{13}) = 0.0003$, determinando un error de 10^{-4}.

Por tanto, la raíz buscada por el método de bisección es el último punto medio $p = p_{13} = \mathbf{0.35065}$, con un error de estimación de 10^{-4}.

Aplicación del software Thomas Bayes

En primera instancia, se debe realizar la gráfica de la ecuación (2) para tener una idea aproximada de cómo es su comportamiento de la curva, detectar su dominio, sus intervalos donde se encuentra las raíces o ceros de la ecuación $f(x) = 0$.

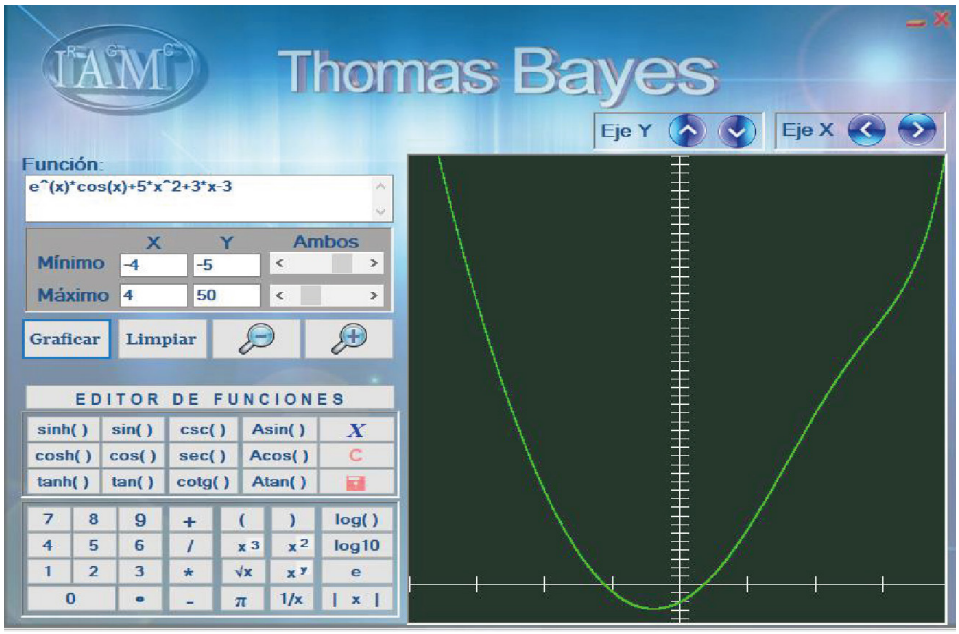
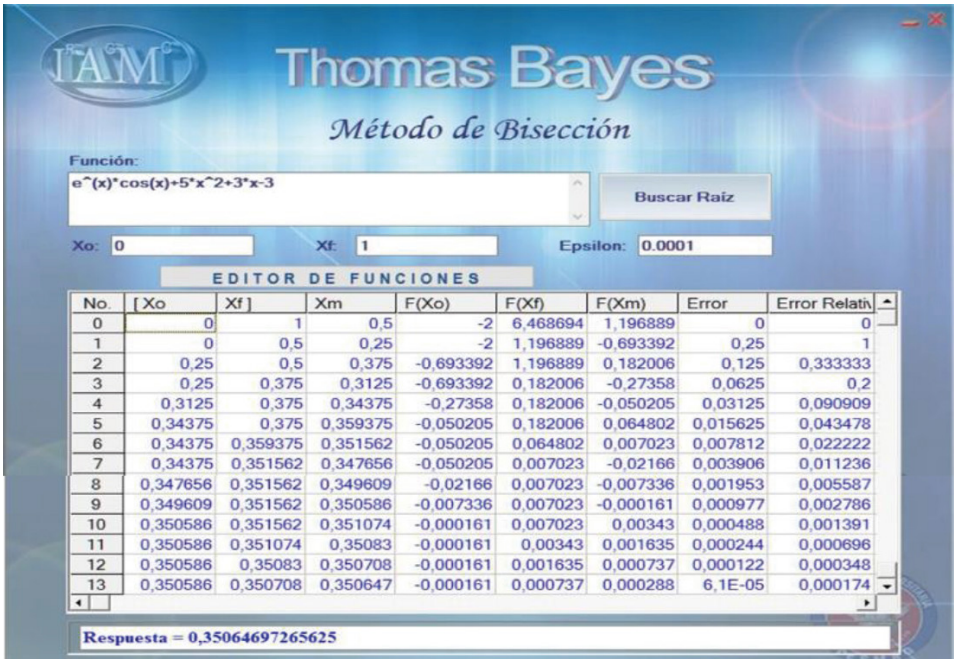


Fig. 5. *Gráfica de la función (1)*

Utilizando la herramienta del software se obtiene la gráfica de la función tal como se ve en Fig.5. Por medio del software, el algoritmo del método de bisección para el desarrollo de la ecuación se observa en Fig.6.



No.	[Xo	Xf]	Xm	F(Xo)	F(Xf)	F(Xm)	Error	Error Relati
0	0	1	0,5	-2	6,468694	1,196889	0	0
1	0	0,5	0,25	-2	1,196889	-0,693392	0,25	1
2	0,25	0,5	0,375	-0,693392	1,196889	0,182006	0,125	0,333333
3	0,25	0,375	0,3125	-0,693392	0,182006	-0,27358	0,0625	0,2
4	0,3125	0,375	0,34375	-0,27358	0,182006	-0,050205	0,03125	0,090909
5	0,34375	0,375	0,359375	-0,050205	0,182006	0,064802	0,015625	0,043478
6	0,34375	0,359375	0,351562	-0,050205	0,064802	0,007023	0,007812	0,022222
7	0,34375	0,351562	0,347656	-0,050205	0,007023	-0,02166	0,003906	0,011236
8	0,347656	0,351562	0,349609	-0,02166	0,007023	-0,007336	0,001953	0,005587
9	0,349609	0,351562	0,350586	-0,007336	0,007023	-0,000161	0,000977	0,002786
10	0,350586	0,351562	0,351074	-0,000161	0,007023	0,00343	0,000488	0,001391
11	0,350586	0,351074	0,35083	-0,000161	0,00343	0,001635	0,000244	0,000696
12	0,350586	0,35083	0,350708	-0,000161	0,001635	0,000737	0,000122	0,000348
13	0,350586	0,350708	0,350647	-0,000161	0,000737	0,000288	6,1E-05	0,000174

Fig. 6. *Método de bisección*

Se ve que la raíz buscada de la ecuación (2) es de 0.350646 con un margen de error de 10^{-4}.

3.3 Método del punto fijo

Para encontrar la raíz aproximada de la ecuación (2) se aplica el algoritmo del método mediante los siguientes pasos:

- **Paso 1.** Graficación de la función

Primero se ha graficado la función (1), como aparece en la figura 6, tal como se hizo con el método anterior donde una de las raíces de la función se encuentra en el intervalo abierto $(0,1)$.

- **Paso 2.** Elección de los valores iniciales

Para encontrar la raíz aplicando el método del punto fijo elegimos el valor inicial $x_0 = 1$, que está cercano al punto de corte de la función con el eje horizontal, como se puede apreciar en la figura 6.

- **Paso 3.** Buscar la función iteradora $g(x)$ adecuada

Al utilizar el método del punto fijo con un margen de error de 10^{-4} y con un valor inicial $x_0 = 1$, es necesario hallar la función iteradora $g(x)$, no sin antes aclarar que la función utilizada para aplicar este método se la encontró después de probar con diferentes formas algebraicas que no siempre convergen, obteniendo así la nueva función para el algoritmo del punto fijo como se indica a continuación.

$$\begin{aligned} f(x) &= 0 \\ e^x\cos x + 5x^2 + 3x - 3 &= 0 \\ 5x^2 + 3x &= 3 - e^x\cos x \\ x(5x + 3) &= 3 - e^x\cos x \\ x &= \frac{3 - e^x\cos x}{5x + 3} \end{aligned}$$

Seguidamente se debe expresar la ecuación dada de la forma $x = g(x)$, para esta ecuación se ha esTable.cido que la función iteradora $g(x) = \frac{3 - e^x\cos(x)}{5x + 3}$.

- **Paso 4.** Iteraciones con la función $g(x)$

Utilizando la función iteradora encontrada $g(x) = \frac{3 - e^x\cos(x)}{5x + 3}$, se aplica el algoritmo matemático del método, obteniendo las siguientes iteraciones:

$$\begin{aligned} x_0 &= 1 \\ g(1) &= 0.191413 \\ g(0.191413) &= 0.457702 \\ g(0.457702) &= 0.299183 \\ g(0.299183) &= 0.380602 \\ g(0.380602) &= 0.334801 \\ g(0.334801) &= 0.359422 \\ g(0.359422) &= 0.345840 \\ g(0.345840) &= 0.353230 \end{aligned}$$

$$\begin{aligned} g(0.353230) &= 0.349178 \\ g(0.349178) &= 0.351390 \\ g(0.351390) &= 0.350180 \\ g(0.350180) &= 0.350841 \\ g(0.350841) &= 0.350480 \\ g(0.350480) &= 0.350677 \\ g(0.350677) &= 0.350569 \\ g(0.350569) &= 0.350628 \end{aligned}$$

Se observa claramente la tendencia de la solución que ofrece el método del punto fijo, donde la última iteración corresponde a la solución aproximada de la raíz de valor de ***0,350628*** con un error de estimación de 10^{-4}.

Aplicación del software Thomas Bayes

Esta misma tendencia se puede apreciar desde el punto de vista gráfico en Fig.7. que representa la función iteradora $g(x) = \frac{3 - e^x \cos(x)}{5x + 3}$ y la función idéntica $y = x$. De igual forma muestra la convergencia del método.

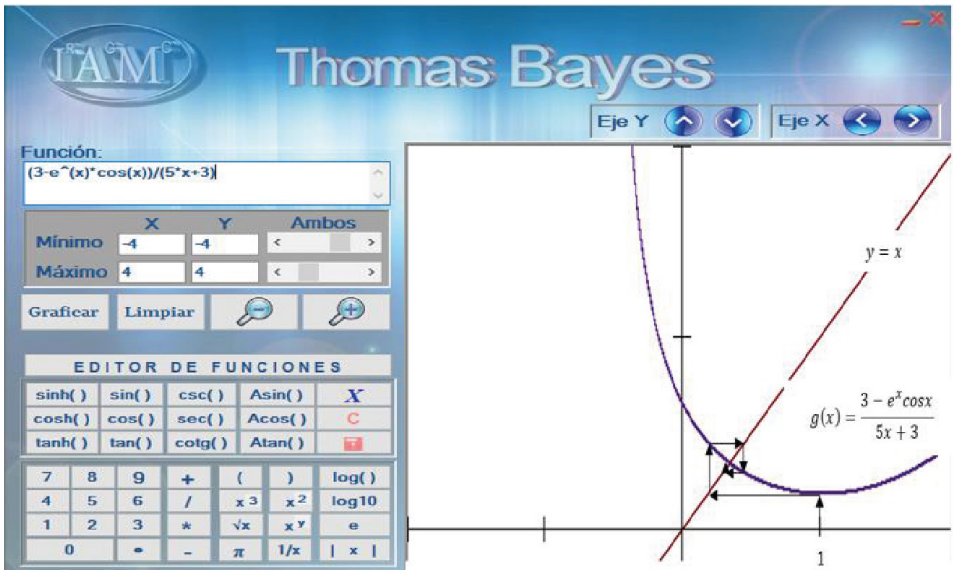


Fig. 7. *Gráfica de las funciones* $g(x) = \frac{3 - e^x \cos(x)}{5x + 3}$ *y* $y = x$

Por medio del software, el algoritmo del método del punto fijo para el desarrollo de la ecuación con un margen de error de 10^{-4} y con un valor inicial $x_0 = 1$ se observa en Fig.8.

No.	Xo	g(Xo)	Error	Error %
1	1	0,1914132575	0	0
2	0,1914132575	0,4577019007	0,2662886432	58,1794925406
3	0,4577019007	0,2991837448	0,1585181559	52,9835456276
4	0,2991837448	0,3806024978	0,0814187531	21,392070086
5	0,3806024978	0,3348016816	0,0458008163	13,6799839468
6	0,3348016816	0,3594219861	0,0246203045	6,8499717478
7	0,3594219861	0,3458405725	0,0135814136	3,9270735359
8	0,3458405725	0,3532298759	0,0073893035	2,0919248331
9	0,3532298759	0,3491787061	0,0040511698	1,1601995696
10	0,3491787061	0,3513905567	0,0022118506	0,6294564611
11	0,3513905567	0,3501801816	0,0012103751	0,3456435134
12	0,3501801816	0,3508417041	0,0006615225	0,1885529796
13	0,3508417041	0,3504799074	0,0003617967	0,1032289433
14	0,3504799074	0,350677706	0,0001977986	0,0564046683
15	0,350677706	0,3505695451	0,0001081608	0,0308528935
16	0,3505695451	0,3506286834	5,91383E-05	0,016866356

Fig. 8. *Método del punto fijo* $f(x)=e^x \cos x + 5x^2 + 3x - 3$

Por medio del método del punto fijo en Fig.10, se ve que la raíz buscada de la función (3) es de 0.350569 con un margen de error de 10^{-4}.

3.4 Método de Newton – Raphson

De acuerdo a la gráfica de la función realizada en el método de bisección se observó que una de las raíces está en el intervalo abierto $(0,1)$, el cual se va a tener presente para aplicar el algoritmo del método en la ecuación (2) y es necesario seguir los siguientes pasos:

- **Paso 1.** Calcular la derivada de $f(x)$

$$f'(x)=e^x\cos(x) + e^x sen(x) + 10x + 3$$

- **Paso 2.** Valor inicial y evaluar

$$x_0 = 0 \text{ y } f'(0) = 4 \neq 0$$

Nota: Es necesario que la evaluación de la derivada sea un valor lejano de cero para garantizar la convergencia del algoritmo.

- **Paso 3.** Proceso iterativo

$$x_0 = 0$$
$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 0 - \frac{(-2)}{4} = 0.5$$
$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 0.5 - \frac{1.1968}{8.6564} = 0.3617$$
$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 0.3617 - \frac{0.0821}{7.4518} = 0.3506$$
$$x_4 = x_3 - \frac{f(x_3)}{f'(x_3)} = 0.3506 - \frac{(-0.00005)}{7.3518} = 0.3506$$

Luego, la raíz aproximada buscada por el método es de 0.3506 con un margen de error 10^{-4}.

Aplicación del software Thomas Bayes

De igual forma, utilizando el software para el algoritmo del método de NewtonRaphson, con valor inicial $x_0 = 0$, , se mira en Fig.9.

Fig. 9. *Método de Newton para la función (1)*

La raíz aproximada buscada de la función (1) es de 0.350607 con un margen de error de 10^{-4}.

3.5 Ventajas y desventajas de la aplicación en la solución de la función con las formas manual y con la aplicación del software.

FORMA MANUAL	
Ventajas	***Desventajas***
• *Afianzamiento de los conocimientos de las matemáticas básicas.* • *Afianzamiento de los conocimientos del álgebra.* • *Proceso de graficación.* • *Fortalecimiento de procesos aritméticos.* • *Utilización de recursos didácticos como colores, regla, calculadora, entre otros.* • *Análisis del comportamiento de la raíz aproximada en cada proceso iterativo.* • *Mejorar la capacidad de atención y concentración en todo el proceso de ejecución.*	• *Los métodos numéricos exigen de numerosos cálculos, que son dispendiosos para encontrar la raíz.* • *Utiliza tiempo considerable del estudiante para encontrar la respuesta final.* • *Por la parte operativa es factible cometer varios errores.* • *Si hay error en alguna parte del proceso iterado éste se vuelve más grande y por ende la raíz buscada es incorrecta.* • *El proceso es mecánico lo cual no permite identificar posibles errores cometidos.*

Con aplicación del Software THOMAS BAYES	
Ventajas	***Desventajas***
• *Optimiza el tiempo para encontrar la solución de una ecuación matemática.* • *Evita cometer errores de cálculo en el proceso iterado.* • *Como la solución aproximada es inmediata esto permite dedicarle tiempo valioso al estudio del comportamiento de las funciones matemáticas y de esta manera determinar las condiciones iniciales de cada método.* • *Permite analizar el comportamiento de la raíz o cero de la función matemática.* • *Por medio del software permite detallar de manera gráfica el comportamiento de la curva.* • *Además, el software por los resultados presentados en la aplicación de cada uno de los métodos se puede analizar la eficiencia y la exactitud de la raíz encontrada (resultados).*	• *La agilidad mental es afectada en la aplicación de los conocimientos de la matemática, del álgebra y la aritmética.* • *Disminuye la habilidad de realizar la representación gráfica de las funciones.* • *Permite la dispersión de la atención y concentración del estudiante en la ejecución de obtener la solución aproximada.*

IV. CONCLUSIONES

De acuerdo al comportamiento de los algoritmos de cada uno de los métodos para la solución de las ecuaciones no lineales, se realizó un estudio comparativo sobre las variables de Eficiencia y Exactitud viendo que el método de Newton es más eficiente, teniendo en cuenta que todos los métodos llegan a una misma solución aproximada.

Mediante el software se permite hacer un estudio más detallado del comportamiento de la curva matemática identificando las condiciones iniciales que se requiere para la aplicación del algoritmo de cada método numérico para la resolución de ecuaciones no líneas univariadas.

La utilización de la herramienta didáctica Thomas Bayes elimina la parte tediosa de los cálculos numéricos que exige cada iteración de los métodos.

Con la aplicación del software Thomas Bayes se mira la aplicabilidad de cada uno de los métodos para encontrar la solución.

Dentro de la matemática hay software de aplicación como GEOGEBRA que presenta las bondades de un procesador gráfico y WOLFRAMALPHA que es un buscador de procesos matemáticos, los cuales exigen un conocimiento previo de comandos y programación. En la investigación se desarrolló un software didáctico que integra el procesador gráfico y procesos matemáticos que se requieren en los algoritmos de los métodos numéricos para su entendimiento y aplicación de modelos matemáticos.

Una vez terminada la investigación, teniendo como resultado el producto tecnológico software Thomas Bayes, se recomienda generar más herramientas de este tipo con el fin de mejorar el proceso de enseñanza – aprendizaje.

Es de vital importancia resolver ecuaciones de la forma $f(x) = 0$, en muchas áreas de la ciencias e ingenierías, como en las matemáticas, la física, la química, entre otras, y también en el cálculo de algunas funciones o constantes matemáticas como las raíces cuadradas, raíces cúbicas u otras funciones polinómicas y trascendentales.

REFERENCIAS

[1] *Brito, M., Jara, C., Larenas, C.: El software educativo en el aula, tesis universitaria, Universidad Academia, Santiago de Chile (2009).*

[2] *Cantoral, R., Montiel, G.: Visualización y pensamiento matemático. Acta Latinoamericana de Matemática Educativa (16), 694-701 (2003).*

[3] *Guerrero, H., Revelo, I., Casanova, M.: Estudio comparativo sobre la eficiencia, precisión, exactitud y pertinencia de los métodos de bisección, punto fijo y Newton Raphson para solucionar ecuaciones no lineales de una variable. 1ra edición, Institución universitaria CESMAG, Pasto Colombia (2017).*

[4] *GEOGEBRA, https://www.geogebra.org/, accedido: 2018/06/20.*

[5] *WOLFRAMALPHA, https://www.wolframalpha.com/, último acceso: 2018/07/15.*

[6] *Quijano, A.: Guía de investigación cuantitativa. 1ra edición. Tecnografic, San Juan de Pasto (2009).*

Biografía Autor(es)

a. Magister en Matematicas Aplicadas. Docente tiempo completo, Universidad Cesmag. E-mail: haguerrero@unicesmag.edu.co. Pasto, Colombia.

b. Especialista en Multimedia Educativa. Docente tiempo completo, Universidad Cesmag. E-mail: rjhernandez@unicesmag.edu.co. Pasto, Colombia.

EnglishPlay: Videojuego para el apoyo a la enseñanza del inglés

Luis Fernando Quintero Castaño (a)
Jorge Leonardo Zapata Castaño (b)
Andres Mauricio Bueno Peña (c)
Javier Simón Naranjo Herrera (d)
Carlos Andrés Delgado Saavedra (e)
Yuri Mercedes Bermúdez Mazuera (f)
Edwin Gamboa (g)

Resumen

English Play es un videojuego para la enseñanza del idioma inglés para estudiantes universitarios de primer semestre enfocado en lectura de textos académicos en inglés. Se compone de cuatro mundos virtuales y un conjunto de minijuegos llamados desafíos.

La creación del videojuego nace de la necesidad de dar un recurso que sirva como apoyo al aprendizaje, ya que se han evidenciado dificultades en la enseñanza del idioma inglés a los estudiantes de primer semestre de la universidad. Se busca que con la incorporación a los procesos de enseñanza de este videojuego, se motive a los estudiantes en el aprendizaje.

Con respecto a las dificultades de aprendizaje, se ha evidenciado que los estudiantes tienen la percepción de que las clases de inglés son repetitivas y aburridas. Dicho problema, puede resolverse parcialmente mediante la inclusión de diferentes herramientas interactivas o mecanismos de enseñanza de un segundo idioma. Usando los videojuegos como herramienta interactiva para el apoyo a la enseñanza del idioma inglés, los estudiantes pueden jugar y divertirse mientras aprenden, olvidando la rigurosidad de las metodologías clásicas de enseñanza.

Como metodología de desarrollo del juego se utiliza Game-Scrum, la cual consiste en 3 etapas: Pre-Producción, para el diseño del videojuego; Producción, para la codificación y Post-Producción, para las pruebas y despliegue.

Por último, se realiza evaluación heurística para medir los aspectos de Jugabilidad y Usabilidad del videojuego, con el fin de analizar y saber lo que sienten los usuarios ante la interacción con el videojuego y evaluar la calidad del juego en términos de su funcionamiento. Los resultados obtenidos permiten concluir que las estrategias de aprendizaje interactivas, como los videojuegos en el aprendizaje de una segunda lengua tienen buena recepción e influyen positivamente en su proceso de aprendizaje.

Palabras clave: Evaluación heurística, experiencia de jugador, juego serio, jugabilidad, metodologías de aprendizaje, usabilidad, videojuego.

EnglishPlay: Video game for learning in English

Abstract

English Play is a video game for teaching English to students first semester college students focused on reading academic texts in English. It is made up of four virtual worlds and a set of mini-games called challenges.

The creation of the video game was born from the need to provide a resource that serves as learning support, since difficulties have been evidenced in language teaching English to first-semester college students. It is sought that with the incorporation to the teaching processes of this video game, students are encouraged to The learning.

With regard to learning difficulties, it has been shown that students They have the perception that English classes are repetitive and boring. Said problem, can be partially solved by including different tolos interactive or second language teaching mechanisms. Using video games As an interactive tool to support English language teaching, students can play and have fun while learning, forgetting the rigor of the classical teaching methodologies.

Game-Scrum is used as a game development methodology, which consists of 3 stages: Pre-Production, for the design of the video game; Production, for coding and Post-Production, for testing and deployment.

Finally, heuristic evaluation is carried out to measure the aspects of Gameplay and Usability of the video game, in order to analyze and know what users feel to the interaction with the video game and evaluate the quality of the game in terms of its functioning. The results obtained allow to conclude that the strategies of interactive learning, such as video games in learning a second language they are well received and positively influence their learning process.

Keywords: Heuristic evaluation, gamer experience, serious gameplay, playability, learning methodologies, usability, video game.

I. INTRODUCCIÓN

La generación de personas nacidas a partir del año 2001 es conocida como nativos digitales[9], los cuales han crecido en medio de tecnologías como computadoras, celulares, entre otras. Su destreza cognitiva[7] se ha desarrollado en medio de estas tecnologías, por lo que han adquirido gran parte de sus experiencias a través de ellas. En consecuencia, los métodos de enseñanza actuales se ven enfrentados a la necesidad de una transformación en los métodos pedagógicos si desean convocar y ser inspiradoras para las nuevas generaciones de jóvenes, ya que éstos tienen como preferencia los gráficos a los textos y aprender jugando en lugar de ser obligados a memorizar información [8].

Para enfrentar este problema, se han desarrollado diferentes estrategias para la educación desde como Las Tecnologías de la Información y Comunicación (TIC's), lo que ha permitido el desarrollo de metodologías para el aprendizaje [10], la búsqueda de innovación en la pedagogía y la forma de organizar la enseñanza - aprendizaje por medio de escenarios virtuales y dispositivos tecnológicos. Uno de estos son los videosjuegos, que han servido como herramientas para aprender conductas y actitudes necesarias para el eficiente desempeño sociocultural, proporcionando a los jugadores habilidades y destrezas que facilitan el aprendizaje de procesos complejos con eficacia [11]. En el caso especifico de videojuegos para el aprendizaje de idiomas, se destaca Ruby Rei [5], el cual es divertido, con una interfaz y gráficos llamativos, colorido, con posibilidad de recorrer y explorar el mundo aprendiendo mediante ejercicios.

En este trabajo se realiza la implementación de un videojuego que sirva como apoyo para los estudiantes de la asignaturas de Inglés en los primeros semestres, teniendo en cuenta las estadísticas de la Universidad en estudio, la cual trabaja con una escala de notas entre 0 a 5, en donde se requiere una nota mayor o igual a 3 para aprobar. En datos históricos entre 2012 y 2016 de la universidad, se observa que en el año 2015 un 48,94 % y en el 2016 un 41,48 % de los estudiantes perdieron la asignatura, El objetivo es buscar motivar el aprendizaje del idioma y dar una herramienta adicional que apoye el proceso de enseñanza de la asignatura, buscando mejorar estas estadísticas.

El videojuego diseñado, considera los temas que más dificultades tienen los estudiantes de la asignatura, de acuerdo a una valoración realizada por los docentes a su cargo. Esta propuesta, la cual llamamos EnglishPlay, es un juego compuesto por 4 mundos, en cada mundo el estudiante debe solucionar un conjunto de desafíos, algunos son de seleccionar una respuesta de múltiples opciones como el desafío Laser, Plataformas, Electroshock; otro llamado Camino Correcto trata de traducir una oración completa del español al inglés, en Shooter el estudiante debe buscar al enemigo cuya palabra ubicada encima de su cabeza, sea la correcta ante la pregunta planteada. Para el análisis de impacto del videojuego se seleccionaron un conjunto de heurísticas de usabilidad y jugabilidad, que son evaluadas mediante los siguientes métodos: evaluación heurística, observación de campo, y cuestionario.

II. METODOLOGÍA

2.1. Elaboración del proyecto

Para la elaboración del videojuego se utilizó una adaptación a Scrum enfocado en el desarrollo de videojuegos conocida como GameScrum, debido a que tiene un enfoque

de gestión ágil, facilitando de esta manera la administración del proyecto, además de la comunicación del equipo de trabajo. Se tienen las siguientes etapas:

- *Pre-Producción:* En esta etapa, se definió el contenido y las estrategias didácticas a utilizar en el videojuego. Con el contenido definido, se creó un documento de diseño de juego que incluye aspectos como la historia, la mecánica y las actividades. Paralelamente se definieron las heurísticas de usabilidad y experiencia de jugador. documento.

- *Producción:* En esta etapa se implementó el videojuego, se desarrolló su mecánica e interfaz y se desarrollaron reglas para que se convirtiera en un juego funcional según lo planificado en la etapa anterior.

- *Post-Producción:* En esta etapa el videojuego se validó usando pruebas de usabilidad y jugabilidad con un grupo de estudiantes que representaba la audiencia objetivo. Estas pruebas se realizaron con el fin de validar el videojuego desde la perspectiva del jugador. Sobre la base de la retroalimentación dada por los estudiantes y el docente, se hicieron los cambios y mejoras propuestos.

2.2. Estudio de dificultades de aprendizaje

Se realiza un sondeo con los profesores del curso de inglés en el primer semestre, para identificar los temas que más dificultades de aprendizaje presentan.

Los temas seleccionados son:

- *Tiempos verbales simples:* Presente, pasado, futuro.

- *Tiempos continuos:* Presente, pasado, futuro.

- *Tiempos perfectos:* Presente, pasado, futuro, presente continuo, pasado continuo.

- *Condicional:* Tipos de condicional.

- *Voz Pasiva:* Presente, pasado, futuro, voz pasiva personal, voz pasiva impersonal.

- *Preposiciones:* In-at-on, de lugar, de tiempo, de movimiento-dirección

- *Conectores:* Tipos de conectores.

- *Adjetivos:* Comparativos (regulares, irregulares, de inferioridad, de igualdad, de superioridad), superlativos (regulares, irregulares, de superioridad, de inferioridad).

Se definen las siguientes estrategias de aprendizaje:

- Table.ro virtual (Libro), en el cual se pueden consultar las reglas gramaticales y ejemplos sobre alguno de los temas

- Desafíos, ejercicios de distinto tipo para que el estudiante no mecanice las actividades y practique la gramática del inglés.

- Comprensión lectora (NPC Test), dentro del juego existen NPC (personajes no jugables) que realizan una serie de pruebas sobre un texto que se debe leer previamente.

Historia

En un reino llamado Amog siempre ha existido paz y seguridad gracias al libro de la sabiduría, el cual emana una luz protectora que se dispersa por todo el reino, evitando que los seres de oscuridad (oscuros) puedan ingresar. Un día la luz protectora es apagada mediante un hechizo oscuro realizado por Ogam, el hechicero más poderoso del reino y considerado mano derecha del rey. El hechizo consistía en robar todas las palabras que contenía el libro y las dispersaba por todo el reino, quedando el libro vacío y sin poder alguno. Ogam confiesa ser líder de los oscuros, derroca y exilia al rey del castillo. El rey se acerca a una aldea en donde se encuentra con el héroe, quien al escuchar la historia decide partir en búsqueda de las palabras dispersadas y volver a introducirlas dentro del libro para restaurar la paz del reino. El camino que recorre el héroe es largo, lleno de obstáculos, trampas y enemigos, pero él nunca se rinde, se hace más fuerte, más ágil, y más inteligente. Cuando logra reunir todas las palabras se enfrenta con Ogam y lo derrota, introduce nuevamente las palabras en el libro, y éste mágicamente las acomoda y estructura, restaurando nuevamente su poder y liquidando a todo ser oscuro que se encuentre dentro del reino a través de su luz, la cual protege nuevamente a todo el reino. El rey vuelve a ocupar su trono y en agradecimiento al héroe, lo nombra caballero del reino.

NPC Test

La finalidad del NPC Test es reforzar la compresión lectora de los estudiantes, para superar el test el jugador deber ‘a leer un texto completo en Inglés en el cual se incorporan los elementos enseñados en el nivel desde el que se accede, después de leer el texto deber ‘a contestar una serie de preguntas referentes al mismo.

El libro: Es un refuerzo de las clases, una especie de Table.ro virtual en donde los estudiantes podrán consultar la estructura de los temas, ejemplos, reglas gramaticales y excepciones de éstas reglas. Está dividido en cuatro capítulos, en el primero se abordan los tiempos simples y continuos, en el segundo capítulo tiempos perfectos y voz pasiva, en el tercero las preposiciones y en el cuarto se enseñan los adjetivos.

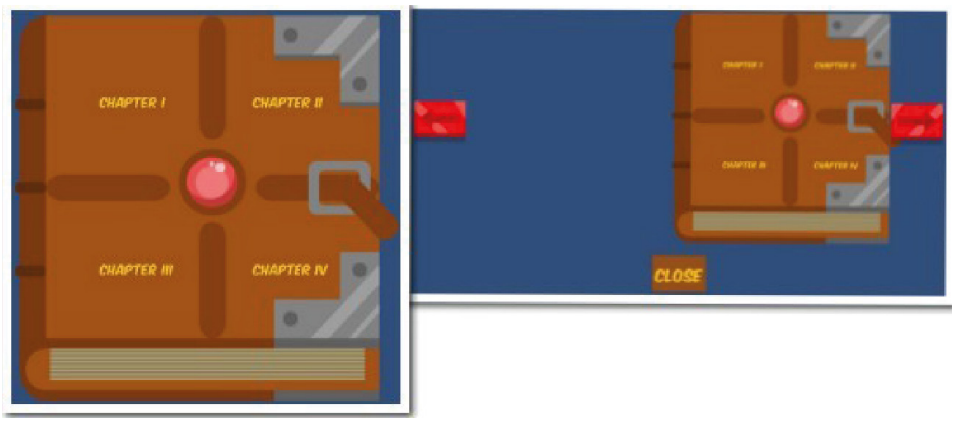


Fig. 1. *Libro de EnglishPlay*

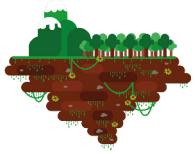
Personajes. EnglishPlay cuenta con personajes tanto masculinos como femeninos, cada personaje puede ocupar un rol, el cual define sus parámetros y habilidades; los roles son:

Table. 1. *Personajes EnglishPlay*

Rol	Características
Guerrero	*Destaca en defensa y vida, su estilo de combate es cuerpo a cuerpo utilizando su espada. Es fácil de utilizar, recomendado para quienes no posean de mucha experiencia con los videojuegos.*
Mago	*Es el personaje con mayor daño de todos y a la vez el mas vulnerable, ataca a distancia mediante bolas de fuego y realiza poderosos hechizo; es el más difícil de utilizar de todos los personajes.*
Arquero	*Personaje equilibrado, posee un nivel medio en defensa y vida (aunque inferiores al guerrero), realiza ataque a distancia a través de su arco y echas, posee buen daño (aunque inferior al mago), y es sencillo de utilizar.*

Mundos. EnglishPlay está formado por cuatro mundos, cada uno de ellos cubre diferentes temas de enseñanza y poseen distinta ambientación.

Table. 2. *Mundos EnglishPlay*

Mundo	Características
Aldea	*Aldea estilo medieval; en este mundo se enseña el presente simple, pasado simple, futuro simple, presente continuo, pasado continuo y futuro continuo con oraciones negativas, afirmativas e interrogativas.*
Bosque 	*Bosque mágico habitado por criaturas fantásticas; los temas que se enseñan en él son los tiempos perfectos y la voz pasiva.*
Cueva	*En este mundo el terreno es irregular, con obstáculos como pozos de agua y de lava, seguido de diferentes desniveles; los temas a enseñar son conectores y preposiciones.*
Castillo	*Tiene un ambiente con una mezcla entre lúgubre (reflejado en los enemigos) y clásico (fondo y plataformas). En este mundo se tratan los siguientes temas: comparativos de inferioridad, de igualdad, de superioridad, otro tipo de comparativos, superlativos de inferioridad y superlativos de superioridad.*

III. DESAFÍOS

Los desafíos son el elemento principal de aprendizaje de EnglishPlay, en ellos el jugador deberá abordar una serie de minijuegos o retos en los cuales deberá aplicar los conceptos vistos en clase.

Las preguntas que el jugador deberá responder varían dependiendo del mundo desde el cual es accedido el desafío, es decir, el mismo desafío tiene diferentes preguntas si es accedido desde el mundo 1 o el mundo 3.

Para superar los desafíos es muy importante la utilización del Libro, así el jugador podrá afianzar los conceptos prácticos/teóricos necesarios en cada desafío.

Table. 3. *Desafíos EnglishPlay*

Desafío	*Dinámica*
Combat	*Completar textos correctamente para atacar a los enemigos, en caso de error el jugador no podrá golpear al enemigo.*
CorrectBox	*Arrastrar cajas hasta la ubicación correcta, la caja contiene diferentes opciones de respuesta donde solo una es correcta y debe ser arrastrar hacia la pregunta que corresponde.*
Platforms	*Se tiene una pregunta y dos opciones de respuesta, una en cada plataforma, el jugador deberá saltar a la plataforma con la respuesta correcta.*
ElectroShock	*El jugador deberá ubicarse en la plataforma que contenga la palabra que complete la frase, si falla será electrocutado.*
LinkedWord	*Enlazar las palabras del lado izquierdo con las frases del lado derecho.*
Shooter	*Buscar y derrotar al enemigo que tenga adherida a su cuerpo la respuesta correcta.*
Láser	*Atravesar el laberinto contestando las preguntas correctamente, en caso de error, el láser ataca al personaje.*
CorrectKey	*Consiste en traducir una oración del español al inglés, para ello, el jugador debe seleccionar las palabras que se encuentran en teclas de piano.*

IV. MÉTODOS DE EVALUACIÓN

Para la evaluación del videojuego, se utiliza los métodos propuestos en *Heuristic And Usability Guidelines For The Creation And Evaluation Of Fun In Video Games* [2], *Using Heuristics to Evaluate the Playability of Games* [1] y *Jugabilidad. Caracterización de la experiencia del jugador en videojuegos* [4]. Estos métodos son sintetizados en la tesis doctoral Jugabilidad. Caracterización de la experiencia del jugador en videojuegos, debido a que en esta se encuentran recopiladas las heurísticas de los otros autores expertos. Además, dicha tesis doctoral cuenta con una batería de heurísticas agrupadas por atributos y facetas.

Los atributos, permiten caracterizar la experiencia del jugador a la hora de jugar un videojuego [6]:

- **Satisfacción:** Agrado que tiene el jugador ante el videojuego y proceso de jugarlo.
- **Aprendizaje:** Facilidad para comprender y dominar el juego y la mecánica del mismo.
- **Efectividad:** Tiempo y recursos utilizados para ofrecer diversión al jugador mientras juega.
- **Inmersión:** Característica del videojuego que define la capacidad de este para hacer creer al jugador que forma parte del mundo virtual que representa.
- **Motivación:** Capacidad del videojuego que mueve a la persona a realizar determinadas acciones y persistir en ellas para su culminación.
- **Emoción:** Impulso involuntario del jugador que se origina como respuesta a los estímulos del videojuego.
- **Socialización:** Atributo que hace apreciar el videojuego de forma distinta al jugarlo en compañía.

Por otro lado, las Facetas de la Jugabilidad, son una descomposición que permite identificar y medir fácilmente los distintos atributos de la experiencia de jugador:

- **Jugabilidad Intrínseca:** Jugabilidad medida en la naturaleza del juego. Se analiza cómo se presentan las reglas, objetivos, el ritmo y las mecánicas del juego.
- **Jugabilidad Mecánica:** Jugabilidad asociada a la calidad del videojuego como sistema de software. Hace hincapié en características como fluidez de las escenas, cinemáticas, iluminación, sonido y el comportamiento de los personajes del juego y el entorno.
- **Jugabilidad Interactiva:** Asociada a lo relacionado con la interacción con el usuario, como el diseño de la interfaz de usuario, mecanismos de dialogo y sistemas de control.
- **Jugabilidad Artística:** Asociada a la calidad artística y estética de los elementos del videojuego. Entre ellos está la calidad gráfica y visual, efectos sonoros, banda sonora, historia y la ambientación realizada de todos los elementos dentro del videojuego.
- **Jugabilidad Intrapersonal o Personal:** Tiene como objetivo la percepción que tiene el usuario del videojuego y los sentimientos que a este le produce.
- **Jugabilidad Interpersonal o Social:** Muestra las sensaciones de los usuarios cuando se juega en compañía ya sea competitiva, cooperativa o colaborativa.

Teniendo en cuenta la batería de heurísticas seleccionada, se procede a aplicar tres métodos de evaluación de sistemas interactivos definidos dentro de la Interacción Persona- Ordenador: *Evaluación heurística*, *Cuestionario*, *Observación de campo* [3].

- **Evaluación Heurística:** Consiste en analizar la conformidad del sistema interactivo con unos principios reconocidos de usabilidad (la \heurística") mediante la inspección de 1 o varios evaluadores.

- **Cuestionario:** Consiste en obtener información a través del dialogo con los usuarios mediante preguntas escritas.
- **Observación de campo:** Consiste en hacer que los usuarios interactúen con el sistema, los evaluadores observarían sus reacciones y a partir de estas podrían hacer preguntas para complementar la información.

V. RESULTADOS Y ANÁLISIS

5.1. Evaluación heurística

Para las pruebas por evaluación heurística, se realiza el siguiente proceso

- Interactuar con el videojuego.
- Analizar cada una de las heurísticas definidas en la plantilla de evaluación heurística.
- Contrastar y validar cada una de las heurísticas con el videojuego
- La valoración de cada pregunta en la heurística esta entre 1 y 5, donde 1 es la menor y 5 es la mayor.

5.2. Evaluación por cuestionario

Para la evaluación por cuestionario se tuvieron en cuenta dos etapas:

Pre-Test Se realizó un pre-test a los usuarios que participaron en la prueba con el fin de hacer un perfil del usuario y sus experiencias con otros juegos. En la Tabla 4 se muestra un resumen de los datos recolectados.

Table. 4. *Perfil de jugador*

Pregunta	*Respuesta (%)*
Edad	*Entre 16-19 (100 %)*
Género	*Masculino (100 %)*
¿Hace cuánto juega videojuegos?	*Entre 10 -15 años (90 %)*
¿Cuál es su frecuencia de uso de videojuegos?	*Todos los días (80 %)*
¿Alguna vez ha utilizado juegos para aprender algún idioma?	*No (80 %)*

Post-Test Después de que los usuarios interactuaron con el videojuego se procede a realizar el cuestionario post-test el cual permitió evaluar la experiencia y emociones que tuvieron los usuarios al jugar.

Table. 5. *Cuestionario Post-Test*

#	Pregunta	Faceta
1	Estaba absorto por el juego.	J. Intrínseca
2	Sabía cómo manipular el juego para seguir adelante.	J. Intrínseca
3	Sabía lo que se suponía que deba hacer para ganar el juego.	J. Intrínseca
4	Entendí las reglas del juego.	J. Intrínseca
5	Sabía todas las acciones que se podrían realizar en el juego.	J. Intrínseca
6	El escenario del juego era interesante.	J. Intrínseca
7	El juego me mantuvo constantemente motivado para seguir jugando.	J. Intrínseca
8	Recuerdo las acciones que los controladores realizaron.	J. Interactiva
9	Pude ver en la pantalla todo lo que necesitaba durante el juego.	J. Interactiva
10	Tenía una buena idea de mi estado en el juego.	J. Interactiva

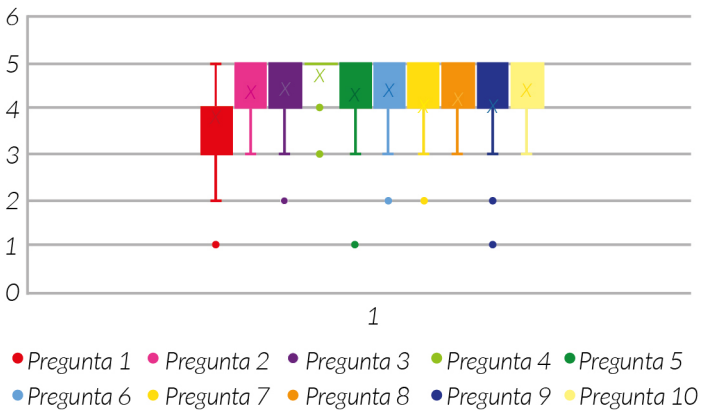


Fig. 2. *Análisis estadístico de cuestionario Post-Test*

Como se observa en la figura 2, la mayoría de los resultados tienen una media entre los valores 4 y 5 lo que permite deducir que durante la interacción con el videojuego los estudiantes tuvieron una buena percepción de él.

Por último, teniendo en cuenta los datos recolectados en el cuestionario, se realizó un promedio de calificaciones asignadas a cada pregunta y se prosiguió a completar la plantilla de evaluación heurística.

5.3. Evaluación por observación de campo

En esta evaluación, se realizó una observación del comportamiento de los usuarios del videojuego mientras interactuaban con el mismo, dando énfasis en sus reacciones, emociones y modo de uso del videojuego.

Para esta evaluación se tuvieron en cuenta los siguientes aspectos:

- Entendimiento de los objetivos del videojuego.
- Emociones generadas (frustración, emoción, felicidad, ira).
- Tiempo invertido y concentración durante el juego.
- Número de intentos.
- Memorización de las reglas del juego.
- Número de veces que se consulta el tutorial.
- Uso del *Libro*.
- Número de elementos de juego conseguidos.
- Rendimiento del jugador ante los retos.
- Una vez culminada la observación, se prosiguió a completar la plantilla de evaluación heurística.

5.4. Discusión de resultados

Después de realizar las pruebas por los 3 métodos de evaluación, se tuvo la siguiente realimentación por cada método:

Realimentación Observación de campo. Según lo analizado en la interacción de los usuarios con el videojuego se generaron las siguientes observaciones:

- El videojuego fue muy bien aceptado, los usuarios disfrutaban al jugar, se observaban reacciones como felicidad al completar retos, risas por eventos ocurridos dentro del juego, concentración al tratar de superar retos o cruzar entre plataformas.
- El uso de los tutoriales fue mínimo (de 1 a 2 veces máximo).
- Algunas veces se generaba frustración por la dificultad de algún reto o el cruce entre plataformas móviles.
- Pocos usuarios usaron el libro, sin embargo, quienes lo utilizaron lo definían como muy práctico y útil, además se observaba una mejora en la solución de retos.

Realimentación Evaluación heurística. Se suministraron las siguientes sugerencias y recomendaciones.

- Se recomienda agregar objetivos secundarios para crear diversidad de objetivos en el juego.
- Se sugiere reducir la altura de los terrenos para poder llegar a ellos fácilmente con los saltos.

- Se recomienda reducir el volumen del sonido de algunos elementos como las monedas.
- Se sugiere aumentar la velocidad de algunas plataformas móviles.

Realimentación por cuestionarios. A continuación, se presentan algunos de los comentarios positivos, negativos y recomendaciones señalados por los usuarios acerca del videojuego.

Aspectos positivos:

- El juego es entretenido
- La historia es interesante
- Los controles son sencillos
- El libro es de gran apoyo en los retos
- Forma entretenida de aprender

Aspectos negativos:

- Debe mejorar la movilidad del personaje
- Debe mejorarse los gráficos
- No hay un personaje afro-descendiente
- No es posible aumentar la velocidad del personaje

Recomendaciones:

- Dar opciones de personalizar los controles
- Agregar más desafíos
- Mejorar los modelos de personajes, dándole más expresividad
- Colocar un jefe final que permita probar los conocimientos adquiridos

Valoración final Jugabilidad EnglishPlay. La plantilla de evaluación heurística genera una valoración promedio por los atributos y facetas de jugabilidad. Cabe aclarar que no se hizo una evaluación de la faceta Interpersonal o Social debido a que EnglishPlay no cuenta con opciones de juego multijugador. La valoración se realiza de acuerdo a la siguiente categorización:

- Crítico: entre 1.0 y 2.9
- Medio: entre 3.0 y 3.9
- AcepTable.: entre 4.0 y 4.4
- Excelente: entre 4.5 y 5.0

Table. 6. *Valoración de Jugabilidad – EnglishPlay*

Atributos / Facetas	Satisfacción	Aprendizaje	Efectividad	Inmersión	Motivación	Emoción
Intrínseca	*3.99*	*4.8*	*3.33*	*4.17*	*4.28*	*4.45*
Mecánica	*4.43*	*2.75*	*5*	*4.86*	*2.75*	*5*
Interactiva	*4.38*	*4.14*	*4*	*4.26*	*3.66*	*3.64*
Artística	*4.1*	*4.13*	*5*	*3*	*2.17*	*3.56*
Personal	*4.42*	*4.33*	*4.67*	*4.5*	*4.63*	*4.25*

Para el análisis se toman los promedios de las facetas y atributos. Para el caso de los atributos

- Satisfacción: 4.26
- Aprendizaje: 4.03
- Efectividad: 4.4
- Inmersión: 4.16
- Motivación: 3.5
- Emoción 4.18

Y las facetas:

- Intrínseca: 4.17
- Mecánica: 4.13
- Interactiva: 4.01
- Artística: 3.66
- Personal: 4.47

La valoración final del videojuego es ***4.09***

De acuerdo con los resultados vistos en la tabla 6 y a los promedios calculados, la valoración de EnglishPlay en base a su jugabilidad se encuentra en una acepTable. experiencia de jugador.

Por último, se dejan algunas conclusiones sobre algunos de los valores críticos mostrados en la valoración:

- **Jugabilidad Mecánica - Aprendizaje:** El resultado se debe a que el videojuego no tiene cambios de dificultad entre los distintos retos.
- **Jugabilidad Mecánica - Motivación:** Se encuentra que los retos solo tienen una forma de superarse o completarse.
- **Jugabilidad Intrínseca - Efectividad:** Esto se debe a que el videojuego no cuenta con retos secundarios, se sugiere crear algunos retos extra que permitan diversificar los objetivos del videojuego.

- **Jugabilidad Interactiva - Emoción:** Indica que el personaje puede ser mejorado integrando más expresiones en el personaje.

- **Jugabilidad Artística - Inmersión y Motivación:** Se muestra la falta de una narrativa.

VI. CONCLUSIONES

Se desarrollaron 8 desafíos y 4 mundos virtuales donde son integrados dichos desafíos. Estos minijuegos incluyen reglas y mecánicas que permiten a los estudiantes aprender mientras juegan, lo que les ayuda a salir del ambiente normal de aula, para entrar a un ambiente virtual donde se pueden divertir y estar más motivados en el aprendizaje.

Se evidencia que el diseño de este tipo de videojuegos debe permitir que el mismo usuario aprenda las mecánicas jugando, ya que el uso de tutoriales en las pruebas realizadas fue mínimo.

La plantilla de evaluación heurística fue utilizada para la evaluación de los mundos y minijuegos de EnglishPlay y además puede ser usada de referencia para la creación de plantillas de evaluación heurística aplicadas a otras propuestas de videojuegos.

Mediante la realimentación proporcionada por los estudiantes, se concluye que la inclusión de los videojuegos en la enseñanza de inglés podría convertirse en un factor clave para aumentar la motivación de los mismos. También, durante las pruebas se evidencio que los estudiantes se divertían jugando y aprendiendo, que estaban motivados para culminar el videojuego y en lugar de aprender gramática de una manera rutinaria, comenzaron a aplicarla de forma interactiva y didáctica.

TRABAJOS FUTUROS

- EnglishPlay cuenta con 4 mundos y 1 nivel por mundo, teniendo en cuenta los diseños realizados en la selección de niveles de cada mundo, se evidencia la necesidad de ampliar la oferta de niveles, utilizando temas de inglés no incluidos aún o ampliando los ya incluidos.

- Desarrollar más desafíos con otras mecánicas, estrategias didácticas y metodologías para la enseñanza de inglés tales como desafíos que impliquen la enseñanza de pronunciación o la escucha en inglés.

- Incluir una narrativa en el videojuego que permita describir sucesos de la historia a medida que se avanza, además, integrar objetivos secundarios que estén ligados a nuevas historias o a una continuación de la narrativa.

REFERENCIAS

[1] *Desurvire, H., Caplan, M., Toth, J.A.: Using heuristics to evaluate the playability of games. In: CHI '04 Extended Abstracts on Human Factors in Computing Systems. pp. 1509-1512. CHI EA '04, ACM, New York, NY, USA (2004). https://doi.org/10.1145/985921.986102*

[2] *Federoff, M.A., Federoff, M.A.: Heuristics and usability guidelines for the creation and evaluation of fun in video games. Tech. rep., Indiana University, Bloomington (2002)*

[3] *Gonzales, M.P., Pasucal, A., Lorés, J.: This is a test entry of type @ONLINE (2001), https://aipo.es/libro/pdf/15-Evaluacion-Heuristica.pdf, accedido 22- 04-2018*

[4] *González Sánchez, J.L.: Jugabilidad. Caracterización de la experiencia del jugador en videojuegos. Ph.D. thesis, Universidad de Granada, http://digibug.ugr.es/handle/10481/5671 (2010)*

[5] *Inc., W.: Ruby rei: Juego de aprendizaje de idiomas. https://www.wibbu.com (2018), accedido 22-04-2018*

[6] *Nacke, L.: Affective Ludology: Scientific Measurement of User Experience in Interactive Entertainment. Ph.D. thesis, Blekinge Institute of Technology, Karlskrona, Sweden (2009)*

[7] *OREALC: Enfoques estratégicos sobre las TICS en educación en América Latina y el Caribe:. Unesco (2014)*

[8] *Prensky, M.: Digital game-based learning. Comput. Entertain. 1(1), 21-21 (Oct 2003). https://doi.org/10.1145/950566.950596*

[9] *Prensky, M.: Listen to the natives. Educational Leadership 63(4) (dec 2005)*

[10] *Smeets, E., Mooij, T.: Pupil-centred learning, ict, and teacher behaviour: observations in educational practice. British Journal of Educational Technology 32(4), 403-417 (2001). https://doi.org/10.1111/1467-8535.00210*

[11] *Yalena Aileth Venega Cañizares, Carlos Alberto Gavilla Cruz: Videojuegos serios como herramientas para la enseñanza. Serie Científica de la Universidad de las Ciencias Informáticas 9(3) (2016), http://publicaciones.uci.cu/index.php/SC/article/view/1733*

Biografía Autor(es)

a. Ingeniero de Sistemas. Docente, Universidad del Valle. E-mail: luis.fernando.quintero@correounivalle.edu.co. Cali, Colombia.

b. Ingeniero de Sistemas. Docente, Universidad del Valle. E-mail:jorge.l.zapata@correounivalle.edu.co . Cali, Colombia.

c. Ingeniero de Sistemas.Docente, Universidad del Valle. E-mail: andres.bueno@correounivalle.edu.co. Cali, Colombia.

d. Ingeniero de Sistemas.Docente, Universidad del Valle. E-mail: javier.naranjo@correounivalle.edu.co. Cali, Colombia.

e. Magister en Ingeniería énfasis en Ingeniería de Sistemas. Docente, Universiddad del Valle. E-mail: carlos.andres.delgado@correounivalle.edu.co. Cali, Colombia.

f. Magíster en administración.Docente, Universidad del Valle. E-mail: yuri.bermudez@correounivalle.edu.co. Cali, Colombia.

g. Magíster en Ingeniería énfasis en Ingeniería de Sistemas.Docente, Universidad del Valle. E-mail: edwin.gamboa@correounivalle.edu.co. Cali, Colombia.

Estudio de caracterización de la Web Colombiana

Arturo Eraso Torres (a)

Resumen

Uno de los más grandes éxitos de Internet es basarse en un modelo abierto, transparente y colaborativo lo que le ha permitido obtener un gran crecimiento. Dentro de ésta organización la ICANN (Internet Corporation for Assigned Names and Numbers) tiene la responsabilidad de coordinar la asignación de nombres y números como identificadores únicos en todo el mundo, en términos más técnicos se encarga de los Sistemas de Nombres de Dominio (DNS), el Protocolo IP, los dominios genéricos (gTLD) y los correspondientes a los dominios de nivel superior geográfico (ccTLD), entre otros. El dominio de nivel superior para Colombia es el ".co", que actualmente es gestionado por concesión del Estado Colombiano a la entidad .CO INTERNET SAS. Esta Web nacional se encuentra caracterizada y abarca una gran variedad de aspectos como las políticas de gobierno, las capacidades tecnológicas, los conocimientos específicos, entre otros; temas importantes y relevantes de estudio, de los cuales se tiene referentes como los de Brasil[1], España[2], Argentina[3] y Chile[4].

Palabras clave: Cable, minería web, rastreador.

Characterization study of the Colombian Web

Abstract

One of the greatest successes of the Internet is to be based on an open, transparent model and collaborative which has allowed it to obtain great growth. Within this organization ICANN (Internet Corporation for Assigned Names and Numbers) has the responsibility to coordinate the assignment of names and numbers as identifiers unique in the world, in more technical terms it is in charge of the Systems of Domain Names (DNS), the IP Protocol, generic domains (gTLD) and corresponding to the geographic top level domains (ccTLD), among others. He top level domain for Colombia is ".co", which is currently managed by concession of the Colombian State to the entity .CO INTERNET SAS. This national website is characterized and encompasses a wide variety of aspects such as policiesgovernment, technological capabilities, specific knowledge, among others; important and relevant study topics, of which there are references such as from Brazil [1], Spain [2], Argentina [3] and Chile [4].

Keywords: Wire, web mining, crawler.

I. INTRODUCCIÓN

El estudio de investigación sobre la Web de Colombia se realizó mediante la herramienta JWIRE, la cual funciona en base al crawler WIRE (Web Information RE-trieval) y una interfaz gráfica implementada en Java, el primero fue desarrollado por Ricardo Baeza-Yates y Carlos Castillo en la Universidad de Chile, posteriormente se lo dotó de una interfaz gráfica proporcionándole así un manejo sencillo y rápido, este aporte lo realizó Luis Alberto García Hernández de la Universidad de Salamanca, con ello se ha podido esTable.cer unas características más específicas del ciberespacio Colombiano, partiendo desde el tipo de páginas, lenguajes y documentos soportados, hasta aspectos generales del colectivo denominados sitios y terminando con algunas características específicas de lo que se han denominado dominios de Internet.

II. LA SEMILLA EL PUNTO DE PARTIDA

El proceso de obtención de la información de la Web de Colombia se realizó mediante el crawler, que es un software que permite descargar el contenido de una lista de sitios predeterminada a la cual se le denomina semilla.

En junio de 2016 la Web Colombia registró 2.225.640 dominios distribuidos como se observa en la Tabla 1, el que presenta un mayor número es el dominio de segundo nivel (.co), con un porcentaje del 92.92%. En segundo lugar se encuentran las agencias comerciales (.com.co), con el 5.68%. En tercer lugar el correspondiente a dominios educativos (.edu.co) con el 0,425%. En cuarto lugar están los gubernamentales (.gov.co), con el 0,3655%. Posteriormente los proveedores de servicios de internet (.net.co) con el 0,3039%.

Table. 1. *Distribución de dominios de la Web Colombia*

Nombre	Público	Número de dominios	%
.co	Dominio de segundo nivel	2.068.065	92,92
.com.co	Comercial	126.41	5,68
.edu.co	Instituciones educativas	9.459	0,425
.gov.co	Agencias gubernamentales	8.135	0,3655
.net.co	Proveedores de servicios de internet	6.764	0,3039
.org.co	Organizaciones no gubernamentales	2.820	0,12669
.nom.co	Personal	2.410	0,1083
.mil.co	Agencias militares	1.561	0,07012
.cctld.co	Servidores del dominio	6	0,00025
.info.co	Información	2	0,0001
.int.co	Internacional	2	0,0001
.nic.co	Administración del dominio	1	0,00004

En la sexta posición se encuentran las organizaciones no gubernamentales (.org.co), que representa el 0,12669%, en la séptima los dominios de tipo personal (.nom.co) con el 0,1083%, en la octava las agencias militares (.mil.co) con el 0,07012%, en la novena posición los dominios correspondientes a los servidores para Colombia (.cctld.co) con el 0,00025%, en la décima y onceava, los correspondientes a información (.info.co) e internacional (.int.co) con el 0,0001% cada uno y finalmente el dominio correspondiente al nic Colombia (.nic.co) con el 0,00004%.

Para la realización del presente estudio relacionado con los dominios registrados para la Web de Colombia .co, se debe mencionar que se recibió información por parte del administrador .CO INTERNET SAS correspondiente al total de dominios registrados, pero no se proporcionó el listado de dominios que éste lo componen, por lo cual fue necesario:

Recurrir a OSINT (inteligencia de fuentes abiertas), término que hace referencia a cualquier información desclasificada y públicamente accesible en Internet de forma gratuita, usando para ello herramientas como theharvester y recon-ng, configurando sus opciones frente a buscadores como: google, googleCSE, bing, bingapi, pgp, linkedin, jigsaw y googleplus.

Posteriormente se cruzó esta información con un listado de direcciones web adquiridas en la empresa Country Ip Blocks (https://www.countryipblocks.net), para aumentar y mejor las características y calidad de la semilla.

De acuerdo a lo anterior se esTable.ció la siguiente distribución de la semilla:

Table. 2. *Distribución de la semilla*

Nombre	Público	Número de dominios	%
.co	Dominio de segundo nivel	1.709.777	99,5681
.com.co	Comercial	3.976	0,2315
.edu.co	Instituciones educativas	901	0,0525
.gov.co	Agencias gubernamentales	1.806	0,1052
.net.co	Proveedores de servicios de internet	67	0,0039
.org.co	Organizaciones no gubernamentales	590	0,0344
.nom.co	Personal	5	0,0003
.mil.co	Agencias militares	71	0,0041
.cctld.co	Servidores del dominio	0	0,0000
.info.co	Información	0	0 0,0000
.int.co	Internacional	0	0 0,0000
.nic.co	Administración del dominio	1	0,0001

Como se puede observar en la Tabla 2, el dominio con mayor cantidad de sitios es el dominio de segundo nivel (.co) con 1.709.777 direcciones, que representan el 99,5681% de la semilla. En segundo lugar está el dominio comercial (.com.co) con 3.976 direcciones, que forman el 0,2315%. En tercer lugar está el dominio de agencias gubernamentales (.gov.co) con 1.806 direcciones, que representan el 0,1052% del total de la semilla. En el cuarto lugar se tiene a las instituciones educativas (.edu.co) con 901 direcciones, que constituyen el 0,0525% de los sitios.

En el quinto lugar se tiene a las .org.co con 590 direcciones y una representación del 0,0344% en relación con las direcciones del tipo en la semilla. En el sexto lugar se tiene a las agencias militares con 71 direcciones, que constituyen el 0,041%. En el séptimo lugar los proveedores de servicios de internet con 67 direcciones (.net.co) que corresponden al 0,0039% de la semilla, en el octavo está el dominio (.nom.co) correspondiente a dominios personales con el 0,0003% y con una cantidad de 5 direcciones. Finalmente los dominios pertenecientes a internacional (.int.co), información (.info.co), administración del dominio (.nic.co) y direcciones de servidores de dominio (.cctld.co) constituyen el 0,0001% de las direcciones en la semilla.

III. DESCARGA DE INFORMACIÓN

Una vez definida la semilla, se pasó a determinar las características necesarias del crawler para llevar a cabo la caracterización de la Web Colombia, en ésta se tuvieron en cuenta dos aspectos: las características de la máquina y la configuración del crawler.

3.1 Características de la máquina

El crawler es un programa que no es difícil en su instalación y ejecución, sin embargo dependiendo de la cantidad de información que se vaya a descargar, se convierten en aplicativos exigentes frente a los recursos que la máquina debe soportar con a los objetivos trazados. Teniendo en cuenta los aspectos necesarios para la instalación del crawler, se procedió en base a la semilla a realizar los recorridos de recolección de los datos correspondientes a la Web de Colombia, por lo cuál los requerimientos mínimos para realizar dicho proceso, fueron los siguientes:

- Procesador: Intel core i3 o superior.
- Memoria RAM: mínimo 2 Gb, recomendable 4 Gb.
- Espacio en disco duro: Mínimo 50 Gb de espacio libre, dependerá de la cantidad de sitios a descargar.
- Sistema operativo: Recomendable Linux (Para el estudio aquí relacionado se utilizó Ubuntu GNU/Linux).
- Configuración del Crawler

La herramienta JWIRE basa su funcionamiento en el crawler WIRE, el cual se compone de un conjunto de aplicaciones las cuales se usan tanto para iniciar el propio crawler como para procesar los datos recogidos por éste, analizarlos posteriormente y darles un formato de salida, algunos de los aspectos considerados fueron:

1. Editar el archivo de configuración del crawler, que contiene una serie de propiedades cuyos valores son atributos de la ejecución de sus programas.
2. Editar el archivo start_urls.txt con las URLs de inicio para el crawler.
3. Iniciar el repositorio donde serán almacenados los datos extraídos.

4. Ejecutar el número de ciclos necesarios para la realización de la tarea correspondiente al crawler.

5. Una vez finalizado el trabajo de las aplicaciones de recolección de datos, estos se almacenan en varios directorios y archivos que componen el repositorio de datos. Se pueden usar varias herramientas de JWIRE para procesar y analizar los datos. Las alternativas son:

   a. Usar las opciones del JWIRE para ejecutar una serie de aplicaciones para obtener los diferentes resultados en modo texto.

   b. Ejecutar la herramienta de extracción de datos de JWIRE para mostrar los datos obtenidos tras la ejecución del crawler en modo texto y por la salida estándar.

   c. Ejecutar la herramienta de análisis de datos de JWIRE que formatea los datos originales a archivos en formato CSV (comma-separated values) y después usar otra herramienta JWIRE adicional para, a partir de los archivos CSV, generar informes en PDF y LaTeX con tablas y gráficos de las relaciones de los datos obtenidos.

Hay más características que se podrían explicar y enumerar en la configuración del crawler, pero las más significativas son las relacionadas en los ítems anteriores, y se pueden observar con mayor detalle en la dirección: http://it.luisalbertogh.net/jwire.htm/.

IV. CARACTERÍSTICAS DE LA DESCARGA

De acuerdo con las sección anterior respecto a la máquina y configuración del crawler, se debe decir que cuando éste inició la descarga, pudo haber situaciones que no dependieron ni de la maquina, ni de las configuraciones esTable.cidas, sino de estados de los sitios que hicieron que fuese imposible descargarlos, en algunos casos se debió a que las direcciones relacionadas correspondían a páginas que ya no existen o que fueron dadas de baja. Este proceso a nivel técnico lo maneja el crawler con un código de estado, el cual indica si la página existe o no, o si detectó un motivo por el cual no se pudo entregar el documento pedido. La Tabla 3 muestra algunos de estos estados.

Table. 3. *Código de estados*

Categorías de Códigos	***Significado***
Ok	*Respuesta estándar para peticiones correctas.*
Error Connect	*Conexión rechazada.*
Not Found	*El recurso solicitado no se pudo encontrar, pero puede estar disponible de nuevo en el futuro.*
Moved	*Al recurso solicitado se le ha asignado una nueva URI permanente y todas las referencias futuras deben utilizar este recurso.*
Forbidden	*La solicitud fue legal, pero el servidor se rehúsa a responderla.*

Table. 4. *Código de estados en la Web Colombia*

Categorías de Códigos	Cantidad	Porcentaje
Ok	*29.461.278*	*59,25*
Error Connect	*4.655.940*	*9,36*
Not Found	*4.632.870*	*9,32*
Moved	*3.478.968*	*7,0*
Found	*3.381.873*	*6,8*

En la Tabla 4, se puede observar la distribución de las páginas con base en los estados encontrados, las direcciones con respuestas correctas corresponden al 59,25%, que permite concluir que no hubo mayores problemas durante el proceso de conexión y esTable.cimiento de la comunicación, encontrándose dentro de los rangos acepTable.s. El 9,36% de la descarga se encontró con errores de conexión, el 9,32% corresponde a recursos que fueron movidos o cambiados permanentemente, el 7,0% referencia direcciones que no han sido encontradas y el 6,8% corresponde a recursos movidos temporalmente, el resto responde a otro códigos como errores de DNS, internos, parciales, entre otros.

V. RESUMEN GENERAL DE LA DESCARGA

Las características generales de la web de Colombia obtenidas mediante la herramienta JWIRE se presentan en la Tabla 5, de la cual se destacan las siguientes observaciones:

- Se recolectaron alrededor de 50.000.000 de páginas.
- El número de sitios referenciados en el proceso es de 460.535.
- El tamaño medio de un sitio en la web Colombia es de 2,32 MB.
- El promedio de páginas por sitio es de alrededor de 133 documentos.
- El promedio de las páginas estáticas y dinámicas por sitio es alrededor de 105 y 28 documentos respectivamente.
- La profundidad promedia máxima por sitio es de 1,49.
- La edad media de una página en meses es de alrededor de 21,41.
- El idioma predominante es el español con el 38,65%.
- Los formatos gráficos más utilizados son el jpg y el png con el 46,28% y el 25,60% respectivamente, las extensiones de audio más empleado son es el mp3 (77,47%), seguido del formato au con el 17,01%.

En general se descargaron 49.725.459 páginas web, desde 460.535 sitios. Si se compara estos valores con estudios de otras web nacionales, se encuentra que la cantidad

de páginas es superior a Chile con 7.403.840[4], España con 6.171.267[2] y Paraguay con 214.548[5].

A nivel de sitios es superior a Chile 171.213[4], España 308.822[2] y Paraguay con 2.812 Sitios [5].

Table. 5. *Resumen general de la Web Colombia*

Páginas web recolectadas	*49.725.459*
Número de sitios Ok	*460.535*
Número de sitios con una edad de páginas validas	*188.3047*
Promedio de enlaces internos	*1.018,49*
Promedio de páginas por sitio	*133,14*
Promedio de páginas estáticas por sitio	*104,82*
Promedio de páginas dinámicas por sitio	*28,33*
Promedio máximo de profundidad de un sitio	*1,49*
Promedio del tamaño por sitio en MB	*2,32*
Promedio de edad de la página más antigua en meses	*33,37*
Promedio de edad de la página media en meses	*21,41*
Promedio de edad de la página más reciente en meses	*19,40*
Páginas web recolectadas	*49.725.459*
Número de sitios Ok	*460.535*
Número de sitios con una edad de páginas validas	*188.304*
Promedio de enlaces internos	*1.018,49*

VI. CARACTERÍSTICAS DE LAS PÁGINAS

En esta sección se presenta un análisis a nivel de las páginas, sin considerar ningún tipo de agrupación. Se disponen de diferentes criterios de estudio como edad, idioma, tipos, lenguajes, documentos multimedia, binarios, comprimidos, entre otros.

6.1 Edad de las páginas en años

Para determinar la edad de la páginas, se analiza el valor del encabezado "last-modified" del protocolo HTTP. En algunos casos la fecha es errónea, esto se debe a que los servidores no tienen sus relojes sincronizados con la hora actual del país o simplemente no han sido configurados. Por ejemplo, un caso es: 0000-00-00, que indica que no se tiene su registro.

De los datos recolectados se puede interpretar que más del 79,74% de la Web de Colombia ha sido creada o modificada en el último año, lo que indica que es un espacio con un alto nivel de crecimiento. Argentina reportó el 72%[3].

6.2 Idiomas

La distribución de los idiomas utilizados en documentos que componen el espacio de la Web de Colombia presenta un porcentaje de el 51,9% de idioma en inglés, seguido por el idioma español con el 46%. El español es el idioma oficial para Colombia y presenta niveles inferiores a otros países. En Chile se reportó con el 81,34% [4], en España con 62%[2] y en Argentina del 63%[3] del idioma español. Lo que demuestra la apertura del dominio Colombiano al resto del mundo, fruto de la gobernanza dispuesta por el Ministerio de las Tecnologías de la Información y las Comunicaciones de Colombia, el Comité Asesor y .CO INTERNET SAS.

6.3 Tipo de páginas

Para el estudio de la Web de Colombia se dividió el conjunto de datos obtenido en dos grupos, uno denominado páginas dinámicas, las cuales tienen las características de ser generadas bajo petición, son capaces de responder de manera inteligente a las demandas de un cliente y permiten automatizar gran cantidad de tareas. Por lo general son el resultado de una consulta a una base de datos y son manejadas por programas, estas páginas contienen formularios y métodos de envío de datos, según sea éste, se pueden observar variables viajando junto con los enlaces, separadas por el símbolo de interrogación "?" y de enlace "&" para concatenar elementos y sus valores.

Para su detección se utilizó como criterio analizar la extensión de los documentos relacionados con los script de los lenguajes utilizados más frecuentemente como PHP, ASP, CGI, etc.

Las páginas estáticas son las mas comunes y simples, aquellas en las cuales su contenido poco o nada cambia en el tiempo. Sus extensiones son generalmente HTML y HTM.

En el análisis de datos expuesto en la Tabla 6, se puede observar que la mayoría de páginas son de tipo estático (75,43%), mientras el 24,57% corresponde a páginas dinámicas. La Web Argentina reportó el 52%[3] de páginas dinámicas, en contraste España el 22%[2] y Chile el 42,5%[4]. El porcentaje de la Web de Colombia de páginas estáticas es bastante elevado.

Table. 6. *Distribución de dominios dinámicos y estáticos.*

	Documentos	***Porcentaje***
Total	*49.725.459*	*100*
Estáticas	*75,43*	*75,43*
Dinámica	*24,57*	*24,57*

6.4 Lenguajes

Complementariamente, se analizó la distribución de los enlaces a documentos con las extensiones utilizadas para construir páginas dinámicas. Se aprecia una importante

participación del lenguaje de preprocesamiento de hipertextos PHP con un 72,61%, éste porcentaje es inferior al estudio de la Web de Chile con el 75%[4] y superior a los de España con el 46,24%[2], Brasil con el 73%[1] y Argentino con el 52%[3]; seguido al PHP está el asp con 24,32% el cuál es propietario, entre otros.

6.5 Documentos multimedia

En cuanto a enlaces de archivos multimedia, se organizaron en grupos de audio, video e imágenes con relación a los tipos de archivos mas representativos de cada uno. Respecto a las imágenes se puede interpretar que los formatos más usados en la Web de Colombia son los jpg, png y gif con el 94,7%; esto está acorde con la Web chilena con el 97%[4] y la Web Argentina con el 97,8%[3].

Los formatos de audio más empleados en la Web de Colombia son el mp3 (77,47%) y el au (17,01%). Sus equivalentes en el estudio de la Web Argentina son 35,9% y 39,4%[3] respectivamente.

Los formatos de video más empleados en la Web de Colombia son el formato avi (43,16%) seguido del mov (27,31%). En Argentina se reportaron 31,7% para wmv y el 21,9% para mpg [3].

6.6 Binarios y comprimidos

El mayor porcentaje de archivos comprimidos en la Web de Colombia son de tipo zip alcanzando el 46,3%, seguido del bz2 con 40,42%. En Chile el formato más usado es el gz (53,8%), seguido del zip con el 37,7%[4]. En Argentina el más usado es el zip con el 70,6% seguida del gz con el 15,5%[3].

La distribución de archivos binarios en la Web de Colombia en su mayoría está bajo formato exe ó ejecuTable. (99,95%). En Argentina el exe alcanza el 86%[3].

VII. CARACTERÍSTICAS DE LOS SITIOS WEB

En esta sección se caracteriza los Sitios Web ó conjunto de páginas que comparten la parte del nombre del servidor de la URL.

7.1 Nombres de los sitios

También se estudió la cantidad de partes que tienen los nombres de sitios, es decir, cada componente separada por un punto. Por ejemplo, el sitio www.ef.com.co contiene 4 partes mientras que acueductodebucaramanga.com.co tiene 3.

Se observa que en la Web de Colombia más del 99% de los nombres de los sitios poseen 2 partes ya que Colombia utiliza un dominio de segundo nivel, lo cuál se puede apreciar en la estructura de la semilla.

7.2 Promedio de páginas por sitio

El promedio de páginas por sitio para la Web de Colombia es de 133 el cuál es superior a los estudios realizados en Chile con 43,24[4], España con 52,08[2] y Argentina con 65[3].

De los sitios encontrados se observa que muchos contienen una sola página, dentro de los motivos están:

- La página es sólo una redirección a otro sitio.
- La página es efectivamente un sólo sitio.
- La página contiene solamente enlaces externos.
- La navegación de la página está basada en Javascript.

7.3 Profundidad por sitio

La profundidad determina los niveles jerárquicos, en los cuales se puede estructurar un sitio, sirven para dar el orden necesario para un buen manejo y administración, esta en manos del administrador de un sitio Web. La página principal de un sitio por lo general esta en el nivel uno.

La mayor parte de los sitios esTable.cen su contenido en el primer nivel (75.91%), seguido del segundo nivel con el 10,63%, esto puede denotar que se tienen poca organización de contenidos, lo que facilita su administración o por el contrario sobre explotación de un recurso.

7.4 Cantidad de enlaces

Este apartado hace referencia al estudio de las relaciones esTable.cidas a nivel de enlaces. El espacio Web es modelado como un grafo dirigido sobre el cual se analizan diferentes características.

Grado entrante

El número de enlaces que recibe una página Web se denomina grado entrante (indegree). Esta característica determina el ranking, el cuál es una medida de popularidad de las páginas. De acuerdo con la información recolectada con el crawler se obtuvo

que la Web de Colombia a partir de la semilla aplicada tiene 60.091 enlaces entrantes. La Tabla 7, muestra los 10 primeros sitios de mayor grado entrante.

Table. 7. *Distribución de los diez primeros Sitios con mayor grado entrante.*

Num	Sitio	Grado Entrante
1	www.portafolio.co	98
2	www.cz100.co	66
3	www.90vs.co	66
4	www.1soccer.co	66
5	www.66365.co	60
6	www.55365.co	60
7	www.39365.co	60
8	www.38365.co	60
9	www.37365.co	60
10	www.35365.co	60

Grado saliente

El número de enlaces que recibe una página Web se denomina grado entrante (indegree). Esta característica determina el ranking, el cuál es una medida de popularidad de las páginas. De acuerdo con la información recolectada con el crawler se obtuvo que la Web de Colombia a partir de la semilla aplicada tiene 60.091 enlaces entrantes. La Tabla 8, muestra los 10 primeros sitios de mayor grado entrante.

Table. 8. *Distribución de los diez primeros Sitios con mayor grado entrante.*

Num	Sitio	Grado Entrante
1	www.portafolio.co	98
2	www.cz100.co	66
3	www.90vs.co	66
4	www.1soccer.co	66
5	www.66365.co	60
6	www.55365.co	60
7	www.39365.co	60
8	www.38365.co	60
9	www.37365.co	60
10	www.35365.co	60

VIII. CARACTERÍSTICAS DE LOS DOMINIOS

Finalmente en esta sección se presenta los resultados de los dominios de la Web de Colombia.

8.1 Estructura macroscópica

Como se puede interpretar en la Tabla 9, la Web de Colombia se encuentra débilmente interconectada. Esto se debe a que el MAIN solo alcanza el 0,0093%, en comparación con la Web Argentina que posee el 54,23%[3], quizá se puede ver valores un poco mas más cercanos en la Web chilena 21,76%[4] y en Brasil que alcanza el 25,27%[1]. Los Sitios que componen el OUT representan el 0,0074% que es un valor inferior al de Chile (26,12%[4]) y al de Brasil con el 45,33%[1].

Table. 9. *Componentes de la estructura macroscópica.*

Componente	Sitios	Porcentaje
PRINCIPAL – NORMAL	*0*	*0 0,0000*
PRINCIPAL – PRINCIPAL	*33*	*0,0072*
PRINCIPAL - ENTRANDO	*0*	*0,0000*
PRINCIPAL – SALIENDO	*10*	*0,0022*
PRINCIPAL	*43*	*0,0093*
ENTRADA	*9*	*0,0020*
SALIDA	*34*	*0,0074*
Tentáculos de Entrada	*0*	*0 0,0000*
Tentáculos de Salida	*1*	*T0,0002*
Túnel	*0*	*0,0000*
Islas	*460.448*	*99,9811*
Total	*460.535*	*100*

Los Sitios correspondientes a IN e ISLANDS se los accede a partir de sus páginas iniciales, debido a que pueden ser páginas nuevas o no estar bien conectadas. En la Web de Colombia estos valores se representan por 0,093% para IN y 99,98% para ISLANDS. En Chile se representa por 6,65% para el componente IN y 46,16% en ISLANDS[4]. En Brasil por 12,95% y 12,35%[1] respectivamente.

8.2 Software usado como servidor

Mediante la recolección de la información con el Crawler, se pudo identificar la información con respecto a los diferentes tipos de servidores Web que se encontraron.

El servidor que presenta mayor uso es Apache con el 72,61%, luego sigue Microsoft IIS (asp) con el 14,92%, y luego otros tipos de servidores no tan conocidos. En Chile el 66,7% de los servidores que ejecutan Apache, mientras que el 32,8% ejecuta Microsoft IIS[4] y en España la presencia de Apache es de 46,89% por 38,88% en Microsoft ISS[2].

IX. CONCLUSIONES

El presente estudio ha conllevado procesos básicos de recolección y análisis de información, mediante los cuales se pudo llegar a especificar las características de la web colombiana, los criterios que se tuvieron en cuenta en el estudio fueron:

- En primera medida se presentan los resultados en forma individual, es decir a nivel de páginas, sin tener en cuenta algún factor de agrupamiento y con base a ellos se analiza sus metadatos como son: el tamaño, contenido y enlaces.

- Posteriormente se interpreta los resultados a nivel de sitios, es decir se analiza un conjunto de páginas que comparten parte del nombre de un servidor mediante su URL, como por ejemplo: http://www.iucesmag.edu.co y http://uvirtual.iucesmag.edu.co, y con base a ellos se obtienen estadísticas por medio de este factor de agrupamiento.

Acorde a lo expuesto anteriormente se pudo determinar las principales características de la Web Colombia, con base a una población de 460.535 sitios y 49.725.459 páginas, con lo que se pudo determinar que la distribución del idioma está compuesta en un porcentaje del 51,9% en inglés, seguido del español con un 46% que es el idioma oficial del país, lo que demuestra que un gran número de dominios registrados corresponden a empresas o personas que están fuera de Colombia o de empresas o personas con expansión internacional. Además aproximadamente el 79,74% de las páginas han sido creadas o modificadas en el último año, lo que demuestra que el espacio Web colombiano está creciendo como ocurre en otros países.

En cuanto a los aspectos tecnológicos el 75,43% del total de las páginas descargadas son de tipo estático y el 24,57% dinámico. Los cuales se encuentran construidas en gran parte por herramientas de tipo libre como PHP con el 72,61% y Asp con 24,32%.

Se puede interpretar y acorde a la muestra que la Web Colombia se encuentra débilmente interconectada. Un indicador es que el componente MAIN solo alcanza el 0,0093% que está por debajo de los indicadores de otras Web de estudio como la Argentina. Sumado a esto existe un alto porcentaje correspondiente a IN e ISLANDS 0,0020% y 99,9811% que refuerza esta idea. Por lo cual se debe mejorar las estructuras de enlaces entre los sitios de la Web nacional colombiana para lograr mayor visibilidad.

La web es un espacio cambiante y Colombia no es ajena a este tipo de modificaciones, la cantidad de información presente en ella hace que esta varíe rápidamente obedeciendo a las tendencias de mercado gestionadas por el administrador de la misma. Se observa que si bien la web crecido en su número de enlaces, la mayoría corresponden a dominios de segundo nivel, que sirven básicamente como redireccionadores a otros dominios adquiridos a otros administradores de la web y que se adquieren por tener acceso a nombres más cortos que faciliten su memorización y uso.

Se puede observar que los resultados obtenidos presentan características similares a otros estudios realizados en otros países, lo que indica que los resultados son al menos parcialmente extrapolables a la realidad de la red global.

Mediante el estudio se pudo recolectar información verdaderamente importante referente al proceso de análisis de una web nacional, entre los cuales cabe destacar la información referente a la configuración del crawler, el cual se convierte en una guía para futuros estudios ya que permite tener una referencia en la asignación de valores a las variables involucradas en el estudio. De igual manera se coloca a disposición la información correspondiente a la semilla, que puede ser un insumo para la realización de estudios posteriores, ya que la esta información es escasa.

Finamente, respecto al uso de JWIRE presenta una gran ventaja la utilización de una herramienta con un entorno gráfico, ya que:

- Facilita la comprensión del proceso, aprendizaje y uso.
- El diseño mediante menús y pestañas permite un rápido acceso.
- Las interacciones se realizan con el uso de botones, iconos y mensajes.
- Existe un complemento adecuado respecto a las ayudas y consultas.
- Permite personalizar el trabajo mediante el agrupamiento de funciones en archivos xml.

REFERENCIAS

[1] *MODESTO, Marco, et. al. Um novo retrato da web brasileira. In Proceedings of SEMISH, São Leopoldo, Brazil, 2005.*

[2] *BAEZA-YATES, Ricardo; CASTILLO, Carlos y LÓPEZ, Vicente. Características de la web de España. El professional de la información.*

[3] *TOLOSA, Gabriel, et. al. Caracterización del espacio Web de Argentina. Universidad Nacional de Luján. 2007.*

[4] *BAEZA-YATES, Ricardo; CASTILLO, Carlos. Características de la Web Chilena 2007. Technical Report, Center for Web Research, University of Chile, 2008.*

[5] *BORDIGON, Fernando; LAVALLEN, Pablo; TOLOSA, Gabriel. "El Estado de la Web de Paraguay y la Sociedad de la Información". Universidad Nacional de Luján. 2006.*

Biografía Autor

a. Doctor en Ciencias de la Electrónica. Docente tiempo completo, Universidad Cesmag. E-mail: aeraso@unicesmag.edu.co. Pasto, Colombia.

Evaluación del componente práctico en la enseñanza de programación I en el programa de Ingeniería de Sistemas de la Universidad de Nariño

Manuel Ernesto Bolaños González (a)
Nelson Antonio Jaramillo Enríquez (b)

Resumen

Las características de un buen ingeniero de sistemas se reflejan en la forma como plantea, analiza, diseña e implementa posibles soluciones a un problema sistémico o computacional, y es el resultado del análisis, el que determina el éxito de la solución, ya que es éste el que esTable.ce los componentes sistémicos y las relaciones entre estos para obtener una solución, en donde, el diseño y la implementación, fundamentan su acción en los planteamientos especificados por el análisis.

En la Universidad de Nariño en el programa de Ingeniería de Sistemas, conscientes de lo anterior, se hace una evaluación de las actividades académicas relacionadas con soluciones computacionales a problemas propuestos, con el objetivo de determinar el nivel de análisis alcanzado por los estudiantes, encontrando, que el nivel demostrado está determinado por la forma como estructuraron en la mente, los fundamentos de programación.

La evaluación permite analizar dos procesos de aprendizaje, determinados estos, por la forma como se realiza el componente práctico, el primero utilizando Materiales Educativos Computacionales y el otro utilizando el lenguaje de programación C. Los resultados llevan a determinar que, el nivel de análisis alcanzado por los estudiantes que utilizaron los materiales educativos computacionales (MEC) es superior al nivel de análisis alcanzado por estudiantes que utilizaron el lenguaje de programación como el Lenguaje C.

El presente estudio muestra la forma como se llega a esta conclusión utilizando para su explicación las teorías del procesamiento de información y la importancia de los MEC en la enseñanza de la programación.

Palabras clave: Materiales educativos computacionales, análisis, procesamiento de información.

Component evaluation practical in teaching programming I in the program Systems engineer the University of Nariño

Abstract

The features of a good systems engineer are reflected in the way he or she poses, analyzes, designs and implements possible solutions to a systemic or computational problem. Such features are the result of the analysis, which determines the success of the solution, since establishing the systemic components and the relationships among them to obtain a solution, which the design and the implementation are the basis.

In the Systems Engineering program at University of Nariño, considering the above, an evaluation of the academic activities related to computational solutions to proposed problems is made, in order to determine the level of analysis reached by the students, finding the level demonstrated is determined by the way they structured the mind in programming fundamentals.

The evaluation allows us to analyze two learning processes, determined by the way in which the practical component is carried out, the first one is using Computational Educational Materials and the second one is using the C programming language. The results lead to determine that the level of analysis achieved by the students who used the computational educational materials–CEM– is higher than the level of analysis reached by students who used the programming language such as C Programming Language. The present study shows the way in which this conclusion is reached, using for its explanation the theories of the information processing and the importance of the CEM in the teaching of the programming fundamental.

Keywords: Computational educational materials, analysis, information processing.

I. INTRODUCCIÓN

La Universidad de Nariño con su programa de Ingeniería de Sistemas, busca formar profesionales capaces de desarrollar soluciones a problemas de sistemas y computación, para lo cual en su plan de estudio en los primeros semestres tienen temáticas que llevan a los estudiantes a desarrollar competencias que les permita planear, analizar, diseñar y desarrollar soluciones a un problema planteado, en donde, las habilidades analíticas, son las mas importantes al momento de entrar a detallar el procedimiento aplicado en la solución.

Los fundamentos de programación presentes en la asignatura Programación I de este programa, se presentan como la primera experiencia que tiene el estudiante con la computación, y depende de la forma como se guíe la apropiación de esta, el nivel de análisis alcanzado. Para determinar esta expectativa, se revisó como se llevan a cabo la asignatura programación I encontrándose diferencias en la metodología, fundamentadas estas, principalmente en la forma como se abarca el componente práctico; aspecto que incide, según el estudio, en forma significativa en la solución de un problema planteado y por tanto en el nivel de análisis encontrado.

El estudio muestra que en el componente teórico se desarrolla y ajusta a unos lineamientos esTable.cidos, pero, el componente práctico se lo realiza de dos formas, con materiales educativos computacionales como el DFD [1], Pseint [2], por un lado y con lenguajes de programación C# por otro.

Con este conocimiento, los grupos de Investigación del programa de ingeniería de Sistemas, determinan las características mentales que el estudiante desarrolla en las metodologías encontradas. El estudio muestra que dentro de estas características, se presenta una combinación de lo heurístico y lo algorítmico, mas acentuado lo heurístico, en las metodologías que asumen la práctica con un lenguaje de programación y lo algorítmico, en quienes asumen la práctica con materiales educativos computacionales (MEC).

Entender el por qué? de este comportamiento y dar una explicación de lo que sucede, es muy complicado, pues hay que adentrarse en las capacidades mentales desarrolladas por el estudiante, en especial, en las relacionadas con el análisis, para lo cual, se utilizan la teoría del procesamiento de información, ACT de Anderson (Control Adaptativo del Pensamiento) [3], y el principio de la percepción humana, con las que se trabajan el concepto de "unidades de información" presentes en la memoria de trabajo cuando se quiere dar solución a un problema planteado.

La estrategia de formación por cada metodología lleva a plantear que cada solución trabaja con tres modelos: mental, computacional y algorítmico, y cada modelo resultante está determinado por un número de unidades de información producto de los conceptos que el estudiante maneja, lo cual permite llegar a un resultado cuantificable de la solución.

Con este conocimiento, se enfoca el estudio en dos grupo de estudiantes, un grupo por cada metodología encontrada, con el propósito de evaluar la asimilación de conceptos básicos, la aplicación de los mismos y la capacidad analítica alcanzada en estos momentos, obteniéndose que la metodología más recomendable, es la que utiliza materiales educativos computacionales Dfd o PSeint en su práctica.

Este artículo, va a detallar las prácticas normales para obtener una solución computacional, los enfoques metodológicos, las estrategias de aprendizaje, para obtener una solución computacional, relacionados con el procesamiento de información y los resultados encontrados, primero en la solución de ejercicios de asimilación de

conceptos básicos, segundo en la solución de ejercicios de aplicación de estos conceptos básicos y tercero el comportamiento del componente analítico, presente en la solución de todos los ejercicios. La explicación de los resultados y las conclusiones se plantean a la luz de las teorías del procesamiento de información.

II. PRÁCTICAS METODOLÓGICAS DE APRENDIZAJE DE LOS FUNDAMENTOS DE PROGRAMACIÓN EN EL PROGRAMA

2.1 Estrategias de aprendizaje

Dentro del proceso de formación de los futuros profesionales en el campo de la programación se incorpora para este estudio a los autores Luis Joyanes y Osvaldo Cairó quienes plantean: Según Joyanes [5] "La principal razón para que las personas aprendan lenguajes y técnicas de programación es utilizar la computadora como una herramienta para resolver problemas. La resolución de problemas exige al menos los siguientes pasos: Definición, Análisis del problema, Diseño del algoritmo, Transformación del algoritmo en un programa, Ejecución y validación del programa".

Por otro lado Cairó, [6] plantea lo siguiente, "No existen reglas claras que nos permitan resolver un problema. Sin embargo, creemos que podemos ofrecer un conjunto de técnicas y herramientas metodológicas que permitan flexibilizar y estructurar el razonamiento en la solución de un problema. Eso provocará finalmente la construcción de algoritmos eficientes. Las etapas que debe seguir la construcción de un algoritmo son: Problema, Análisis profundo del problema, Construcción del algoritmo, Verificación del algoritmo".

Los pasos presentados llevan a esTable.cer dentro del desarrollo de una solución a un problema, una serie de modelos que se van transformando en el avance de esta, de ahí que, se plantea para el estudio y como estrategia de aprendizaje, que para llegar a la solución de un problema, se deben construir los modelos mental, computacional, algorítmico, producto de un análisis profundo y para el caso de la metodología con lenguajes de programación otro modelo más, el programa como solución de un problema.

Dónde: el modelo mental es el planteamiento de una solución del problema sin involucrar elementos computacionales, es un ejemplo de una solución, el modelo computacional es el modelo anterior sumado componentes computacionales (instrucciones de entrada, salida, decisiones, ciclos etc.), el modelo del algoritmo es la organización de los anteriores pasos en un orden lógico, y el modelo del programa es la codificación en el lenguaje de programación del algoritmo resultante.

2.2 Practicas normales para obtener una solución computacional en el programa.

Después de realizar la respectiva observación del programa que tienen en su plan de estudios formación en programación, se encuentra, que los fundamentos de la lógica de programación, presentes en la asignatura Programación I, se asume por

parte de los docentes siguiendo una estrategia de aprendizaje común, pero, el componente práctico se lleva a cabo en forma diferente, tal como se planteó en el punto anterior. De esta revisión se identifica las siguientes metodologías:

- Metodología con Material Educativo Computacional – Dfd o PSeint.
- Metodología con Lenguaje de Programación C#.

2.3 Enfoques de las metodologías

La observación del desarrollo del curso de programación I, permite determinar que la forma en que se desarrolla el componente práctico del curso favorece dos enfoques mentales fundamentados en la heurística y en lo algorítmico y por tanto dos metodologías. Estas son:

- **Metodología con enfoque algorítmico-heurístico.** En esta metodología predomina el enfoque algorítmico [4] en donde el estudiante trata de seguir una serie de pasos que le permiten llegar en a una solución. Pero combina también elementos heurísticos para abordar algunas etapas del diseño, en especial, las entradas y salidas, las cuales las realiza utilizando muchas de las características del lenguaje de programación.
- **Metodología con enfoque algorítmico.** En esta metodología el estudiante sigue una serie de pasos claramente determinados que le permiten llegar a una solución. Los MEC le permite al estudiante llegar hasta la prueba del algoritmo siguiendo lo que se construye en clase, y aunque brinda algunas posibilidades para abordar las entradas y salidas, son mínimas.

III. TEORÍA DEL PROCESAMIENTO DE INFORMACIÓN

Para el desarrollo del presente estudio se tienen los siguientes componentes teóricos pertenecientes a la teoría cognoscitiva del procesamiento de información:

3.1 Teoría del Control Adaptativo del Pensamiento (ACT) de Anderson

La teoría de Control Adaptativo del Pensamiento (ACT) de Anderson [3], se convierte en el referente de éste estudio. Con base en ésta se esTable.ce que, cada una de las metodologías encontradas estructuran los nuevos aprendizajes de acuerdo con la estrategia esTable.cida, con el propósito de incorporarlos a las redes de conocimiento presentes en la memoria permanente del estudiante, mas concretamente en las memorias declarativa y procedural, para que éste, sea capaz de activarlos en la memoria de trabajo cuando necesite dar solución a un problema. El componente teórico asociado con la estrategia de aprendizaje del curso de programación I y las características prácticas del lenguaje de programación y los materiales educativos computacionales (MEC) crean en la memoria permanente del estudiante, redes de nodos declarativas y de producciones con elementos estrechamente relacionados con las características

de las herramientas, compuestas por diversas unidades cognoscitivas jerárquicamente ordenadas que van a ser activadas en la memoria de trabajo cuando la solución de un problema lo requiera.

3.2 Principio de la teoría de procesamiento de la información y percepción humana

Para definir las características básicas de los conceptos presentes en la solución de cada ejercicio, realizado por cada metodología, se hace uso del principio de la teoría del procesamiento de la información relacionado con la percepción humana y la memoria de corto plazo, conocida también como memoria de trabajo. El principio de percepción plantea que en la memoria de trabajo es limitada y que tan solo puede mantener a la vez siete "unidades de información" [7]. Lo que contiene cada "unidad" depende de la forma como la persona captó u organizó la información.

Las "unidades de información" se las define como el espacio de memoria de trabajo que se debe destinar para poder entender y trabajar con un concepto.

Para el desarrollo del presente estudio se elabora la tabla 1, la cual determina las unidades de información que puede tener cada concepto que va a participar en la solución de un problema, y estas unidades, van para los tres modelos que deben desarrollarse en la solución de un problema computacional, según la estrategia de aprendizaje planteada, el modelo mental, el modelo computacional y el modelo algorítmico.

Por cada ejercicio desarrollado para el estudio, y después de la evaluación y aplicación de la tabla 1 se va a obtener una cantidad de unidades de información, producto de la aplicación de los diferentes conceptos que participan en la solución, obteniendo así, un valor cuantitativo que va a permitir comparar las dos metodologías. Para el caso de la práctica con lenguaje de programación se obtiene también los elementos del modelo programa y juega con la columna elementos del lenguaje.

Table. 1. *Unidades de información de conceptos básicos.*

	Unidades de Información			
Concepto	***Elementos Mentales***	***Elementos Computacionales***	***Elementos Algoritmo***	***Elementos Lenguaje***
Variable	*1*	*1*	*1*	*1*
Definición	*2*	*2+*	*2+*	*2+*
Asignación	*2+*	*2+*	*2+*	*2+*
Entrada	*1+*	*1+*	*1+*	*5+*
Salida	*1+*	*1+*	*1+*	*5+*
Condición	*4+*	*3+*	*3+*	*3+*
Decisión	*5+*	*5+*	*5+*	*5+*
Contador	*3*	*4+*	*4+*	*4+*
Acumulador	*4+*	*4+*	*4+*	*4+*
C. mientras	*6+*	*6+*	*6+*	*6+*
Ciclo para	*6+*	*5+*	*5+*	*6+*
Arreglo	*4*	*2+*	*2+*	*2+*
Otros				*1 por cada variable*

IV. EVALUACIÓN Y RESULTADOS

La apropiación de los conceptos básico con la estrategia esTable.cida y la aplicación de los mismos en la solución de problemas más complejos, se lo realiza de la siguiente forma: para la evaluación de la asimilación de los conceptos básicos el desarrollo de siete (7) ejercicios por 10 estudiantes de cada metodología; para la evaluación de la aplicación de los mismos, el desarrollo por los mismos estudiantes de 3 ejercicios. Con estos ejercicios se identifican los modelos mental, computacional y algorítmico utilizados para la solución del problema y por cada modelo, se determinan las unidades de información utilizadas, para obtener una serie de valores que van a ser sometidos a interpretación y llegar así a unos valores consolidados que van a permitir esTable.cer comparativos y conclusiones.

La aplicación de la metodología de evaluación de los conceptos básicos su aplicación y el nivel de análisis generan los siguientes puntos.

4.1 Evaluación de la asimilación de conceptos básicos y su aplicación

Con esta clasificación se tienen elementos para determinar que pasa en la solución de cada uno de los ejercicios planteados y los modelos mental, computacional, algorítmico y de programas, que se generan con la solución si se sigue la estrategia de aprendizaje esTable.cida.

Para la interpretación de los resultados se determina que el grupo 1 corresponde a la metodología que incorpora el lenguaje de programación en el componente práctico del curso y el grupo 2 quienes utilizan los MEC.

De la aplicación de la metodología se sacan las tablas de resultados y su respectivo análisis, tabla 2 para el grupo 1, y tabla 3 para el grupo 2.

El puntaje total de la tabla hace referencia a la cantidad de unidades de información presentes en los modelos mental, computacional y algorítmico de la solución planteada por los estudiantes a los siete ejercicios y el porcentaje representa el porcentaje con respecto a la cantidad de unidades que se esperaban sean utilizados.

La evaluación tiene valores de MB para Muy Bien aplicados, B para Bien, R para regular y M para Malo.

Las tablas 2 y 3 muestran la evaluación de conceptos básicos, del Grupo 1 y Grupo 2 respectivamente. Presentan un consolidado de la evaluación.

Table. 2. *Tabla de evaluación conceptos básicos. Grupo 1*

No. Estudiante	Puntaje Total	Porcentaje	Evaluación
1	74	60	R
2	77	62	R
3	101	81	B
4	120	97	MB
5	31	25	M
6	99	80	B
7	37	30	M
8	99	80	B
9	77	62	R
10	42	34	M

Table. 3. *Tabla de evaluación conceptos básicos. Grupo 2*

No. Estudiante	Puntaje Total	Porcentaje	Evaluación
1	111	90	B
2	109	88	B
3	72	58	R
4	122	98	MB
5	69	56	R
6	38	31	M
7	122	98	MB
8	108	87	B
9	123	99	MB
10	42	34	M

Análisis de resultados y comparativos de la asimilación de conceptos básicos. De las tablas 2 y 3 se obtiene la tabla 4 que contiene un resumen de estos resultados, la cual permite esTable.cer un comparativo entre los dos grupos.

Table. 4. *Resumen resultado conceptos básicos*

Grupo	MB	B	R	M
Grupo 1	1	3	3	3
Grupo 2	3	3	2	2

De la tabla 4 se obtienen las figuras 1 y 2 en donde se esTable.ce por cada grupo el porcentaje de participación en el rango de evaluación esTable.cido (ver Fig. 1 y 2). También de esta tabla se realiza la figura 3 en la cual se puede observar el rendimiento de los dos grupos (ver Fig. 3).

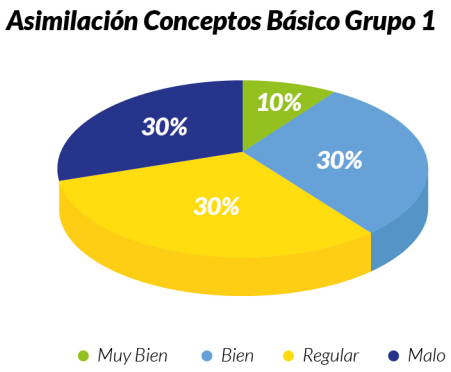


Fig. 1. *Porcentaje de participación grupo1.*

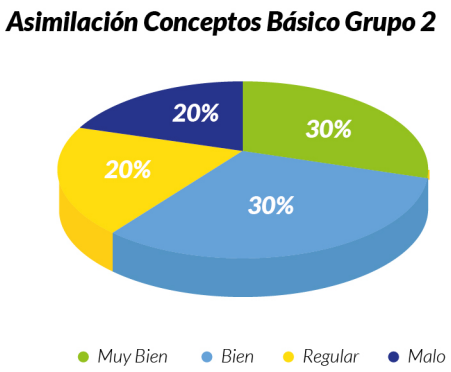


Fig. 2. *Porcentaje de participación grupo2.*

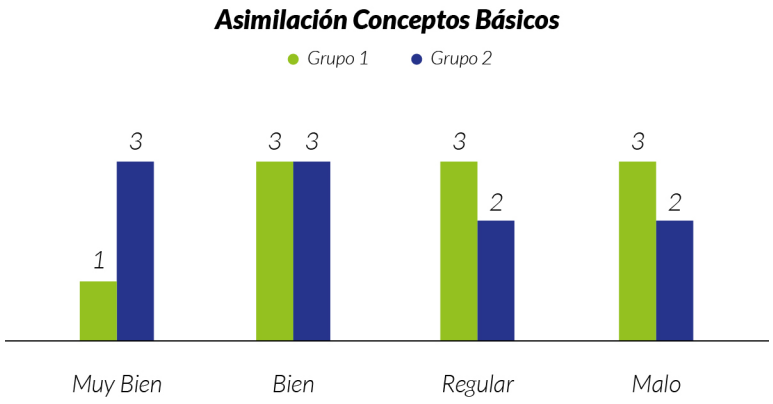


Fig. 3. *Comparación de asimilación conceptos básicos grupo 1 y grupo 2.*

La tabla resumen 4 y las gráficas anteriores permiten esTable.cer lo siguiente:

- El rendimiento presentado en los conceptos básicos por los estudiantes del grupo 2, es mejor que el mostrado por el grupo 1.

- Los estudiantes ubicados en las evaluaciones de MB y B son los que demuestran la asimilación y aplicación correcta de los conceptos básicos de acuerdo a la metodología propuesta.
- Los estudiantes con resultados de R y M no alcanzan a dominar los conceptos básicos de acuerdo a la metodología propuesta.

4.2 Evaluación de aplicación de conceptos

Las tablas 5 y 6 muestran la evaluación de conceptos aplicados y procedimiento de solución de los grupos 1 y grupo 2 respectivamente.

Table. 5. *Tabla de evaluación de aplicación de conceptos. Grupo 1*

No. Estudiante	Puntaje Total	Porcentaje	Evaluación
1	84	48	R
2	87	50	R
3	153	88	B
4	165	95	MB
5	38	22	M
6	127	73	B
7	44	25	M
8	152	87	B
9	97	56	R
10	45	26	M

Table. 6. *Tabla de evaluación de aplicación de conceptos. Grupo 2*

No. Estudiante	Puntaje Total	Porcentaje	Evaluación
1	152	87	B
2	154	89	B
3	100	57	R
4	157	90	B
5	50	29	R
6	41	24	M
7	168	97	MB
8	152	87	B
9	172	99	MB
10	35	21	M

Análisis de resultados y comparativos de conceptos aplicados. De las tablas anteriores se obtiene la tabla 7 que contiene un resumen de estos resultados, la cual permite esTable.cer un comparativo entre los dos grupos.

Table. 7. *Resumen de resultados de ejercicios de aplicación.*

Grupo	MB	B	R	M
Grupo 1	1	3	3	3
Grupo 2	2	4	1	3

De esta tabla se obtienen las gráficas 4 y 5 donde se esTable.ce por cada grupo el porcentaje de participación en el rango de evaluación esTable.cido (ver Fig. 4 y 5). También de esta tabla se realiza la gráfica 6 en la cual se puede observar el rendimiento de los dos grupos (ver Fig. 6)

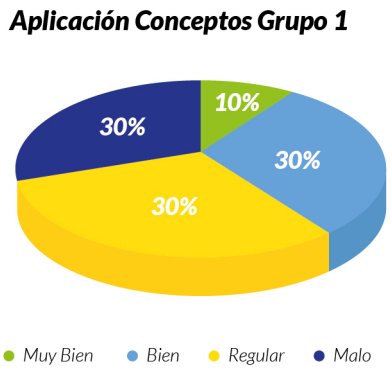


Fig. 4. *Ponderados aplicación de conceptos grupo1*

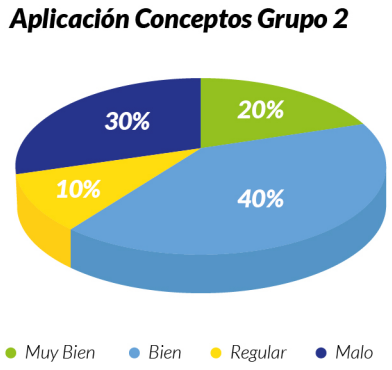


Fig. 5. *Ponderados aplicación de conceptos grupo 2*

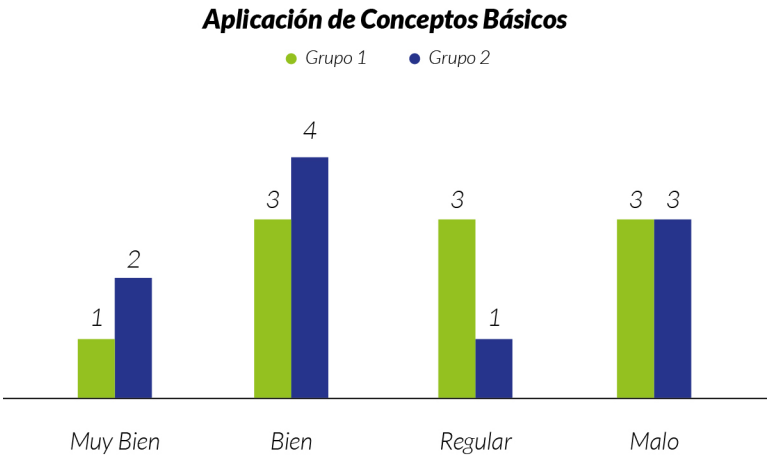


Fig. 6. *Comparación aplicación de conceptos grupo 1 y grupo 2*

La tabla resumen y las gráficas anteriores permiten esTable.cer lo siguiente:

- El rendimiento presentado en la aplicación de conceptos por los estudiantes del grupo 2, es mejor que el mostrado por el grupo 1.
- Los estudiantes ubicados en las evaluaciones de MB y B son los que muestran la aplicación correcta de los conceptos en la solución de problemas de acuerdo con la metodología propuesta.
- Los estudiantes con resultados de R y M no aplican los conceptos de acuerdo con la metodología propuesta.

4.3 Evaluación del análisis en la solución del problema

Teniendo en cuenta que el análisis es el proceso de pensamiento que incluye las siguientes operaciones mentales, por parte del estudiante: descomposición de un todo en sus partes, determinación de las relaciones entre las partes y comprensión de la forma como están organizadas.

La evaluación del análisis se enfoca en los tres modelos planteados por la estrategia de aprendizaje y en los resultados obtenidos con el punto anterior en la evaluación de la asimilación de conceptos básicos y la aplicación de los mismos para obtener: análisis de elementos, análisis de las relaciones identificadas en la solución y análisis de la construcción relacional de la solución, presentes en una solución en sus tres modelos.

De la aplicación de la metodología se sacan las tablas 8 y 9 correspondientes al análisis presente en cada solución y los comparativos de los dos grupos, donde MB es Muy Bueno, B es Bueno, R es Regular y M es malo.

Table. 8. *Tabla de proceso de análisis grupo 1.*

No. Estudiante	Identificación de Elementos	Relación de Elementos	Construcción Relacional
1	M	R	R
2	M	R	R
3	B	B	B
4	MB	MB	MB
5	M	M	R
6	R	B	B
7	M	M	M
8	R	B	R
9	R	R	B
10	M	M	M

Table. 9. *Tabla de proceso de análisis grupo 2.*

No. Estudiante	Identificación de Elementos	Relación de Elementos	Construcción Relacional
1	R	B	B
2	B	B	B
3	R	R	R
4	MB	MB	MB
5	M	M	M
6	M	M	M
7	MB	B	B
8	B	B	B
9	MB	MB	MB
10	M	M	M

Con estas tablas se obtienen las tablas comparativas 10, 11 y 12, y de estas las figuras comparativas de los grupos 1 y 2 del proceso de análisis (ver Fig. 7, 8, 9).

Table. 10. *Tabla para Identificación de elementos*

Grupo	MB	B	R	M
Grupo 1	1	1	3	5
Grupo 2	3	2	2	3

Table. 11. *Tabla de relación de elementos*

Grupo	MB	B	R	M
Grupo 1	1	3	3	3
Grupo 2	2	4	1	3

Table. 12. *Tabla para la construcción relacional.*

Grupo	MB	B	R	M
Grupo 1	1	3	3	3
Grupo 2	2	4	1	3

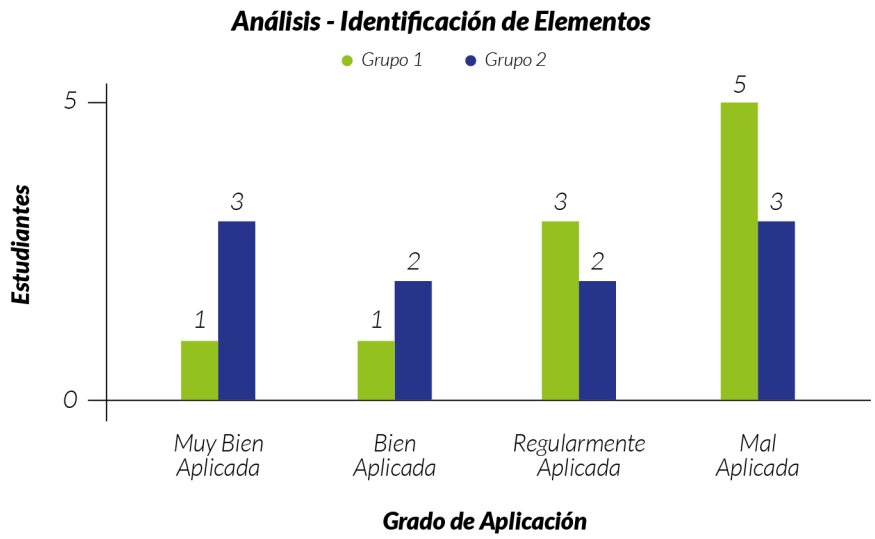


Fig. 7. *Proceso de análisis - Identificación de elementos.*

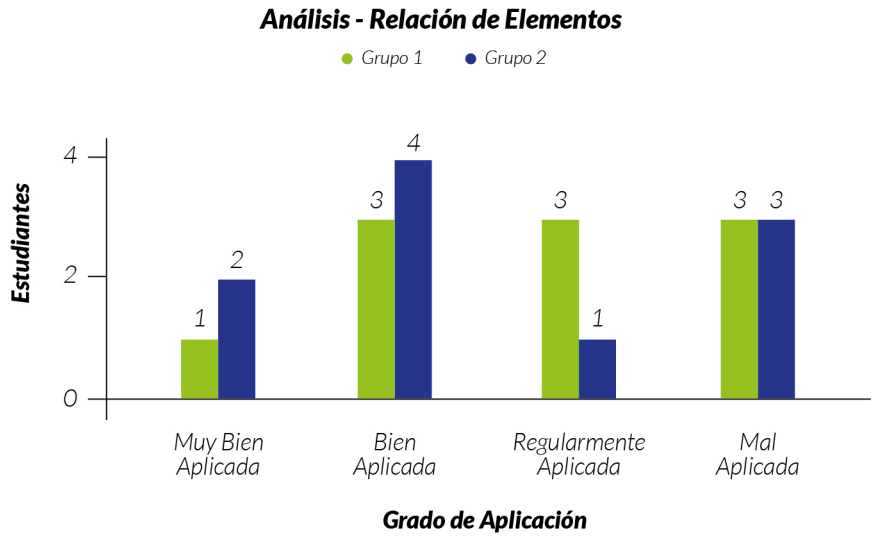


Fig. 8. *Proceso de análisis - Identificación de elementos.*

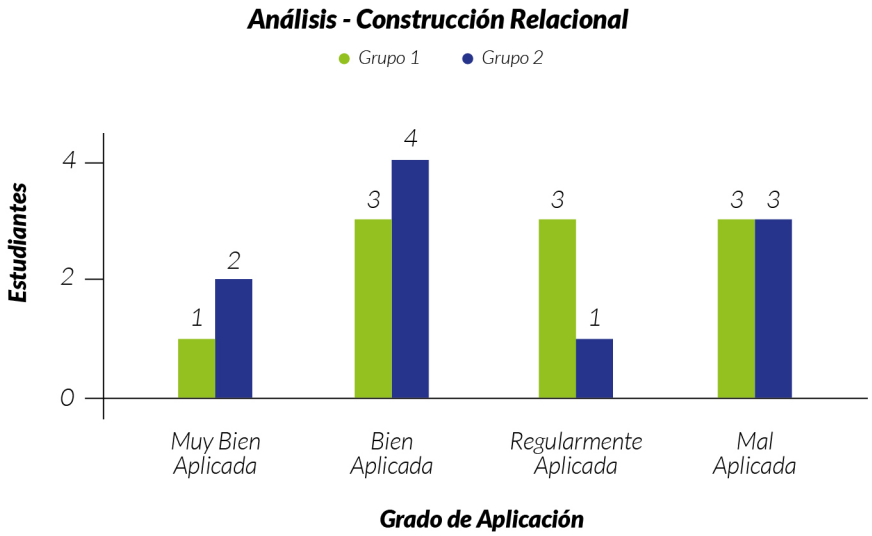


Fig. 9. *Proceso de análisis - Identificación de elementos.*

De las gráficas resultantes se puede esTable.cer lo siguiente:

- El grupo 2 que utiliza MEC para la parte práctica identifica los elementos de solución de un problema más eficientemente que el grupo que utiliza lenguaje de programación. Lo cual indica que el grupo 2 identifica de mejor forma los elementos presentes en ejemplo, el modelo mental y el modelo computacional que se deben plantear para la solución de un problema.

- Las relaciones que se presentan entre el ejemplo solución de un problema y el modelo computacional son mejor trabajadas por el grupo 2.

- De acuerdo con la figura 3, los algoritmos resultantes realizados por el grupo 2 presentan mejores características relacionales que los presentados por el grupo 1.

- Las tres comparaciones realizadas muestran que los estudiantes del grupo 2 presentan en sus ejercicios aplicativos, mejores características analíticas que los ejercicios presentados por los estudiantes del grupo 1.

V. CONCLUSIONES

El ambiente en el cual se mueve el estudiante determina en gran medida las estructuras que se van a relacionar en la memoria permanente.

El tener menos elementos dentro del ambiente de aprendizaje como en el caso de la utilización de MEC, permite que se dedique más tiempo a las actividades mentales de análisis de un problema determinado antes de adentrarse al computador para probar su solución.

El estudio muestra que utilizar un MEC específico como el Dfd o Pseint para el componente práctico de un curso, permite al docente enfocarse única y exclusivamente en la construcción del algoritmo.

La enseñanza de los inicios de la programación con un lenguaje de programación formal como el C#, necesariamente exige una dedicación del tiempo para su enseñanza y asumir una serie de pasos adicionales, lo que va en contra del proceso de análisis del problema que se desarrolla.

Se debe favorecer un ambiente en donde la práctica se realice con herramientas genéricas como el Dfd o Pseint, que sea flexible a cambios tecnológicos y que aproveche en forma óptima todo el conocimiento.

Desde el punto de vista de una construcción eficiente y eficaz de una solución a un problema específico, se recomienda la utilización de MEC específicos en los inicios de la programación.

REFERENCIAS

[1] *https://sites.google.com/site/informaticaieensma/home/diagramas-de-flujo/dfd-1-0*

[2] *http://pseint.sourceforge.net/*

[3] *J. Pozo (1995).Teorías Cognitivas del Aprendizaje, Universidad Autónoma de Madrid 1995. POZO JUAN, Ignacio. Pág.119 • Salma.*

[4] *Marisol Santos Silva. Diferencia entre método algorítmico y heurístico. Consultado el 20 de Junio del 2016 en http://www.academia.edu/15831318/*

[5] *L. Joyanes (1999), Fundamentos de Programación. McGraw Hill/Interamericana de España S.A. España. pp 40.*

[6] *Cairó Osvaldo (1995) Metodología de la Programación. Computec. México DF. pp 5.*

[7] *Panqueva Galvis. Ingeniería de Software Educativos. Pág. 169*

Biografía Autor(es)

a. Doctor en Ciencias de la Electrónica. Docente tiempo completo, Universidad de Nariño. E: mail: mbolanos@udenar.edu.co. Pasto, Colombia.

b. Magister en Comunicación Educativa. Docente tiempo completo, Universidad de Nariño. E-mail: njaramillo@udenar.edu.co. Pasto, Colombia.

Evaluación heurística de usabilidad sobre el sistema gestor de base de datos (SGBD) de Sevenet de la Industria Licorera del Cauca (ILC)

Miller Andrés Mera Inga
Lisbeth Jurany Zemanate Zuñiga
Gabriel Elías Chanchí Golondrino
Katerine Márceles

Resumen

El auge de las nuevas tecnologías y el desarrollo exponencial de la red de redes ha hecho posible hoy aumentar el número de aplicaciones desplegadas en Internet, así como la necesidad de que estas aplicaciones se centren en los usuarios finales, desde la perspectiva de la usabilidad. Según ISO 9241-11, se define la usabilidad como la medida por la cual un producto puede ser utilizado por ciertos usuarios para lograr objetivos con eficacia, eficiencia y satisfacción en un contexto de uso concreto. La principal ventaja de la usabilidad es que permite a las empresas del área de software desarrollo y usuarios finales para ser más productivos. Uno de los más utilizados Los métodos de inspección para evaluar la usabilidad es la evaluación heurística, mediante la cual un grupo de los expertos analizan, exploran y evalúan los diferentes elementos de un producto de software, teniendo en cuenta un conjunto de principios heurísticos previamente definidos. Este documento presenta el desarrollo de una evaluación heurística de la gestión de la base de datos de Sevenet Sistema de la Industria Licorera del Cauca. Lo anterior con el fin de determinar las razones por qué dicho software se ha subutilizado dentro de la empresa.

Palabras clave: Diseño, evaluación, heurística, usabilidad.

Heuristic evaluation of usability about the system database manager (SGBD) of Sevenet of the Cauca Liquor Industry (ILC)

Abstract

The rise of new technologies and the exponential development of the network of networks has made it possible today to increase the number of applications deployed on the Internet, as well as the need for these applications to be focused on end users, from the perspective of usability. According to ISO 9241- 11, usability is defined as the measure by which a product can be used by certain users to achieve specific objectives with effectiveness, efficiency and satisfaction in a context of concrete use. The main advantage of usability is that it allows companies in the area of software development and end users to be more productive. One of the most commonly used inspection methods to evaluate usability is heuristic evaluation, whereby a group of experts analyze, explore and evaluate the different elements of a software product, taking into account a set of previously defined heuristic principles. This paper presents the development of a heuristic evaluation of the Sevenet Data Base Management System of the Liquor industry of Cauca. The above in order to determine the reasons why such software has been underutilized within the company.

Keywords: Design, evaluation, heuristics, usability.

I. INTRODUCCIÓN

Hoy día, casi todos somos usuarios de sistemas informáticos, desde una anciana que nivela su libreta de ahorro en el banco hasta el dueño de un pequeño comercio que hace un pedido a su proveedor a través de internet. La expansión del uso de ordenadores a una gran parte de la sociedad ha hecho que la usabilidad del software adquiera una gran importancia en el desarrollo de sistemas informáticos [1]. Dentro de las ventajas y aportes de la evaluación heurística al campo del desarrollo del software se encuentran: es un atributo fundamental en el desarrollo de aplicaciones interactivas, ya que permite tener un desarrollo competitivo dentro de la industria del software al considerar a las personas como parte fundamental del éxito o fracaso de un producto software, permite el incremento en la productividad del usuario, cuenta con medidas para poder evaluarla, principios de diseño para lograr productos más usables y métodos para aplicarla [2]. De acuerdo al estándar ISO 9241-11, la usabilidad es entendida como el grado en el que un producto puede ser utilizado por usuarios específicos para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un determinado contexto de uso. La anterior definición describe los tres atributos (eficiencia, eficacia y satisfacción) que se deben considerar en cuenta para estimar el grado de usabilidad de un producto software [3].

La evaluación heurística es un método que consiste en verificar la calidad de una serie principios llamados principios heurísticos o simplemente "Heurística" previamente es-Table.cidos. Es llevada a cabo por evaluadores expertos en usabilidad de que actúan imitando las relaciones que tendría un usuario promedio al interactuar con el sistema que se está evaluando en este caso es el Sistema Gestor de Base de Datos SEVENET [4-5].

En la Industria Licorera del Cauca es una empresa que cuenta con 70 empleados en la parte administrativa en diferentes áreas de trabajo, los empleados deben ingresar diariamente al sistema gestor de base de datos (SGBD) SEVENET, con el fin de cumplir determinadas tareas en la parte de gestión documental. A pesar de las ventajas ofrecidas por el sistema de gestión documental se ha detectado falencias o problemas en cuanto a que es una herramienta subutilizada, esto podría estar asociado a que posiblemente los usuarios presenten dificultad para entenderla y aprenderla; debido a esto el número de accesos es bajo. La Industria Licorera del Cauca (ILC) con base en su política de cero papel, se manifiesta inconforme respecto a la usabilidad del sistema, ya que es una herramienta que le genera a la empresa un gran gasto en cuanto a su servicio; y se pretende que sus empleados mejoren sus actividades en la parte de gestión documental a consecuencia de esto, se plantea realizar una evaluación heurística de usabilidad al sistema gestor de base de datos (SGBD) SEVENET de la Industria Licorera del Cauca (ILC), para así determinar cuáles son los aspectos a mejorar en cuanto a la usabilidad, funcionamiento y rendimiento del software dentro de la empresa.

En virtud de los resultados de la realización de la evaluación heurística, se espera lograr un impacto tecnológico, ambiental y económico, promoviendo el uso del software SEVENET y de este modo conseguir una reducción en la utilización de papel y tinta; que a su vez representa un costo significativo para la Industria Licorera del Cauca (ILC). Finalmente, este artículo está organizado de la siguiente forma: En la sección 1 se detalla la metodología usada para la evaluación heurística; en la sección 2 se presentan conjunto de conceptos y temáticas que se tuvieron en cuenta para el desarrollo del trabajo; en la sección 3 se describe la estructura metodológica para la evaluación heurística; en la sección 4 se presenta recomendaciones de la evaluación heurística por expertos al sistema gestor de base de datos (SGBD) SEVENET; en la sección 5 se presenta las conclusiones y trabajos futuros derivados del presente trabajo.

II. METODOLOGÍA

Durante el proceso del presente proyecto se utilizó una metodología para el desarrollo y ejecución de evaluaciones heurísticas de usabilidad. Dentro de las fases de la metodología (ver Fig. 1), se presentan las 7 fases de la evaluación heurística propuestas para esta metodología desarrollada en la Institución Universitaria Colegio Mayor del Cauca por ingenieros en informática con conocimientos en usabilidad. [6].

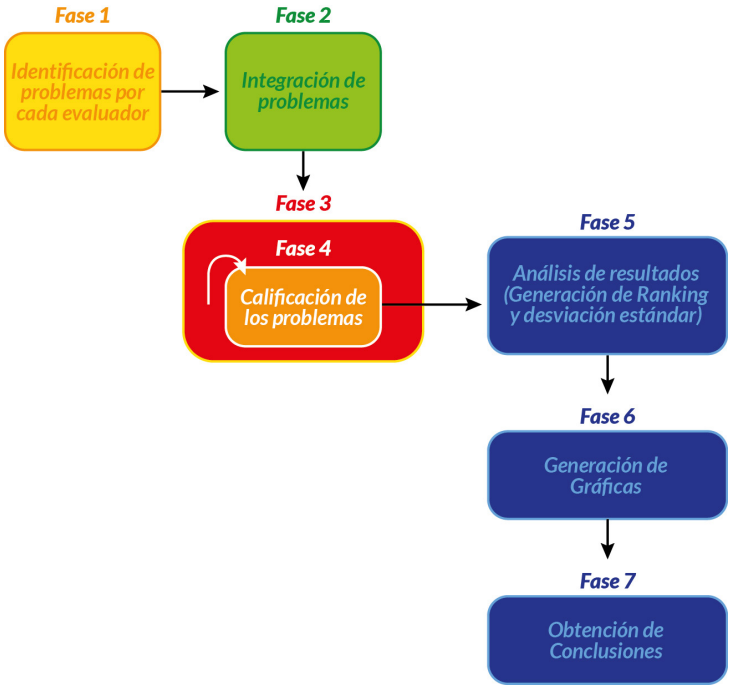


Fig. 1. *Fases de la evaluación heurística*

III. MARCO CONCEPTUAL

En esta sección se presentan un conjunto de conceptos y temáticas que hacen parte del desarrollo del presente trabajo. Dentro de esos conceptos están: Usabilidad, Sevenet, Heurísticas de Nielsen y Evaluación Heurística.

3.1 Usabilidad

La Organización Internacional para la Estandarización (ISO) ofrece dos definiciones de usabilidad. Según la ISO/IEC 9126: "La usabilidad se refiere a la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario,

en condiciones específicas de uso". Del mismo modo, ISO/IEC 9241: "Usabilidad es la eficiencia y satisfacción con la que un producto permite alcanzar objetivos específicos a usuarios específicos en un contexto de uso específico" [7]. En sí, el concepto proviene del inglés "usability", hace referencia a la facilidad con que un usuario puede utilizar una herramienta fabricada por otras personas con el fin de alcanzar un cierto objetivo.

La usabilidad está vinculada, por lo tanto, a la simpleza, la facilidad, la comodidad y la practicidad. En otras palabras, la noción tiene relación con la eficacia percibida de un objeto y la posibilidad de aprovechar todo su potencial [8]. Más concretamente aún, podemos determinar que uno de los principios básicos que se deben seguir a la hora de diseñar y crear un sitio web para que goce de la mejor usabilidad es pensar en el usuario. Es decir, tiene que estar conformado por y para el usuario. Se soporta que, si el sistema en cuestión no puede ser utilizado por el usuario en todo su potencial, no es útil y, por lo tanto, es deficiente. El grado de usabilidad se mide a partir de pruebas empíricas (con trabajo de campo entre múltiples usuarios) y relativas (la comparación con sistemas similares) [9].

3.2 Sevenet

Es un software especializado para la administración de documentos que le permite optimizar las funciones de archivo y flujo de procesos a partir de su digitalización. Los documentos son considerados el activo más importante dentro de una organización ya que constituyen la memoria institucional de todas sus actividades, por tal razón, se hace imprescindible brindarles el tratamiento correcto.

Es un componente especializado en la administración digital de archivo de forma fácil, confiable y segura que le permite: Conservar sus documentos en forma digital en una base de datos centralizada; implementar las TRD (Tablas de Retención Documental) en su empresa controlando el ciclo vital de la información y la disposición final de los documentos; crear estructuras de almacenamiento, índices de captura y consulta de acuerdo a sus necesidades; manejo del archivo de gestión, central e histórico; con el fin de cumplir con la normatividad vigente en Colombia [10].

3.3 Heurísticas de Nielsen

En 1990 Jacob Nielsen propuso 10 heurísticas de usabilidad para el diseño de aplicaciones interactivas, las cuales fueron obtenidas a partir del análisis de 249 problemas identificados en proyectos de diseño centrado en el usuario [11]. De este modo, las heurísticas propuestas por Nielsen son una respuesta a un conjunto de problemas comunes de usabilidad. Estas heurísticas en la actualidad se siguen utilizado para guiar el proceso de diseño de aplicaciones de escritorio y han sido adaptadas para ser aplicadas en diferentes contextos de aplicación (aplicaciones web, videojuegos, entornos educativos, etc.) [12]. A continuación, en la tabla 1 se presentan los diez principios heurísticos de Nielsen.

Table. 1. *Heurísticas de Nielsen*

Principios Heurísticos de Nielsen		
Id	**Principio**	**Descripción**
H1	*Visibilidad del estado del sistema*	*El sistema debe mantener siempre informados a los usuarios sobre lo que está pasando, mediante una realimentación apropiada en tiempos razonables.*
H2	*Relación entre el sistema y el mundo real*	*El sistema debe ser familiar para el usuario, hablar su idioma y orientarse hacia el perfil del usuario, mostrando la información de una manera natural y fácil de entender.*
H3	*Control y libertad de usuario*	*El sistema debe permitir deshacer y rehacer acciones al usuario, además de otorgarle salidas de emergencia. El usuario debe tener la posibilidad de deshacer acciones que hayan sido seleccionadas por error, y salir de estados no deseados.*
H4	*Consistencia y Estándares*	*Los usuarios no deberían preguntarse si diferentes palabras, situaciones o acciones significan lo mismo. Se deben seguir las convenciones de la plataforma.*
H5	*Prevención de Errores*	*El sistema debe diseñarse de tal forma para evitar que los errores ocurran, mostrando advertencias a los usuarios antes de que realicen alguna acción que pueda llevar a estados no deseados por el usuario, ocultando aquellas funciones que no se encuentran disponibles y puedan generar errores.*
H6	*Minimizar la carga de memoria del usuario*	*El usuario no debe verse obligado a recordar información de una parte del sistema a otra. Las opciones importantes deben ser claramente visibles, indicando instrucciones fáciles de recordar usando interfaces intuitivas.*
H7	*Flexibilidad y eficiencia de uso*	*El sistema debe diseñarse para que cualquier usuario pueda utilizarlo de forma fácil y sea capaz de alcanzar y lograr lo que desea. El sistema debe acomodarse a las distintas formas en que un usuario puede realizar una misma tarea, haciéndolo intuitivo y obteniendo el mismo resultado.*
H8	*Diseño estético y minimalista*	*El sistema debe mostrar información relevante, ocultando aquella que sea raramente necesaria. La información debe estar correctamente distribuida, sin sobrecarga de elementos, mostrando una interfaz simple al usuario.*
H9	*Ayuda al usuario para reconocer, diagnosticar y recuperarse de errores*	*Luego que ocurre un error, el sistema debe mostrar mensajes de errores sencillos, fáciles de entender para el usuario sugiriendo una solución adecuada al problema.*
H10	*Ayuda y documentación*	*El sistema debe ofrecer ayuda y documentación sobre cómo funciona la aplicación. Esta ayuda debe ser fácil de encontrar y de entender, debe ser precisa, no muy extensa y centrada en las tareas que realiza el usuario, mostrando una lista de pasos concreta sobre cómo utilizar el sistema.*

3.4 Evaluación Heurística

El primer paso de una evaluación heurística es la planificación. Durante toda esta etapa es fundamental tener en cuenta que cada contexto de uso (Universidades, Compañías de seguros, Entidades Bancarias, Administraciones Públicas, Etc.) posee sus normas o convenciones, las cuales deberían reflejarse en la interfaz de sus sitios y aplicaciones y en la forma de trabajo de sus usuarios. Es por ello que un paso crucial dentro de la etapa de planificación de la evaluación Heurística es la adecuación de cada criterio que se utilizara al contexto de uso al cual pertenece el sistema que se desea evaluar. [13]. La metodología de la evaluación Heurística fue presentada inicialmente por Nielsen y Molich. Esta metodología consiste en analizar la conformidad de la interfaz con unos principios reconocidos de usabilidad (la "heurística") mediante la inspección de varios evaluadores. Los autores solo plantearon que los evaluadores deberían poseer algún conocimiento sobre los principios de la usabilidad [14].

IV. EVALUACIÓN HEURÍSTICA REALIZADA

Las evaluaciones heurísticas son un método seguro para detectar que tan usables son determinados aplicativos, para que un usuario interactúe con ellos, dicha evaluación se realiza teniendo en cuenta los diez principios de Nielsen, los cuales permiten que la evaluación sea precisa y segura en cuanto a los aspectos más relevantes de la aplicación. Por ello hemos elegido estas heurísticas para hacer la evaluación del sistema gestor de base de datos (SGBD) Sevenet de la Industria Licorera del Cauca (ILC), debido a que las evaluaciones heurísticas son el método de inspección más usado y se realiza con un grupo de evaluadores expertos mínimo 3 máximo 5, los cuales evalúan una aplicación software y así verificar si cumplen los principios heurísticos de usabilidad. Nielsen dice "add more and more users, you learn less and less" debido a que se encontraran los mismos problemas una y otra vez, y no es obligatorio mantener la observación de las mismas cosas. Asegura también que después del quinto usuario no se encontrara nada nuevo y además dice que si se da el caso de ser heterogéneos se tendrá que poner más usuarios a realizar pruebas [15]. A continuación, se presenta las 7 fases en para el desarrollo de la evaluación heurística.

Fase 1: Análisis y Detención de Problemas Individuales.

En esta fase cada evaluador trabaja independientemente 1-2 horas inspeccionando el sistema basándose en los principios heurísticos, registrando de la manera más clara posible el problema de usabilidad detectado y agrupando de acuerdo al principio de usabilidad que incumple. Como resultado de este paso los evaluadores generaron el listado de problemas de la tabla 2 a partir del estudio del software Sevenet.

Table. 2. *Tabla de definición de Problemas*

Definición de Problemas Heurísticos			
Id	***Definición del problema***	***Comentarios Explicaciones***	***Principios Incumplidos***
P1	*La información no es presentada de forma simple, natural y en orden lógico.*	*La información no es clara para el usuario, pero hay que mejorar cierta y organizar de manera lógica la información.*	*H2 – Relación entre el sistema y el mundo real.*
P2	*El sistema no está diseñado en el idioma que indica.*	*Adecuar la parte de vocabulario a un contexto del mundo real.*	*H2 – Relación entre el sistema y el mundo real.*
P3	*El sistema no utiliza metáforas y controles de interfaz que corresponden con la realidad.*	*Hay que mejorar en ciertas partes en la interfaz e iconos de navegabilidad.*	*H2 – Relación entre el sistema y el mundo real.*
P4	*Las metáforas no son fáciles de comprender.*	*Corregir la duplicidad en información para así brindar una interacción más simple con el sistema.*	*H2 – Relación entre el sistema y el mundo real.*
P5	*En la aplicación no se explican elementos complejos.*	*No existe evidencia de alguna explicación de elementos del sistema.*	*H2 – Relación entre el sistema y el mundo real.*
P6	*El uso del vocabulario no es consistente en todo el sistema.*	*No es consistente ni entendible en su totalidad debido a que hace uso de términos desconocidos y en otro idioma.*	*H4 – Consistencia y estándares.*

P7	*Las interfaces que conforman el sistema no tienen un aspecto visual coherente.*	*Hay que mejorar la interfaz principal del sistema para que el usuario le sea más amigable.*	*H4 – Consistencia y estándares.*
P8	*Los símbolos utilizados no son comprensibles para el usuario y facilitan la interacción con el sistema.*	*No son tan claros los símbolos en la parte de ayuda o acciones.*	*H4 – Consistencia y estándares.*
P9	*No se usan colores estándares para vínculos (rojo para advertencia, verde para acción exitosa).*	*Hay que mejorar los colores en la interfaz y cambiar ese color rojo que no es adecuado.*	*H4 – Consistencia y estándares.*
P10	*La ubicación de las barras de navegación, herramientas de búsquedas y controles (botones) no siguen los estándares comunes establecidos.*	*Estandarizar los menús principales y de búsqueda que tiene la interfaz, botones.*	*H4 – Consistencia y estándares.*
P11	*No se le pide confirmación al usuario antes de realizar una acción crítica (cancelar, eliminar, aceptar).*	*Tener en cuenta para agregar al momento de realizar acciones importantes.*	*H5 – Prevención de Errores.*
P12	*El sistema no provee mensajes fáciles de entender que previenen posibles errores.*	*Hay que mejorar posibles errores o alertas a la hora de realizar acciones.*	*H5 – Prevención de Errores.*
P13	*Las opciones y/o funciones no son fáciles de encontrar.*	*En la página de inicio no muestra muchas opciones y encontrarlas es complejo.*	*H6 – Minimizar la carga de memoria del usuario.*
P14	*No existen atajos para realizar tareas frecuentes.*	*No existen atajos por tanto evidencia.*	*H7 – Flexibilidad y eficiencia de uso.*
P15	*Las acciones que realiza el usuario toman mucho tiempo.*	*El sistema se demora en responder a las acciones del usuario.*	*H7 – Flexibilidad y eficiencia de uso.*

Elementos positivos del sistema:

- *Agilización en los procesos de documentación interna de la empresa.*
- *Los procesos de consulta de un documento son más ágil y eficaz.*
- *Ayuda a la disminución del papel y almacenamiento de los documentos.*
- *La rapidez en la información y la posibilidad de responder inmediatamente.*

Fase 2: Concertación del Listado de Problemas.

Una vez todos los evaluadores han realizado la evaluación individual, estos se reúnen para generar una lista única de problemas de usabilidad encontrados en el software SEVENET de la Industria Licorera del Cauca (ILC), agrupándolos de acuerdo al principio de usabilidad que incumplen.

Fase 3 y 4: Calificación de Problemas.

En la fase 3, luego de que el coordinador haya unificado en una lista los problemas de usabilidad encontrados por los evaluadores, esta es entregada a cada evaluador nuevamente, para estimar la severidad y frecuencia con que aparecen los problemas. Se utiliza la tabla 3 para determinar el grado de severidad de cada problema y su frecuencia de aparición.

Table. 3. *Tabla de Notas Asignadas por los Evaluadores*

Calificación de Los Problemas															
Id	**Evaluador 1**			**Evaluador 2**			**Evaluador 3**			**Evaluador 4**			**Evaluador 5**		
Problema	**Severidad**	**Frecuencia**	**Criticidad**	**Severidad**	**Frecuencia**	**Criticidad**	**Severidad**	**Frecuencia**	**Criticidad**	**Severidad**	**Frecuencia**	**Criticidad**	**Severidad**	**Frecuencia**	**Criticidad**
P1	4	5	**9**	3	5	**8**	4	5	**9**	4	3	**7**	3	3	**6**
P2	4	3	**7**	3	3	**6**	4	4	**8**	4	5	**9**	3	5	**8**
P3	3	3	**6**	3	4	**7**	2	3	**5**	4	4	**8**	4	3	**7**
P4	3	1	**4**	4	1	**5**	2	3	**5**	4	3	**7**	3	3	**6**
P5	2	3	**5**	2	3	**5**	4	3	**7**	4	4	**8**	4	1	**5**
P6	4	3	**7**	5	4	**9**	4	3	**7**	2	3	**5**	4	3	**7**
P7	3	2	**5**	2	2	**4**	1	2	**3**	4	1	**5**	2	3	**5**
P8	4	3	**7**	2	3	**5**	4	1	**5**	4	3	**7**	2	3	**5**
P9	5	4	**9**	4	3	**7**	3	5	**8**	4	5	**9**	4	1	**5**
P10	4	3	**7**	4	5	**9**	3	5	**8**	3	5	**8**	4	5	**9**
P11	4	3	**7**	2	3	**5**	4	1	**5**	4	3	**7**	2	3	**5**
P12	3	5	**8**	5	4	**9**	4	3	**7**	3	5	**8**	4	3	**7**
P13	3	4	**7**	2	3	**5**	3	5	**8**	4	5	**9**	3	5	**8**
P14	3	5	**8**	4	5	**9**	4	5	**9**	4	1	**5**	5	4	**9**
P15	4	5	**9**	4	1	**5**	3	5	**8**	3	5	**8**	4	5	**9**

Una vez todos los evaluadores han realizado la calificación de cada problema de acuerdo a su severidad y frecuencia de aparición, se continua con la fase 4, donde el coordinador de la evaluación calcula la criticidad ***(C)*** que equivale a la suma entre la severidad y la frecuencia de cada problema de usabilidad [16].

$$\mathcal{C} = \delta + \mathcal{F} \qquad (1)$$

Posteriormente, el coordinador promedia las calificaciones individuales, calcula la desviación estándar y en compañía de los evaluadores analiza y realiza un ranking de los problemas según:

Severidad. Es el nivel de gravedad del problema identificado. Con esto, es posible identificar aquellos problemas más peligrosos, que comprometen la usabilidad del sistema.

Frecuencia. Es el grado de ocurrencia del problema identificado. Con esto, es posible saber qué tan habitual o repetido es el problema dentro del sistema.

Criticidad. Es la suma de la severidad más la frecuencia. Con esto, es posible determinar qué problemas son más críticos o cruciales, que amerita una solución más rápida, pues involucran aún más la usabilidad del sistema, comparados con otros.

En la tabla 4 se presentan las escalas de evaluación con su nota para su respectiva severidad y frecuencia. Usualmente al evaluador se le asigna una lista de chequeo para llevar el control sobre los principios incumplidos y permita dar calificaciones más precisas [17].

Table. 4. *Tabla de Severidad y Frecuencia*

Escala de Severidad y Frecuencia		
Nota	**Severidad**	**Frecuencia**
0	No es un problema de usabilidad.	< 1%
1	Problema "Cosmético": no necesita ser resuelto a menos que se disponga de tiempo extra en el proyecto.	1-10%
2	Problema de usabilidad menor: arreglarlo tiene baja prioridad.	11-50%
3	Problema de usabilidad mayor: es importante arreglarlo, se le debe dar alta prioridad.	51-90%
4	Problema de usabilidad catastrófico: es imperativo arreglarlo antes de que el producto sea liberado.	> 90%

Fase 5: Análisis de Resultados.

En esta fase se presenta el análisis de resultados obtenidos mediante la evaluación realizada por los evaluadores expertos. A continuación, se presenta el perfil de los evaluadores, rankings según severidad, frecuencia y criticidad.

Perfil de los Evaluadores. Se presenta en la tabla 5 el perfil de los evaluadores expertos encargados de evaluar el sistema gestor de base de datos (SGBD) SEVENET de la Industria Licorera del Cauca (ILC).

Table. 5. *Perfil de Evaluadores*

Perfil de Evaluadores Expertos	
Evaluador	**Perfil**
N° 1	Ingeniero Informático egresado de la Institución Universitaria Colegio Mayor del Cauca en el año 2017. Especialista Tecnológico en Seguridad en Redes de Computadores del Centro de Aprendizaje SENA. Fue uno de los desarrolladores del trabajo de tesis realizado en la Institución Universitaria Colegio Mayor del Cauca [18].
N° 2	Ingeniera Informática egresada de la Institución Universitaria Colegio Mayor del Cauca en el año 2017. Fue uno de los desarrolladores del trabajo de tesis realizado en la Institución Universitaria Colegio Mayor del Cauca [19].
N° 3	Ingeniero de Sistemas egresado de la Universidad del Cauca en el año 2018. Empleado de la compañía mexicana DA Comp SC / Desarrollo y Asesoría en Computación, donde desempeña en los cargos de Líder Técnico y Desarrollador Web.
N° 4	Ingeniera Informática egresada de la Institución Universitaria Colegio Mayor del Cauca en el año 2017. Desarrolladora del proyecto [20].
N° 5	Estudiante de 8 semestres de Ingeniería Informática de la Institución Universitaria Colegio Mayor del Cauca. Como experiencia en usabilidad tiene la aprobación del curso interacción humano computador (HCI) que se dicta en el semestre 7 de la carrera que cursa actualmente.

Rankings de problemas según su criticidad. A continuación, en la tabla 6 se presenta el ranking de los 5 problemas con mayor criticidad ***(C)*** organizados de menor a mayor, donde el problema más crítico es P34, y el menos crítico es P7, se presenta una muestra de los 5 problemas más relevantes.

Table. 6. *Ranking de problemas según criticidad*

Ranking de Problemas de Criticidad				
Problema		Promedio		
Id	Definición	S	F	C
P7	El uso del vocabulario no es consistente en todo el sistema.	2,4	2	**4,4**
P4	Las metáforas no son fáciles de comprender.	3,2	2,2	**5,4**
P8	Los símbolos los utilizados no son comprensibles para el usuario y facilitan la interacción con el sistema.	3,2	2,6	**5,8**
P11	No se le pide confirmación al usuario antes de realizar una acción crítica (cancelar, eliminar, aceptar).	3,2	2,6	**5,8**
P18	La interfaz del sistema está sobrecargada de información, opciones u otros elementos que distraen al usuario.	2,2	2,6	**5,8**
P5	En la aplicación no se explican elementos complejos.	3,2	2,8	**6**
P16	El sistema no informa al usuario si la ejecución de una acción requiere mucho tiempo.	3,8	2,8	**6,6**
P27	El sistema no ofrece una ayuda clara y simple.	3,8	2,8	**6,6**
P29	La ayuda no es fácil de encontrar, entender y aplicar.	3,6	3	**6,6**
P24	Los mensajes de error no orientan al usuario para solucionar el problema.	2,8	3,4	**6,6**

Fase 6: Gráfica Desviación Estándar.

Para la siguiente gráfica (ver Fig. 2) resultante de la prueba, respecto a la desviación estándar se toma para el eje (X) los 35 problemas hallados y para el eje (Y) los valores de la desviación estándar, donde el mayor valor se aproxima a 2.5.

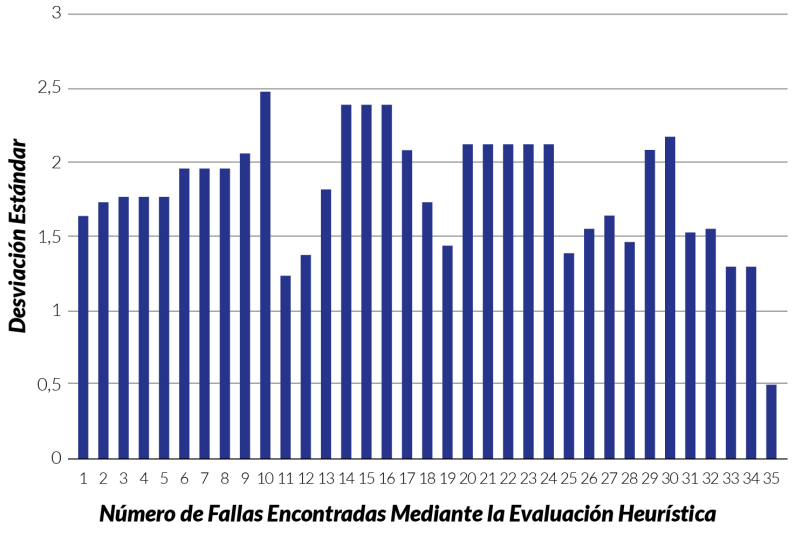


Fig. 2. *Desviación estándar*

El análisis de la anterior figura se observa que el problema con desviación estándar más bajo y con el que los usuarios más coincidieron fue ***(p35)*** de tal modo el problema con desviación estándar más alto fue ***(p10)*** seguido de los problemas ***(p14, p15, p16)*** e ***(p20, 21, p22, p23, p24)*** y terminando con los problemas ***(p29, p30)*** en los cuales menos coincidieron.

Fase 7: Conclusiones de la Prueba.

Los problemas más relevancia fueron:

El problema más severo es P30 (La ayuda no siempre está visible y disponible). Con dificultad los usuarios hallaron la herramienta de ayuda, también notaron que la información disponible en dicha herramienta no aporta soluciones al usuario.

El problema menos severo es P7 (Las interfaces que conforman el sistema no tienen un aspecto visual coherente). Los usuarios mostraron dificultad para interactuar y entender el software.

El problema más frecuente es P34 (En la ayuda la información no es fácil de encontrar). Debido a que no existe consistencia en la información y no existe una guía que oriente a los usuarios en la solución de errores o dudas.

El problema menos frecuente es P7 (Las interfaces que conforman el sistema no tienen un aspecto visual coherente). Por lo que se vuelve necesario mejorar la interfaz principal del sistema para que al usuario le sea más amigable.

El problema más crítico es P34 (En la ayuda la información no es fácil de encontrar). No hay consistencia en la información y es difícil encontrar datos que realmente ayudes u orienten al usuario.

El problema menos critico es P7 (Las interfaces que conforman el sistema no tienen un aspecto visual coherente). El diseño de las interfaces hace compleja la interacción del usuario con el software.

El principio más incumplido por el software fue H10 – Ayuda y Documentación. En este caso la mayoría de los usuarios señalaron que el Sistema Gestor de Base de Datos (SGBD) SEVENET de la Industria Licorera del Cauca(ILC), incumple esta heurística al no brindar claridad y consistencia en la información de ayuda.

El principio menos incumplido fue H1 – Visibilidad del estado del sistema. Este principio heurístico fue el mejor calificado por los evaluadores, lo que permite concluir que el Sistema Gestor de Base de Datos (SGBD) SEVENET de la Industria Licorera del Cauca(ILC), si lo cumple y que la visibilidad del sistema es buena, mostrando así las acciones delos usuarios, sus estados, etc.

V. RECOMENDACIONES OBTENIDAS

Como resultado de la evaluación heurística por experto realizada al sistema gestor de base de datos (SGBD) de la Industria Licorera del Cauca (ILC), en esta sección se presenta un conjunto de recomendaciones (ver tabla 7) teniendo en cuenta los 10 principios expuestos por Heurísticas de Nielsen.

Table. 7. *Heurísticas de Nielsen recomendaciones*

Recomendaciones de Problemas Heurísticos		
Id	**Principio Heurístico**	**Descripción**
H1	Visibilidad del estado del sistema	• El sistema muestra la información recibida o envía. • Cuando se sube un documento al sistema nos indique cuando se está cargando y el tiempo restante. • Mejoramiento acerca de la ubicación del usuario (miga de pan).
H2	Relación entre el sistema y el mundo real	• El software debe mejorar lo correspondiente a un lenguaje apropiados debido a que la información que hallen en el sistema sea similar al que ellos trabajan en físico. Esto con el fin de que el manejo del sistema sea fácil. • Manejar un orden lógico y natural en los módulos del sistema en cada uno de los módulos.
H3	Control y libertad de usuario	• Mejorar aquellos iconos, menús, lenguajes en el cual el usuario permita comprender esas acciones de rehacer y deshacer. • Lo correspondiente a menús, botones e iconos de emergencias deben ser más visibles y entendibles para los usuarios para que de esta manera no cometan errores en el sistema.
H4	Consistencia y Estándares	• Corregir la duplicidad de información. • Los diferentes menús deben manejar vocabulario y entendible. • Mejorar la ubicación de los menús, imágenes e iconografía adecuarlos a los nuevos estándares.
H5	Prevención de Errores	• Mejorar los mensajes de errores, advertencias, acciones (Guardar, Eliminar, • Reenviar, Limpiar) al momento en que los usuarios van a terminar o finalizar una tarea o proceso en el sistema. • Cuando están realizando una búsqueda o creación de un documento al momento de digitar los datos en el formulario mejorar las características que debe tener el campo.
H6	Minimizar la carga de memoria del usuario	• Al momento de consulta y creación de un formularios no sea tan extensos esto con el fin de ahorrar procesos y disminuir la carga de información a modo de resumen.
H7	Flexibilidad y eficiencia de uso	• El sistema no adapto, ya que no es necesario que hayan aceleradores o atajos de teclado con el fin de no confundir a los usuarios.
H8	Diseño estético y minimalista	• Cambiar los menús o interfaces de color rojo por colores pasteles, permitiendo que el usuario no canse la vista al momento de navegar por el sistema. • Mejorar la distribución de cada uno de los menús, módulos y formularios para que el usuario tenga mejor usabilidad. • Modificar la iconografía que sea un poco más visible para el usuario, en ciertos casos cambiar de iconos. • Eliminación de duplicidad de información • Mejorar los diferentes formularios que maneja el sistema .
H9	Ayuda al usuario para reconocer, diagnosticar, recuperarse de errores	• Mejorar los mensajes de error donde el usuario tenga claro acerca del problema causado cuando envié un documento, creación o consulta de un documento.
H10	Ayuda y Documentación	• La opción de ayuda debe mejorar ser amplia y clara la información donde el usuario se informe acerca de las funcionalidades del sistema. • Los iconos que se visualizan en ayuda no son lo suficientemente claro deben ser un poco grandes o claros. • Mejorar la información que se encuentra en cada icono, se recomienda agregar un enlace lo cual lo lleve a un documento PDF donde lo informe sobre la funcionalidad de ese icono.

VI. CONCLUSIONES

Con respecto a la evaluación heurística es un método muy fácil de aplicar a sistemas de información, ya que permite aplicar en cualquier parte del proceso de desarrollo y cualquier lugar físico, muy importante es que se debe contar con evaluadores que tengan conocimientos previos acerca de los principios heurísticos.

Como conclusión de la evaluación heurística realizada por expertos, arrojó como aspectos a mejorar en el software la interfaz, iconografía, vocabulario, aspectos relacionados con el color de la interfaz, adaptar mejor los módulos, y la opción de ayuda ser más clara y explícita para los usuarios.

Se debe corregir la duplicación y sobrecarga de información, así como también abreviar los formularios de consultas que son demasiado extensos y generar una guía de soluciones a posibles errores que sirva de soporte para el usuario.

Con diagnóstico de usabilidad a el sistema gestor de base de datos (SGBD) SEVENET la Industria Licorera del Cauca (ILC), se han conseguido plantear un conjunto de recomendaciones y sugerencias para el manejo, uso y navegabilidad del software y requerimientos propios de los empleados; para que de esta manera se siga implementando el uso de SEVENET internamente para la agilización y optimización en la parte documental.

Las pruebas de usabilidad realizadas a los empleados de la de la Industria Licorera del Cauca (ILC) demostraron que el sistema gestor de base de datos (SGBD) SEVENET debe mejorar en aspectos como: la interfaz, vocabulario, iconografía, estética que son fundamentales para la utilización del software.

Como iniciativa se proyecta que la Industria Licorera del Cauca (ILC) para el año 2019, implemente el uso del sistema gestor de base de datos (SGBD) SEVENET, basados en el diagnóstico y pautas que este proyecto permito realizar, visualizando una mejora en sus procesos internos y en la solución ante fallas.

REFERENCIAS

[1] *Ferré, X.: "Usabilidad: Software pensado para los usuarios". Revista Universia [Online], (Fecha de consulta: 27/10/2018).*

[2] *Sánchez, Walter.: "La usabilidad en Ingeniería de Software: definición y características". Ing-novación. Revista de Ingeniería e Innovación de la Facultad de Ingeniería, Universidad Don Bosco, Agosto (2011), Año 1, No. 2. pp. 7-21. ISSN 2221-1136.*

[3] *Mascheroni, M. A., Greiner, C. L., Petris, R. H., Dapozo, G. N., & Estayno, M. G. (2012). Calidad de software e ingeniería de usabilidad.*

[4] *Libro Evaluación Heurística. ISO Standards No. 9241-11: Guidance on usability. Geneva, Switzerland, 1998.*

[5] *Usability Net: Internacional standards for HCI and usability. Disponible en http://www.usabilitynet.org/tools/r_international.htm(Consultada en Marzo 2019).*

[6] *Males H, Chuvila P, Muñoz C.: Metodología "Guía metodológica para la ejecución de evaluaciones heurísticas en aplicaciones de propósito general, teniendo en cuenta los principios heurísticos de Nielsen". Metodología, Popayán, 2017).*

[7] *Milán, Verónica.: Usabilidad web o experiencia de usuario. Principios básicos y beneficios, 9 marzo, 2018.*

[8] *Merino, Julián Pérez Porto y María.: Definición de usabilidad, https://definicion.de/usabilidad/, 2010.*

[9] *Merino, Julián Pérez Porto y María.: Definición de usabilidad, https://definicion.de/usabilidad/, Publicado: 2010. Actualizado: 2013.*

[10] *Presidencia de la Republica, Directiva Presidencial - MinTic Norma de cero papel (Ley 594 de 2000)*

[11] *Canales, Diego Arancibia Céspedes y Sebastián González.: Heurísticas de Usabilidad y Experiencia del Usuario en Redes Sociales, http://opac.pucv.cl/pucv_txt/txt-4500/UCC4799_01.pdf*

[12] *Gonzales M, Pascual A, Lorés J.: Libro de Evaluación Heurística, Concepto de Heurísticas de Nielsen, Universidad de Lleida.*

[13] *Gonzales M, Pascual A, Lorés J.: Libro de Evaluación Heurística.: Concepto de Evaluación Heurística, Universidad de Lleida.*

[14] *Molich, R., Nielsen, J.: Improving a human-computer dialogue. En Communications of the ACM, num33, vol. 3. ACM Press, pág 338-348, 1990.*

[15] *Reyes, Vera., Liberos A.: Método para la evaluación integral de la usabilidad en sistemas E-Learning Revista Educación en Ingeniería*

[16] *Solano, A., Cardona, Jesús.: Evaluación colaborativa de la usabilidad en el desarrollo de sistemas software interactivos,2015.*

[17] *Males H, Chuvila P, Muñoz C.: Guía Metodológica Para La Ejecución De Evaluaciones Heurísticas En Aplicaciones De Propósito General, Teniendo En Cuenta Los Principios Heurísticos De Nielsen", Escala de Severidad y Frecueuncia, 2017.*

[18] *Males H, Chuvila P, Muñoz C.: Guía Metodológica Para La Ejecución De Evaluaciones Heurísticas En Aplicaciones De Propósito General, Teniendo En Cuenta Los Principios Heurísticos De Nielsen", Perfil de Evaluador N°1, 2017.*

[19] *Males H, Chuvila P, Muñoz C.: Guía Metodológica Para La Ejecución De Evaluaciones Heurísticas En Aplicaciones De Propósito General, Teniendo En Cuenta Los Principios Heurísticos De Nielsen", Perfil de Evaluador N°2, 2017.*

[20] *Diego Felipe Paz.: Propuesta De Una Guía Para La Implementación De Una Mesa De Ayuda (Help Desk) En La Empresa Certicol Sas Basada En El Framework De Itil Lite, 2017.*

Biografía Autor(es)

Correspondencia. Gabriel Elías Chanchí Golondrino. Doctor en Telemática. Docente, Universidad del Cauca. Email: gabrielchanchi@gmail.com .Cauca, Colombia.

How to adapt deep learning models to a new domain: Biomedical relation extraction case

Jefferson Amado Peña Torres (a)
Víctor A. Bucheli (b)
Raúl Ernesto Gutiérrez De Piñerez (c)
Fabio A. González O. (d)

Cómo adaptar los modelos de aprendizaje profundo a un nuevo dominio: Caso de extracción de relaciones biomédicas

Resumen

En este artículo, adaptamos un modelo de la literatura para la extracción de relaciones (RE) que utiliza el aprendizaje profundo. Entrenamos el modelo para extraer relaciones semánticas para el dominio biomédico, es deseable adaptar los modelos de extracción de la evolución a un nuevo dominio que es diferente a sourcev-domain. Presentamos un análisis más detallado de la problema y evaluación preliminar. Existen variantes de arquitecturas de aprendizaje profundo y son ampliamente utilizados en tareas de procesamiento del lenguaje natural (PNL), en particular para distinguir relaciones semánticas entre entidades nombradas. Modelos de Train-ventirely DL en una escala de tiempo corta no es práctico, deseamos que los modelos se adapten rápidamente de diferentes conjunto de datos con varios dominios y sin vdelay. La adaptación es crucial para la inteligencia sistemas que operan en el mundo real, donde los factores cambiantes y no anticipados Las perturbaciones son inusuales.

Palabras clave: Aprendizaje profundo, procesamiento del lenguaje natural, extracción semántica, extracción de relaciones.

Abstract

In this paper, we adapt a model from the literature for relation extraction (RE) that utilizes deep learning. We trained the model forvextracting semantic relations for biomedical domain, is desirable to adaptvrelation extraction models to a new domain that is different to sourcevdomain. We present more detailed analysis of the problem and preliminary evaluation. There are variants of deep learning architectures andvare widely used in natural language processing (NLP) tasks, in particular for distinguishing semantic relations between named entities. Trainventirely DL models on short time-scale is impractical, we desire modelsvto adapt quickly from different dataset with several domain and withoutvdelay. The adaptation is crucial for intelligent systems operating in thevreal world, where changing factors and unanticipated perturbations arevusual.

Keywords: Deep learning, natural language processing, semantic extraction, relation extraction.

I. INTRODUCTION

The goal of relation classification is to predict a semantic relation r between a entity e1 and a entity e2 from a given sentence s. RE plays a key role in knowledge from unstructured text and it has a wide range of applications in many domains both as population and as information extraction from knowledge bases [16, 23, 27]. With the rapid growth of valuable knowledge that is recorded in the form of unstructured text data, there is considerable interest in automatic detection and extraction of semantic relations among the entities [12].

Although many work have been done on developing supervised relation extraction models [10, 3, 18, 2], neural network-based approaches have been proposed for relation extraction, in particular deep learning works such as recursive neural network-based [26, 32, 15], recurrent neural network-based (RNN) [30, 35, 34], and convolutional neural network-based approaches (CNN) [31, 19, 24, 29, 16].

The deep learning (DL) has clearly demonstrated its power in promoting the RE task. Specifically, focused on relations from English, but since deep learning models have the advantage of being able to do training with little or no domain knowledge. Several works has been performed for relation extraction in texts using deep learning methods, but two open challenges according to language and domain of deep learning models for relation extraction are (a) the lack of training samples in some languages (b) the generalization of model in domain with different classes of relations.

Each of those open challenges could potentially generate several research question. Three research questions are meant to help us assess the credibility of our proposed

- How to train a existing DL model which are suiTable. for a domains different to domain knowledge for which was trained?

In terms of training samples in some languages. Combining the shortest dependency path for Chinese relation classification on the CNN model welltrained for English the works on English should be easily extended to other languages.[33]. This keeps open the challenge of transferring well-trained deep models that exist in other domains. Intrigued by this idea, we first propose a baseline DL model for RE from literature, and then prove its performance by joint use of linguistic features on biomedical domain. For addressing this challenges, in this work we describe the deep learning method and compare the performance when add linguistic features to baseline model on biomedical domain.

The remainder of this paper is structured as follows. First, We present a DL model based on [31, 19] for relation extraction. Considering the great potential of deep learning in RE and bioinformatic researchs (biology, bio-medicine and healthcare), We evaluate the baseline model on multiple public PPI, DDI and CPI corpus, followed by our experiments with linguistic features. The results are discussed and summarized in the last two sections.

II. RELATED WORK

Relation extraction is generally approached as a classification task that occurs among two or more named entities within a same sentence with a semantic relationship. According to the number of semantic relation classes RE task can be binary or

multi-class extraction. In this paper, we focus on binary relation extraction task. For RE, the supervised methods has high-performing, need annotated data, which is time consuming and entails a human labor intensive [1, 11, 25, 36]. Recently, models based on deep neural networks such as CNNs and RNNs, have shown promising results for RE. [31] explored the RE without a exhaustive pre-processing. They employed a CNN and observed that any automatically learned features yielded excellent results and can potentially replace the manually designed features. Similar DL approaches have been proposed with the RNN architecture [26, 7] and with CNN and RNN combinations. [32]. We also surveyed the related works in recent years, as shown in Table. 1

Most of the work that we survey is concerned with English, and extending these techniques to non-English languages or another domains is not always easy. Because a variety of possible relations are characterized by its own syntactic and lexical properties, which vary from one domain/language to another. The notion of a relation is inherently ambiguous and there is often an inherent ambiguity about what a relation \means". [20].

Table. 1. *The task of relation extraction with DL approach*

Model F1	***F1***
Research with CNN-based Models	
CNN (Zeng et al., 2014)	*82.7*
CR-CNN (dos Santos et al., 2015)	*84.1*
CNN (Nguyen, Thien Huu, and Ralph Grishman, 2015)	*82.8*
CNN (XU, 2015)	*85.4*
depLCNN + NS (Xu et al., 2015a)	*85.6*
Multi-Attention CNN (Wang et al. 2016)	*88.0*
Attention CNN (Huang and Y Shen, 2016)	*84.3*
Selection-Attention CNN (Yankai Lin, 2016)	*88.0*
Research with RNN-based Models	
matrix-vector RNN (Socher et al., 2012)	*82.4*
Entity Attention Bi-LSTM (Lee et al., 2019)	*85.2*
Hierarchical Attention Bi-LSTM (Xiao and C Liu, 2016)	*84.3*
Attention Bi-LSTM (Zhou et al., 2016)	*84.0*
Bi-LSTM (Zhang et al., 2015)	*82.7*
BRCNN (Cai et al., 2016)	*86.3*
DRNNs (Xu et al., 2016)	*86.1*
SDP-LSTM (Xu et al., 2015b)	*83.7*
DepNN (Liu et al., 2015)	*83.6*
FCN (Yu et al., 2014)	*83.0*
RCNN(Lai, Siwei, et al. 2015)	*96.49*
PCNN(Zeng, D., Liu, K., et al. (2015))	*84.0*

Many efforts have been invested in biomedical relation extraction, the goal of biomedical information extraction is discover valuable knowledge about proteins, drugs, diseases, genes, adverse effects and other interactions biological from unstructured free text. Most of the work that we survey is concerned with English, and extending these techniques to non-English languages or another domains is not always easy. Moreover, with DL, several hand-annotated datasets have been created for biomedical relation extraction such as protein-protein interactions (PPI) [21, 4], drug-drug interactions (DDI) [9], drug-related adverse effects (ADE) [5] and and chemical-protein interactions (CPIs).

Intrigued by the idea extend a DL model, we first propose a baseline DL model for RE from literature, and then prove multi-representation on biomedical domain.

III. METHODOLOGY

We evaluated the performance of a preexisting DL model for relation extraction when is adapt for RE task in another domain, The main steps is the adaptation of a DL model can be summarized as follows. First, as baseline model, we trained a Convolutional Neural Network (CNN) [14], known as a DL architecture for RE that allows the model to automatically find relevant information (features or patterns) from a source sentence while predicting a target relation. The CNN has been applied successfully to many NLP and has achieved significant performance for RE task. Second, a set of features called multi-representation are defined and are combined as input to train a DL model. This is the most crucial part of our work in order to extract explicit semantic relation, we compare results of two models in order to understand adaptability of well-trained model on another domain.

3.1 Baseline model: Convolution, Pooling and Non-linear Layers

We implemented a system based on [31, 19]. The structure of the system has three layers: a convolutional layer, a max pooling layer, and a fully connected layer with dropout and softmax layer. Figure 1 gives an overview of the network.

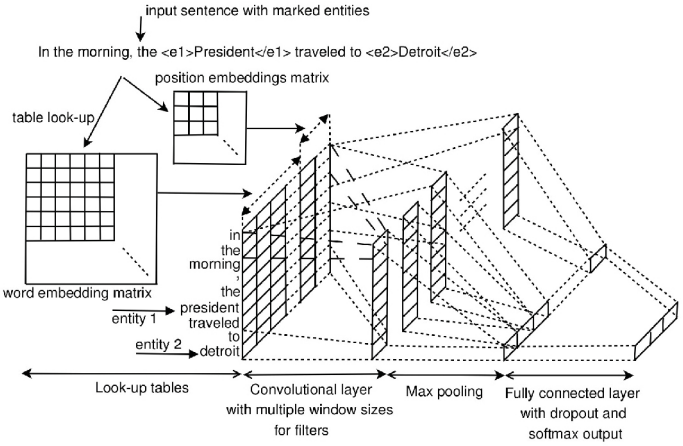


Figure 1: Convolutional Neural Network for Relation Extraction.

Fig. 1. *Convolutional Neural Network for Relation Extraction from [19].*

3.2 Multi Representation

The inputs of the CNN are raw words of the sentence s. Each input word is transformed into a vector via word embedding. The word embedding is one of the most popular representations, characterized by being able to capture the semantic and syntactic properties of words according the context.

Word Embeddings: With the capture syntactic and semantic meanings of the words into distributed representations. A sentence s with m words $s = w1, w2, \ldots, wm$, every word w_i is represented by a real-valued vector. These word representations are encoded in an embedding matrix as $\boldsymbol{V} \in^{da/v}$ where V is a fixed-sized vocabulary. Unfortunately, makes these word representations word usually takes a long time to train and normally are used trained word embeddings that are freely available [28].

Position Embeddings: In RE task, the words close to the target entities are usually informative to determine the relation between entities. Similar to [31] use position embeddings specified by entity pairs. Apparently, it is not possible to capture such structure information only through semantic and syntactic word features. It is necessary to specify which input tokens are the target nouns in the sentence and where is positioned. The position among entities is a relative distances and also are mapped to vector representations.

IV. RESULTS AND DISCUSSION

4.1 Experiments

Our experiments are intended to demonstrate that a DL models (baseline) for relation extraction (RE) can be adapted to another domain using multirepresentation. To this end, we first introduce the datasets and evaluation metrics used in the experiments. Next, we show the parameters of baseline model and our experiments. And then we evaluate multi-representation effects and performance on the data. Finally, we compare the performance of baseline model with our adapt model.

Dataset We explore RE tasks that focus on biomedical domain, specially relations such as protein-protein interactions (PPIs), drug-drug interactions (DDIs) and chemical-protein interactions (CPIs). From several scarce resources may take advantage for adapt a pre-trained model. All are publicly available and detailed in the following.

The adapt model was tested with three Training Corpus Biomedical dataset. For protein-protein interactions(relations), We used the BioInfer dataset [22]. For adverse drug events, we used ADE corpus [5] was used and we used ChemProt corpus [13] for chemical-protein interactions. The detailed information of the each datasets are listed in Table.s 2, 3 and 4, respectively.

Table. 2. *Bioinfer statistics [21]*

Dataset	Sentences	Positive	Negative	Total
BioInfer	1100	2,534	7,132	9,666

- **BioInfer:** is a public resource providing an annotated corpus of biomedical English that aimed at developing information extraction (IE) systems and their components in the biomedical domain.

Table. 3. *ADE corpus from [5]*

Dataset	Sentences	Positive	Negative	Total
ADE	*4,272*	*988*	*2196*	*3184*

- **ADE (Adverse Drug Effect corpus):** [5] consists of MEDLINE case reports, annotated with drugs and conditions (e.g., diseases, signs and symptoms), along with untyped relationships between them.

Table. 4. *ChemProt corpus [13]*

Relations	Training Set	Development Set	Test set
Active	*768*	*550*	*664*
Inhibitor	*2251*	*1092*	*1661*
Agonist	*170*	*116*	*194*
Antagonist	*234*	*197*	*281*
Substrate	*705*	*457*	*643*
Negative	*12461*	*8070*	*11013*
Total	*16589*	*10482*	*14456*

- **ChemProt:** Consisted PubMed abstracts annotated with chemical and protein entities. The relations were annotated with 10 chemical-protein relations. According to the share task description, only 5 out of 10 semantic relation types would be evaluated.

Other dataset are important for our work. The baseline model [19], is evaluate on two datasets SemEval-2010 Task 8 dataset [8] and ACE 2005.

- **SemEval 2010 task 8:** Is focuses on multi-way classification between pairs of nominals. The task was designed to compare different approaches to semantic relation classification

- **ACE-2005 This dataset consists of 6 domains:** broadcast news (bn), newswire (nw), broadcast conversation (bc), telephone conversation (cts), weblogs (wl) and usenet (un).

- **reACE(Edinburgh Regularized Automatic Content Extraction)** [6] consists of English broadcast news and newswire annotated with organization, person, fvw (facility, vehicle or weapon) and gpl (geographical, political or location) entities along with relationships between them. Relationships are classified in five types: general-affiliation, organisation-affiliation, partwhole, personal-social and agent-artifact.

Model configuration For all the experiments below, baseline model was trained using the Adam optimizer with a learning rate set to 0.001, weight decay set to 1e-5 and clip grad set to 5. We train for a minimum of 20 and maximum of 25 epochs. Likewise, we used a batch size of 50, a word embedding with dimension 50, a position embedding with dimension 20. We used filter widths of size 2, 3, 4 and 5 and to avoid overfitting, we use dropout with value is set to 0.5.

Evaluation measure For the task for relation classification, we use the F1- score as our evaluation measure. F1-score is defined as the harmonic mean between precision (P) and recall (R) such that:

$$P = \frac{TP}{TP + FP}, R = \frac{TN}{TN + FN}, and\ F1 = \frac{2P.R}{P + R}$$

where TP, FP and FN are the true positives, false positives and false negatives, respectively.

4.2 Results

The performance the SemEval2010, ADE, BioInfer, Chemprot and reACE datasets are presented in Table. 5. In this Table., we report the mention-level F1-score.

Table. 5. *F1-score for each dataset*

Dataset	*F1-Score*
SemEval2010	*82.76*
ADE	*99.56*
BioInfer	*38.50*
reACE	*25.24*
Chemprot	*59.80*

Overall, we make two observations. First, we observe that baseline model is not sTable. for each dataset. Although, the baseline model predictions not is reason for know of what happen, the performance is meant to provide a neutral benchmark against which to measure the effects of adaptation changes. We observe that baseline training not improves the performance. The ADE and reACE contain imbalanced class distribution, this is perhaps reason of 99.56 and 25.24 performance respectively. Our hypothesis is that as the relation extraction with an unbalanced corpus in our baseline model sensitive and generate a large performance gap. The baseline model with ADE corpus outperforms the baseline model, this is perhaps not surprising since short fragments of texts with scarce grammatical information, where convolutions with small windows are enough to capture necessary information for RE classification. Second, when comparing the CNN performance, we hypothesize that RE task is more complex than the CNN model. When there is domain shift with supervised domain adaptation, the model need transfer of knowledge from the target domain. While we suspect there is still

room for improvement, without utilizing domain specific information, the datasets may differ in ways we cannot account for with our method. There are two main aspects in our baseline model: First, we observe that unbalanced corpus have a large impact on the F1-score. More importantly, we see that using a corpus with proportion of positive to negative relations balanced to makes the better performance. Next, we compare our domain adaptation method with a CNN base model (baseline) and the performance is quite low. We believe our performance can be attributed to 1) vector representation and 2) class imbalance from dataset. For example, word embeddings and positions embeddings combination outperforms to baseline model with word embeddings.

4.3 Discussion

We demonstrated a simple way to train a model for domain target in cases where there is data with another distributions and classes, but with the same task. We also showed that method not performs better than baseline model. Therefore, is need adding much complexity for adapt a well-trained model. We evaluate the DL model on different datasets. We tried a deep learning model for extracting semantic relations between entities and follow similar experimental setting as in Nguyen and Grishman [19].

The DL model no is capable of extract relations from new domain. However, to adapt deep learning models to a new domain the results confirm the need of a balanced dataset and information of target domain. Our approach do not improve the baseline score. The main reason adaptation does not work, is the proportion of positive to negative relations is very different in ADE, BioInfer, reACE, Chemprot and SemEval datasets. The problem of class imbalance between datasets is a known problem for unsupervised domain adaptation methods [17].

We can see that the model is sensitive to embedding and it plays a significant role in model training. There are several embedding representation such as position embedding that represent the position of a entity in the sentence, the medical and biological embedding, that contains specific information domain word, also includes methods that can generate domain-sensitive word embeddings. For future work, we are considering the same approach exploring other representations, we speculate that we can achieve input layer improve with major information from text.

V. CONCLUSIONS

In this paper, we proposed an approach for DL model adaptation in a new domain and which aims to learn a pre-trained model in biomedical domain for RE task. More specifically, we use an architecture to transfer the RE task from one domain to another domain. After pre-processing the dataset, experimental results on several benchmark datasets demonstrate our proposed method does not achieves the same performance that baseline model.

The approach classifies reACE, BioInfer, Chemprot and ADE with the Fmeasure 25.24, 38.50, 59.80 and 99.56 respectively. Due to the low results, we provide a error analysis and discuss about reason of our results. Our work allows an exploration with another representations and the results to prevent duplication of research efforts in the development of future systems.

ACKNOWLEDGMENTS

The authors would like to acknowledge the systems and Computing Engineering school, Faculty of Engineer, The Universidad del Valle of Cali, Colombia.

REFERENCES

[1] *Antti Airola, Sampo Pyysalo, Jari Bj¨orne, Tapio Pahikkala, Filip Ginter, and Tapio Salakoski. All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning. BMC bioinformatics, 9(11):S2, 2008.*

[2] *Michele Banko, Michael J Cafarella, Stephen Soderland, Matthew Broadhead, and Oren Etzioni. Open information extraction from the web. In IJCAI, volume 7, pages 2670-2676, 2007.*

[3] *Razvan C Bunescu and Raymond J Mooney. A shortest path dependency kernel for relation extraction. In Proceedings of the conference on human language technology and empirical methods in natural language processing, pages 724-731. Association for Computational Linguistics, 2005.*

[4] *Rezarta Islamaj Dogan, Andrew Chatr-aryamontri, Sun Kim, Chih-Hsuan Wei, Yifan Peng, Donald Comeau, and Zhiyong Lu. Biocreative vi precision medicine track: creating a training corpus for mining protein-protein interactions affected by mutations. BioNLP 2017, pages 171-175, 2017.*

[5] *Harsha Gurulingappa, Abdul Mateen Rajput, Angus Roberts, Juliane Fluck, Martin Hofmann-Apitius, and Luca Toldo. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. Journal of biomedical informatics, 45(5):885-892, 2012.*

[6] *Ben Hachey, Claire Grover, and Richard Tobin. Datasets for generic relation extraction. Natural Language Engineering, 18(1):21-59, 2012.*

[7] *Kazuma Hashimoto, Makoto Miwa, Yoshimasa Tsuruoka, and Takashi Chikayama. Simple customization of recursive neural networks for semantic relation classification. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1372-1376, 2013.*

[8] *Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions, pages 94-99. Association for Computational Linguistics, 2009.*

[9] *María Herrero-Zazo, Isabel Segura-Bedmar, Paloma Martínez, and Thierry Declerck. The ddi corpus: An annotated corpus with pharmacological substances and drug-drug interactions. Journal of biomedical informatics, 46(5):914-920, 2013.*

[10] *Nanda Kambhatla. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In Proceedings of the ACL 2004 on Interactive poster and demonstration sessions, page 22. Association for Computational Linguistics, 2004.*

[11] *Seonho Kim, Juntae Yoon, Jihoon Yang, and Seog Park. Walk-weighted subsequence kernels for protein-protein interaction extraction. BMC bioinformatics, 11(1):107, 2010.*

[12] *Natalia Konstantinova. Review of relation extraction methods: What is new out there?*

In International Conference on Analysis of Images, Social Networks and Texts, pages 15-28. Springer, 2014.

[13] Jens Kringelum, Sonny Kim Kjaerulff, Søren Brunak, Ole Lund, Tudor I Oprea, and Olivier Taboureau. Chemprot-3.0: a global chemical biology diseases mapping. Database, 2016, 2016.

[14] Yann LeCun, L´eon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278- 2324, 1998.

[15] Sangrak Lim and Jaewoo Kang. Chemical{gene relation extraction using recursive neural network. Database, 2018, 2018.

[16] Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. Neural relation extraction with selective attention over instances. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), volume 1, pages 2124-2133, 2016.

[17] Tzu Ming Harry Hsu, Wei Yu Chen, Cheng-An Hou, Yao-Hung Hubert Tsai, YiRen Yeh, and Yu-Chiang Frank Wang. Unsupervised domain adaptation with imbalanced cross-domain data. In Proceedings of the IEEE International Conference on Computer Vision, pages 4121-4129, 2015.

[18] Raymond J Mooney and Razvan C Bunescu. Subsequence kernels for relation extraction. In Advances in neural information processing systems, pages 171-178, 2006.

[19] Thien Huu Nguyen and Ralph Grishman. Relation extraction: Perspective from convolutional neural networks. In Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, pages 39-48, 2015.

[20] Sachin Pawar, Girish K Palshikar, and Pushpak Bhattacharyya. Relation extraction: A survey. arXiv preprint arXiv:1712.05191, 2017.

[21] Sampo Pyysalo, Antti Airola, Juho Heimonen, Jari Bj¨orne, Filip Ginter, and Tapio Salakoski. Comparative analysis of five protein-protein interaction corpora. In BMC bioinformatics, volume 9, page S6. BioMed Central, 2008.

[22] Sampo Pyysalo, Filip Ginter, Juho Heimonen, Jari Björne, Jorma Boberg, Jouni Järvinen, and Tapio Salakoski. Bioinfer: a corpus for information extraction in the biomedical domain. BMC bioinformatics, 8(1):50, 2007.

[23] Xiang Ren, Zeqiu Wu, Wenqi He, Meng Qu, Clare R Voss, Heng Ji, Tarek F Abdelzaher, and Jiawei Han. Cotype: Joint extraction of typed entities and relations with knowledge bases. In Proceedings of the 26th International Conference on World Wide Web, pages 1015-1024. International World Wide Web Conferences Steering Committee, 2017.

[24] Cicero Nogueira dos Santos, Bing Xiang, and Bowen Zhou. Classifying relations by ranking with convolutional neural networks. arXiv preprint arXiv:1504.06580, 2015.

[25] Isabel Segura-Bedmar, Paloma Martinez, and Cesar de Pablo-Sánchez. Using a shallow linguistic kernel for drug-drug interaction extraction. Journal of biomedical informatics, 44(5):789-804, 2011.

[26] Richard Socher, Brody Huval, Christopher D Manning, and Andrew Y Ng. Semantic

compositionality through recursive matrix-vector spaces. In Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, pages 1201-1211. Association for Computational Linguistics, 2012.

[27] Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. Representing text for joint embedding of text and knowledge bases. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1499-1509, 2015.

[28] Joseph Turian, Lev Ratinov, and Yoshua Bengio. Word representations: a simple and general method for semi-supervised learning. In Proceedings of the 48th annual meeting of the association for computational linguistics, pages 384-394. Association for Computational Linguistics, 2010.

[29] Kun Xu, Yansong Feng, Songfang Huang, and Dongyan Zhao. Semantic relation classification via convolutional neural networks with simple negative sampling. arXiv preprint arXiv:1506.07650, 2015.

[30] Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. Classifying relations via long short term memory networks along shortest dependency paths. In proceedings of the 2015 conference on empirical methods in natural language processing, pages 1785-1794, 2015.

[31] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, Jun Zhao, et al. Relation classification via convolutional deep neural network. In COLING, pages 2335-2344, 2014.

[32] Dongxu Zhang and Dong Wang. Relation classification: Cnn or rnn? In Natural Language Understanding and Intelligent Applications, pages 665-675. Springer, 2016.

[33] Linrui Zhang and Dan Moldovan. Chinese relation classification via convolutional neural networks. In The Thirty-First International Flairs Conference, 2018.

[34] Runyan Zhang, Fanrong Meng, Yong Zhou, and Bing Liu. Relation classification via recurrent neural network with attention and tensor layers. Big Data Mining and Analytics, 1(3):234-244, 2018.

[35] Shu Zhang, Dequan Zheng, Xinchen Hu, and Ming Yang. Bidirectional long shortterm memory networks for relation classification. In Proceedings of the 29th Pacific Asia conference on language, information and computation, pages 73-78, 2015.

[36] Yijia Zhang, Hongfei Lin, Zhihao Yang, Jian Wang, and Yanpeng Li. A single kernel-based approach to extract drug-drug interactions from biomedical literature. PLoS One, 7(11):e48901, 2012.

Biografía Autor(es)

a. Magíster en Ingeniería. Docente, Universidad del Valle. E-mail: jefferson.amado.pena@correounivalle.edu.co. Cali, Colombia.

b. Doctor en Ingeniería. Docente Universidad del Valle. E-mail: victor.bucheli@correounivalle.edu.co. Cali, Colombia.

c. Doctor en Ingeniería. Docente Universidad del Valle. E-mail: raul.gutierrez@correounivalle.edu.co. Cali, Colombia.

d. Doctor enCiencias de la Computación. Docente Universidad del Valle. E-mail: fagonzalezo@unal.edu.co. Cali, Colombia.

Inception Deck un camino hacia el desarrollo de *Sof Skills* requeridas por el sector TI en Colombia

Manuel Pastrana (a)
Hugo Ordoñez (b)
Sandra Guañarita (c)

Resumen

Los Soft Skills también conocidos como “habilidades blandas” se refieren a los rasgos de personalidad y las actitudes que dirigen el comportamiento de una persona. En la actualidad se han convertido en uno de los requisitos exigidos por las organizaciones al momento de seleccionar un candidato a una vacante; no obstante, las empresas y en especial el área encargada de la selección de nuevos colaboradores manifiestan la falta o poco desarrollo de los Soft Skills en muchos de los candidatos evaluados e incluso en muchos de los colaboradores incorporados, causando en ese último caso aumento en los tiempos de incorporación a la organización e impactando sobre los costos de cualificación y en el corto plazo costos de productividad de la organización. En este sentido, este trabajo propone la técnica Inception Deck como un camino para el desarrollo desde la academia de estas habilidades requeridas por el sector TI, a través del currículo de los programas de Ingeniería de Sistemas y afines en el nivel de pregrado en Colombia.

Palabras clave: Habilidades blandas, inception deck, soft skills.

Inception Deck a path towards the development of Sof Skills required by the IT sector in Colombia

Abstract

Soft Skills also known as "soft skills" refer to the personality traits and attitudes that direct a person's behavior. Today they have become one of the requirements demanded by organizations when selecting a candidate for a vacancy; however, companies and especially the area in charge of the selection of new collaborators state the lack or little development of Soft Skills in many of the candidates evaluated and even in many of the incorporated collaborators, causing in the latter case increase in the time of incorporation to the organization and impacting on the qualification costs and in the short term productivity costs of the organization. In this sense, this work proposes the Inception Deck technique as a way to development from the academy of these skills required by the IT sector, through of the curriculum of Systems Engineering and related programs at the undergraduate level In Colombia.

Keywords: Soft skills, starting deck, soft skills.

I. CONTEXTUALIZACIÓN Y ESTADO DEL ARTE

1.1. Planteamiento de la pregunta de investigación.

En las guías para el diseño de *currículos* de la ACM como IT2017 [1], y SE2014 [2] se consideran además de las habilidades técnicas, las denominadas habilidades blandas o Soft Skills, refiriéndose a los rasgos de personalidad y las actitudes que dirigen el comportamiento de una persona. En [3] incluyen el sentido común, la capacidad de tratar con las personas y una actitud flexible positiva o también rasgos de carácter y habilidades interpersonales que caracterizan las relaciones de una persona con otras personas. Estas habilidades blandas (Soft Kills) complementan las habilidades técnicas (Hard Skills), por lo que el miembro ideal del personal de TI debe poseer ambas. Los proyectos de software pueden lograr mejores resultados asignando personas con habilidades blandas particulares a las diferentes fases del proyecto [3].

Los Soft Skills en la actualidad constituyen un tema de interés en la industria del software dado que se han convertido en un factor importante en el momento de la selección de personal por parte de las organizaciones [1] [4] [5] y el desempeño en los equipos de trabajo: cómo realizan sus tareas y cómo interactúan con sus pares [3]. En los trabajos [6] [7] [8] se han analizado las necesidades de la industria en relación con las habilidades del talento humano para su desempeño efectivo dentro de la organización, y se han detectado las habilidades blandas como una de las demandas por parte de las empresas.

Desde el área de la Ingeniería de Requisitos, fase análisis de requisitos, encontramos la técnica Inception Deck difundida en [9] y complementada en [10] en donde se propone como respuesta a la necesidad de mejorar el inicio de los proyectos de software enfocándose en crear una visión unificada del proyecto logrando el consenso de las opiniones de todos los involucrados (cliente, analistas, stakeholders, desarrolladores, etc), a través de 10 dinámicas. No obstante, Inception Deck más que un conocimiento técnico, involucra un trabajo en equipo bajo una filosofía colaborativa en donde cada integrante pone en marcha una serie de habilidades blandas como comunicación, empatía, manejo de conflictos, y una visión orientada hacia las necesidades del cliente.

Durante la revisión de los estudios e informes sobre Soft Skills en la industria se detectó a la academia como el actor responsable de su formación debido a su rol como agente de cambio social y su compromiso frente a los egresados de sus programas de prepararlos para su incorporación exitosa al mundo laboral y en la sociedad.

De lo anterior, se propone la pregunta de investigación: *¿Cómo integrar la técnica Inception Deck en el currículo de los programas de Ingeniería de Sistemas y afines en el nivel de pregrado a fin de lograr egresados con mejores niveles de desarrollo de Soft Skills requeridos por el sector empresarial?*

II. OBJETIVOS

2.1 Objetivo principal.

Integrar la técnica ágil Inception Deck al currículo de los programas de Ingeniería de Sistemas y afines en Colombia con el fin de lograr mayor nivel de desarrollo de los Soft Skills de sus egresados.

2.2 Objetivos específicos

Exploración y selección. Identificar los Soft Skills relevantes requeridos por el sector empresarial para los profesionales TI y cubiertos por la técnica Inception Deck.

Construcción. Definir la estrategia de integración al currículo de los programas de Ingeniería de Sistemas y afines de la técnica Inception Deck para el desarrollo de los Soft Skills identificados.

Evaluación. Validar el impacto de la propuesta desde la visión de la mejora de proceso mediante el uso de escenarios de prueba.

III. DESCRIPCIÓN DE LA PROPUESTA

La construcción de la propuesta se inicia con una revisión sistemática de informes y estudios en el ámbito nacional o internacional sobre los Soft Skills requeridos en el sector empresarial, a partir de los cuales se selecciona el top 5 de estas habilidades usando como criterio principal la frecuencia de aparición en los estudios, además de la importancia dada en estudios claves para la industria TI como el informe CHAOS y los referentes de formación TI y de gestión de proyectos. Los resultados de este estudio muestran las habilidades de trabajo en equipo y comunicación asertiva como las de mayor relevancia [1] [6], seguido por el liderazgo [6] [8], las capacidades de negociación y la inteligencia emocional y/o madurez emocional. Seguidamente se analiza la técnica Inception Deck a la luz de los Soft Skills identificados, y se obtuvo como resultado su alta compatibilidad, dada su filosofía y mecánica de trabajo la cual requiere de los participantes habilidades de trabajo en equipo, comunicación asertiva, manejo de conflictos, inteligencia emocional e interés por la necesidad del cliente.

Posteriormente se define la estrategia de integración empresa-academia. Esta estrategia se basa en el uso de la técnica Inception Deck como método de trabajo para las actividades grupales y el trabajo en equipo y comunicación asertiva como los Soft Skills a monitorear y seleccionar las dinámicas del Inception ¿por qué estamos aquí?, ¿quién está en el auditorio?, Discurso en el ascensor, Caja del producto, Lista del NO, ¿qué nos impide dormir en las noches? y mostrar la solución (usando una de cuatro formas: dar una personalidad a la app, prototipado, hagámoslo fluir y mapa de historias). Para ello se plantean dos fases, en la primera fase, trabajar en el ciclo básico de formación de los profesionales TI para una o máximo tres cohortes del mismo período; en la segunda fase, trabajar en el ciclo especializado y/o de profundización según sea el diseño del currículo, para las cohortes de la fase anterior. En cada fase seleccionar por lo menos dos asignaturas por semestre académico, realizando una evaluación del estado de los Soft Skills al inicio y al final de cada fase para evaluar el éxito de la estrategia y de las prácticas docentes utilizadas para ejecutarla. Algunas de las actividades de enseñanza-aprendizaje que se pueden trabajarse son: proyectos de curso de cualquier asignatura, proyectos transversales o integradores, proyectos de fin de carrera. Para la evaluación de las actividades se podría usar una matriz de observación (lista de chequeo con fines de verificar mediante observación la presencia o ausencia de algo) que involucre una serie de comportamientos relacionados con los Soft Skills sujetos a medición en el curso (la selección de la escala estaría sujeto a los lineamientos curriculares propios de cada institución, del curso o del comité curricular del programa). Si es una actividad colaborativa entonces incluir además criterios que permitan verificar la presencia y efectividad de esta colaboración en cada uno de los equipos de trabajo.

REFERENCIAS

[1] *IEEE-CS and ACM, «Information Techonology Curricula 2017: Curicula Guidelines for Baccalaureate Degree Programs in Information Technology,» IEEE computer society, 2017.*

[2] *IEEE-CS and ACM, «Software Engineering 2014: Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering,» IEEE computer society, 2015.*

[3] *L. F. Capretz y F. Ahmed, «A Call to Promote Soft Skills in Software Engineering,» Psychol Cogn Sci Open J, 2018.*

[4] *L. Zapata, «Planes de estudio, Hablidades blandas y búsqueda de empleo: los casos de la facultad de Ingeniería y Ciencias Sociales de la Universidad de Buenos Aires y Trenes Argentinos: apuntes de una investitación en curso,» Graduate School of Business, nº 16, pp. 119-130, Nov. 2017.*

[5] *Human Age Institute. Manpower Group, «Soft Skills 4 Talent 2016,» 2016.*

[6] *Instituto de la Ingeniería de España. apd., «Informe sobre competencias adicionales requeridas a los ingenieros por las empresas españolas.,» 2018.*

[7] *C. Baluarte Araya, E. Vidal Duarte, L. Delgado Barra y E. Castro Gutierrez, «Integrating soft skills: Writing, communication and ethics in the curriculum of the System Engineering Professional School - UNSA,» de 15th LACCEI International Multi-Conference for Engineering, Education, and Technology: "Global Partnerships for Development and Engineering Education", Boca Raton FL, United States, 19-21 July 2017.*

[8] *«Propuesta de evaluación de habilidades blandas en Ingeniería de Software por medio de Proyectos Universidad-Empresa,» Revista Educación en Ingeniería, vol. 10, nº 19, pp. 131-140, Enero a Junio 2015.*

[9] *J. Rasmusson, The Agile Samurai: How Agile Masters Deliver Great Software, Dallas, Texas, 2010.*

[10] *E. Comba Riepenhausen, Inceptions: Starting a Software Projects, Lean Publishing, 2014.*

Biografía Autor(es)

a. Magister. Perteneciente al grupo de invetigación GrinTic – Universidad Antonio José Camacho. E-mail. mapastrana@admon.uniajc.edu.com. Cali, Colombia.

b. Phd. Perteneciente al laboratorio de Investigación de Ingeniería de Software, Universidad de San Buenaventura. E-mail: haordonez@usbcali.edu.co. Cali, Colombia.

c. Magister. Perteneciente al Laboratorio de Investigación de Ingeniería de Software – Universidad de San Buenaventura. E-mail: sandraguanarita@hotmail.com. Cali, Colombia.

La Elicitación de requisitos mediante la técnica Inception Deck en las empresas de Santiago de Cali

Sandra Guañarita (a)
Hugo Ordoñez (b)
Manuel Pastrana (c)

Resumen

El propósito de este artículo es mostrar el estado actual del proceso de ingeniería de requisitos, específicamente la etapa de elicitación de requisitos, en relación con la aplicación y uso de la técnica Inception Deck en empresas de Santiago de Cali, Valle del Cauca. Los resultados del estudio se basan en la percepción de los profesionales de la industria del software obtenida mediante prácticas talleres de la técnica y encuestas. El estudio asociado a este artículo corresponde a una investigación en curso.

Palabras clave: Ingeniería de software, Ingeniería de requisitos, elicitación de requisitos, Inception Deck.

Elicitation of requirements using the Inception technique Deck in companies Santiago from Cali

Abstract

The purpose of this article is to show the current status of the requirement engineering process specifically the requirements elicitation stage in relation to the application and use of the Inception Deck technique in companies of Santiago de Cali, Valle of the Cauca. The results of the study are based on the perception of professionals in the software industry obtained through practical workshops of the technique and surveys. The study associated with this article corresponds to an ongoing investigation.

Keywords: Software engineering, requirements engineering, requirements elicitation, home cover.

I. INTRODUCCIÓN

El análisis de requisitos se encarga de recopilar o elicitar mediante diversas técnicas como las entrevistas, los cuestionarios, los prototipos y la observación directa todas las necesidades y restricciones implicadas para la construcción de un aplicativo software como indica [1]. Actualmente tanto [1] como [2], clasifican los procesos de desarrollo de software en tradicionales (trabajan en secuencia o cascada) y ágiles (trabajan iterativo incremental). Esto significa que cualquier modificación realizada durante el proyecto impacta en menor escala cuando el proceso de desarrollo es ágil. Basado en esto, [3] revisa varios proyectos de desarrollo de software bajo ambas metodologías e identifica que las principales problemáticas dentro de la industria del desarrollo de software se centran en el análisis de requisitos. Adicionalmente, [3] también concluye que los proyectos agiles han generado mejores resultados en comparación con los tradicionales por su enfoque de participación de todos los interesados del proyecto, lo que también es afirmado por [4] quien justifica esto bajo el uso de la técnica inception deck.

En este artículo, se presenta un trabajo investigativo que permite conocer que tanta difusión tiene la técnica inception deck a nivel académico e industrial dentro de la ciudad de Cali. Lo anterior permitirá determinar la relevancia de incluir esta técnica dentro de los planes curriculares de las instituciones educativas superiores de la región, con el fin de aportar en el conocimiento de la técnica, lo impacta de manera positiva en la industria del software.

II. ESCENARIO DE MOTIVACIÓN

En respuesta a la necesidad de mejorar el inicio de los proyectos de software, específicamente el análisis de requisitos [5] propone la técnica inception deck que se centra en crear una visión unificada del proyecto a partir del consenso de las opiniones de todos los involucrados (cliente, analistas, stakeholders, desarrolladores, etc). En esta reunión, la elicitación de requisitos es guiada a través de 10 dinámicas, que permiten de manera objetiva dimensionar lo que se requiere hacer, los riesgos inherentes del proyecto, las estrategias para resolverlos, que personas son relevantes para lograr con éxito el proyecto, además de dimensionar el tamaño y nivel complejidad del mismo. Esta técnica fue refinada y difundida más ampliamente en [6]. En complemento, [7] expone un manual de las dinámicas utilizadas inception deck que son explicadas a continuación:

Primera dinámica ¿Por qué estamos aquí?: Esta actividad enfoca la visión del producto, permite identificar el contexto inicial del proyecto.

Segunda dinámica, Elevator Pitch: Esta dinámica consiste en identificar rápidamente cuál es la esencia o factores relevantes del proyecto. Al final todas las respuestas son socializadas al auditorio como mecanismo de unificación de puntos de vista, buscando llegar al consenso.

Tercera dinámica, Diseñar la caja del producto: Tiene como objetivo darle identidad al proyecto mediante darle un nombre del producto, una frase que lo identifica y enumerar los beneficios del producto.

Cuarta dinámica, generar la lista del no: Delimita el alcance de lo que será construido especificando las restricciones.

Quinta dinámica, Conoce el vecindario: Esta dinámica permite identificar como las personas que están en el auditorio interactúan entre sí, además, potenciar esas relaciones para lograr el éxito del proyecto.

Sexta dinámica ¿Qué nos impide dormir en las noches?: Esta dinámica se enfoca en identificar de manera temprana los posibles riesgos del proyecto y las estrategias que el equipo debe seguir para impedir que sucedan o en caso de contrario minimizar su impacto.

Séptima dinámica, Mostrar la solución: Esta dinámica permite realizar la definición funcional del sistema.

Octava dinámica ¿qué es lo que hay que dar?: Permite fijar los compromisos de todos los participantes para el éxito del proyecto.

Novena dinámica ¿qué es lo que costara hacerlo?: El objetivo de esta pregunta es que los participantes realicen sus apreciaciones sobre que costara hacerlo pensando en alcance, tiempo y costo frente a todo lo que hay que realizar.

Decima dinámica, Resumen o sumarización: Es el paso final y presenta un resumen de los resultados.

Basado en el trabajo de [6] y [7], [8] se comprueba el impacto del uso de esta técnica en diversos proyectos de desarrollo de software, concluyendo que el uso de inception deck repercute de manera positiva sobre la elicitación de requisitos, generado incremento en la calidad. Este panorama expone que el uso de la técnica puede beneficiar a la industria del desarrollo de software pero se hace necesario identificar que tan conocida es en el medio académico e industrial para ampliar su difusión.

III. PROPUESTA Y RESULTADOS

Para dar valides a la propuesta, se planearon varios talleres prácticos con un total de 200 profesionales, entre los cuales se tiene gerentes de proyectos, programadores y arquitectos. En este sentido, cada taller se realizó con las dinámicas sugeridas por [7] y se dividió en dos momentos. El primero consiste en una presentación de 45 minutos donde se explica la técnica inception deck en detalle y se resuelven inquietudes a los participantes sobre su aplicabilidad y soluciones a posibles problemas para su utilización, el segundo momento consistió en un trabajo práctico grupal en donde el auditorio se organizó en equipos de cinco a siete integrantes y cada equipo utilizando la técnica elicitar los requisitos de un proyecto de software asignado por el facilitador del taller. Posteriormente se aplica una encuesta para identificar si previamente conocían o no la técnica, la opinión sobre la utilidad que le ven dentro del ámbito laboral y los obstáculos identificados para su implementación y aprendizaje.

Del total de participantes el 58.3% conoce la técnica previamente, aunque destacan que capacitación fue por sus propios medios y no desde una institución académica universitaria, el restante 41.3% desconocía la técnica, como se aprecia en la Fig. 1. Los

talleres son ejecutados bajo las especificaciones de la sección anterior y posterior a esto se procede con las encuestas diseñadas previamente. Estas son enviadas al correo electrónico de 200 profesionales de diferentes sectores. El 75% de los encuestados tiene una experiencia profesional superior a 10 años, y apenas un 8.3% posee una experiencia profesional inferior a 3 años.

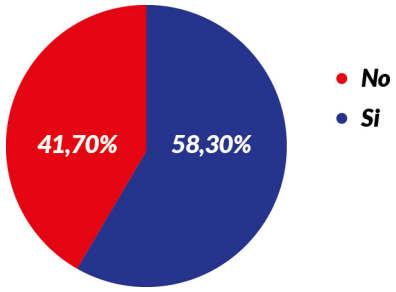


Fig. 1 *Resultados Encuesta de conocimiento previo al taller de la técnica Inception Deck*

Dado que el 41.7% de los encuestados indica que no conocían el inception deck se pregunta qué técnicas conocen y utilizan en sus trabajos. Los resultados indican que la entrevista y los prototipados se encuentran entre las más usadas, seguida por la observación directa, el análisis de dominio y el análisis de documentos. Utilizando una escala donde 1 es el menor valor posible de calificación y 5 el valor más alto, el 30% de los participantes califican la utilidad de la técnica Inception Deck con un valor de 4, el 60% con 5 y el 10% le asignó una calificación de 2.

Sobre los obstáculos para implementar la técnica en el ámbito laboral, se obtienen que las principales dificultades giran en torno a la cultura organizacional de las empresas, a la dificultad de reunir a todos los interesados, justificando a las altas gerencias la razón y costo de la reunión, el desconocimiento de los enfoques propuestos por las metodologías y frameworks agiles, la duración de la técnica y la motivación de los equipos de trabajo.

IV. CONCLUSIONES

Los resultados muestran que aunque algunos profesionales conocen diferentes enfoques para la elicitación de requisitos, la industria sigue trabajando con técnicas tradicionales por temas de cultura organizacional. Pese al impacto positivo en los proyectos de desarrollo de software identificados por varios autores y a la difusión que tiene a nivel mundial, la técnica Inception Deck todavía no se encuentra suficientemente reconocida en la comunidad de profesionales TI de Cali, a lo cual manifestaron que es una buena estrategia, fácilmente adopTable. por la mayoría de las empresas desarrollo, y permite incrementar de alguna forma los atributos de calidad del software a desarrollar.

Lo anterior sugiere que las instituciones universitarias deben difundir esta técnica enfocados en las necesidades de la industria y la mejora de las condiciones de productividad de la región. Esto es posible mediante la inclusión de la técnica en los currículos académicos, con el propósito de que ésta sea difundida y posteriormente adoptada como forma de trabajo en el proceso de elicitacion de requerimientos.

REFERENCIAS

[1] *R. S. Pressman and D. Ph, Ingeniería del software. .*

[2] *I. Sommerville, Ingeniería del Software. 2011.*

[3] *S. Hastie and S. Wojewoda, "Standish Group 2015 Chaos Report - Q&A with Jennifer Lynch," http://www.infoq.com/articles/standish-chaos-2015, 2015. [Online]. Available: http://www.infoq.com/articles/standish-chaos-2015.*

[4] *M. Pastrana, H. Ordóñez, A. Ordonez, L. H. Thom, and L. Merchan, "Optimization of the Inception Deck Technique for Eliciting Requirements in SCRUM Through Business Process Models," Bus. Process Manag. Work., vol. 4928, no. January, pp. 649–655, 2018.*

[5] *J. Rasmusson, "Agile project initiation techniques - The inception deck & boot camp," in Proceedings - AGILE Conference, 2006, 2006, vol. 2006, pp. 337–341.*

[6] *J. Rasmusson, The Agile Samurai–How Agile Masters Deliver Great Software. 2010.*

[7] *Enrique Comba Riepenhausen, "Inception Starting a New Project." .*

[8] *M. Pastrana, H. Ordóñez, A. Ordóñez, and L. Merchan, "Requirements elicitation based on inception deck and business processes models in scrum," in Communications in Computer and Information Science, 2017, vol. 735, pp. 327–339.*

Biografía Autor(es)

a. Magister. Pertenecientes a GrinTic – Universidad Antonio José Camacho. E- mail: sandraguanarita@hotmail.com. Cali, Colombia.

b. Phd. Perteneciente al laboratorio de Investigación de Ingeniería de Software, Universidad de San Buenaventura. E-mail: haordonez@usbcali.edu.co. Cali, Colombia.

c. Magister. Perteneciente a GrinTic – Universidad Antonio José Camacho. E-mail: mapastrana@admon.uniajc.edu.com. Cali, Colombia

Marco de referencia para el diseño de la presentación de localizaciones geográficas

William Joseph Orozco Giraldo (a)
Luisa Fernanda Londoño Celis (b)

Resumen

Este trabajo presenta una propuesta preliminar para abordar la complejidad que subyace al procesamiento de información relacionada con el uso de ubicaciones geográficas en los sistemas de información. Se presenta la motivación que origina este trabajo, se reflexiona el marco conceptual elaborado por el Instituto Geográfico Agustín Codazzi (IGAC) y el estado del arte y finalmente se presenta una propuesta conceptual para abordar esta problemática desde los principios de la Interacción Humano Computador (HCI). Se estudia la plataforma MUISCA específicamente el trámite de actualización del Registro Único Tributario (RUT) y se propone una posible mejora para la tarea de cambio de dirección del contribuyente. Se presenta en este trabajo el resultado de más de una década de investigación en forma de principio denominado "el principios de la granularidad".

Palabras clave: HCI, localización geográfica, granularidad, Gobierno en línea.

Framework for the presentation design geographic locations

Abstract

This paper presents a preliminary proposal to address the complexity that underlies the processing of information related to the use of locations geographic information systems. The motivation that originates this work, the conceptual framework developed by the Agustín Geographical Institute is reflected Codazzi (IGAC) and the state of the art and finally a conceptual proposal is presented to address this problem from the principles of Human Interaction Computer (HCI). The MUISCA platform is studied specifically the processing of update of the Single Tax Registry (RUT) and a possible improvement is proposed for the task of change of address of the taxpayer. The result of more than a decade of research in the form of a principle called "The principles of granularity".

Keywords: HCI, geographic location, granularity, Government online.

I. INTRODUCCIÓN

La gestión de información geográfica es una tarea vital en la mayoría de los sistemas de información por lo que se hace necesario el uso de patrones de negocio para tal fin. El patrón de ubicación geográfica es una propuesta confiable que se puede utilizar siempre que necesite modelar una dirección, ejemplos de dominios problemáticos son las empresas de venta por correo, la oficina postal, las agencias de envíos, las fuerzas de venta y los departamentos de cuentas por cobrar [17]. El patrón de ubicación geográfica es una solución abstracta que debe ser especializada y concretizada en el contexto propio en donde se hace presente la necesidad de gestionar dicha información, para nuestro caso, el territorio colombiano en cabeza del IGAC.

El IGAC gestiona el Sistema Nacional Catastral para lo cual conceptualiza la estructura del número predial nacional (único, permanente e independiente) con doce niveles de granularidad distribuidos en treinta categorías. Este número es utilizado por las instituciones que deseen gestionar la información de ubicaciones geográficas dentro de un marco de responsabilidad política, comercial, logística, etc. El caso de estudio en este trabajo es el trámite de actualización de documentos (TAD) del RUT.

La interacción de los ciudadanos con los TAD, en el contexto de gobierno en línea (GEL), debe estar concebida desde dos dimensiones: (a) la tecnológica, definida por el diseño de la interacción y la experiencia del usuario. (b) La del dominio del negocio, definida por una arquitectura empresarial que se orienta por el diseño del trámite ofreciendo atributos como: agilidad, oportunidad, accesibilidad, eficacia, entre otros.

En este trabajo se discute exclusivamente la problemática relacionada con la presentación de localizaciones geográficas desde la perspectiva del HCI, no se pretende abordar la problemática de los patrones de negocio de ubicación geográfica, principalmente, por considerar que este es un problema ampliamente evaluado y en la actualidad se encuentra estabilizado. Se estudia el caso de la actualización de dirección del RUT y se realiza una propuesta para la mejora de su interfaz de usuario.

En la sección dos se presenta una revisión de la problemática relacionada con la gestión de direcciones en el contexto de gobierno colombiano, luego se presenta la motivación de este trabajo, seguido se presenta el marco de referencia propuesto un caso de estudio y finalmente las conclusiones.

II. ESTADO DEL ARTE EN LA GESTIÓN DE DIRECCIONES

El IGAC (Instituto Geográfico Agustín Codazzi) [3][4], es la entidad encargada de la cartografía básica y la nomenclatura vial de Colombia, tiene como finalidad facilitar la planeación y ubicación de los predios y vías urbanas de una población. Colombia se caracteriza por tener una gran cantidad de accidentes geográficos [1][2] que generan diferencias en el sentido y orientación de las vías principales como: Calles, Carreras y las vías alternas como: Avenidas, Transversales, Circulares y Pasajes.

Por esto, es claro que en la actualidad se presentan problemas en la estandarización de la nomenclatura y para ellos se presentan propuestas de trabajos con el fin de me-

jorar y facilitar la problemática de la estandarización de direcciones, por ello se trae a colación la propuesta metodológica para el manejo de problemas en la estandarización de direcciones [5] que tiene como finalidad de elaborar una guía metodológica para el manejo de las direcciones urbanas, con el propósito de mejorar la normalización y calidad de los datos de las direcciones residenciales. Este trabajo estudia todos los conceptos generales de la nomenclatura vial y residencial mundial en países como: Europa, Australia, nueva Zelanda, reino unido, Japón, china, Beijín, corea del sur estados unidos y finalmente Colombia. Esta guía metodológica estudia la estructura de representación de direcciones residenciales reflejado en el informe realizado por el MEN (Ministerio de Educación Nacional) junto con el IGAC (Instituto Geográfico Agustín Codazzi) [3], en el que se puntualizan cinco tipos de direcciones residenciales basadas en las estructuras de: Malla Vial, Barrio-Manzana-Predio, Malla Vial/BarrioManzana-Predio, Malla Vial/Barrio y Sitios de interés, representada de la siguiente manera:

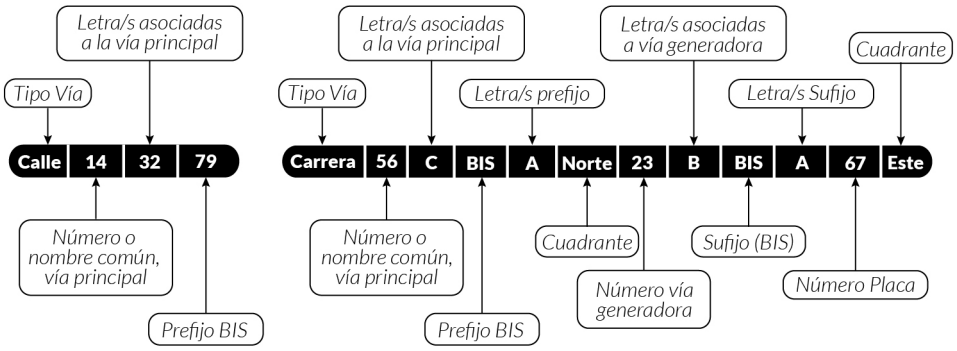


Fig. 1. *Estructura taxonómica propuesta para abordar la gestión de las direcciones [10]*

De esta misma manera se hace referencia a la estandarización de direcciones [10], basada en una estrategia para ajustar y actualizar la dirección de la base de datos de las instituciones educativas de Colombia por medio del SIG[11], que es un software desarrollado para automatizar la cartografía con herramientas como: Geocode [13] y ArcGis [14], para ayudar a resolver las diferentes necesidades relacionadas con la ubicación geográfica de los estudiantes del ITM(Instituto Tecnológico Metropolitano) [12] [15], en cuanto a la Geo-codificación, Localizador de dirección, Nomenclatura urbana, vial, predial, y Transcripción de direcciones urbanas, resaltando además que tampoco se pueden incluir observaciones en el mismo campo de la dirección, por lo que se recomienda crear un campo nuevo de Observaciones en la siguiente tabla.

Por otro lado, Los organismos gubernamentales de Colombia emplean diferentes mecanismos para identificar, ubicar y clasificar a las personas en la búsqueda de las direcciones. Por ejemplo, la DIAN [16] presenta nomenclaturas para el ingreso de direcciones residenciales de algunos servicios ofrecidos como el RUT (Registro Único Tributario) [7], convirtiéndose en referentes para otras entidades como: Dane [9], código postal Colombia [8], Colpensiones, Migración Colombia, Runt, entre otros. Sin embargo, es usual que muchas empresas como: ebay, Facebook, Hotmail, mercado libre, oriflame, sigep, simo, visado de Inglaterra, visa e imigracion no tengan definido un proceso de estandarización en el registro de sus direcciones residenciales y prefieran un modelo menos complejo para evitar cualquier tipo de contratiempo a la hora de validar la dirección, como el ingreso de los datos realizado en un campo de texto libre sin tener ningún orden, dando lugar a diferentes tipos de errores tipográficos, sintácticos, ortográficos, etc.

Finalmente, se puede inferir que el estudio de la nomenclatura de dirección en Colombia no hace que este se vuelva menos compleja, al contrario, son propuestas representadas por filas que pueden ocupar hasta 13 campos separados por bloques dentro de una misma cadena de texto, lo que no permite a las personas entender de una manera comprensible la estructura taxonómica de una dirección. Por tal motivo, es fundamental que las entidades del gobierno promuevan la adaptación de la nomenclatura y, en general, la normalización conceptual que requiere el desarrollo de las mismas, brindando así, calidad en los datos de una manera estandarizada.

III. MOTIVACIÓN

Esta investigación se enmarca bajo la ingeniería de software y tiene como finalidad aportar soluciones para la mejora de los trámites de gobierno en línea. Estos trámites son muy comunes en el contexto de GEL y afectan a una gran cantidad de ciudadanos por lo que es de vital importancia optimizarlos y mejorar la usabilidad de los mismos.

La naturaleza del patrón de diseño para la nomenclatura de dirección que utilizan los trámites del gobierno puede ser compleja debido a las variables involucradas, algunas son: confusión en la forma en cómo están organizados los datos, codificación de caracteres que se utilizan para la creación de una dirección, la información complemento de la dirección, la eliminación de la información existente, la lectura de dirección, entre otras.

Para entender todos los conceptos relacionados en la nomenclatura de dirección, se realiza un benchmarking con el fin de identificar los principales conceptos, las tareas relacionadas, los modelos de interacción asociados y la arquitectura de información, entre otros, utilizados en los trámites del gobierno. Algunos de los trámites estudiados son: RUT, RUNT, Migración Colombia, Colpensiones, Visado de Inglaterra, Dane (Censo), Sigep, Simo, Visas & Inmigración. Otras plataformas estudiadas son utilizadas constantemente por los usuarios como: Ebay, Facebook, Hotmail, mercadolibre, oriflame. El análisis de estos conceptos permite la agrupación de los mismos debido a su diferencia o parentesco, por otro lado, se cuenta con la conceptualización definida por el IGAC y es a partir de ésta en la que se basan la mayoría de trámites del gobierno, principalmente el RUT. En la figura 2 se presenta la interfaz para la tarea de cambio de dirección.

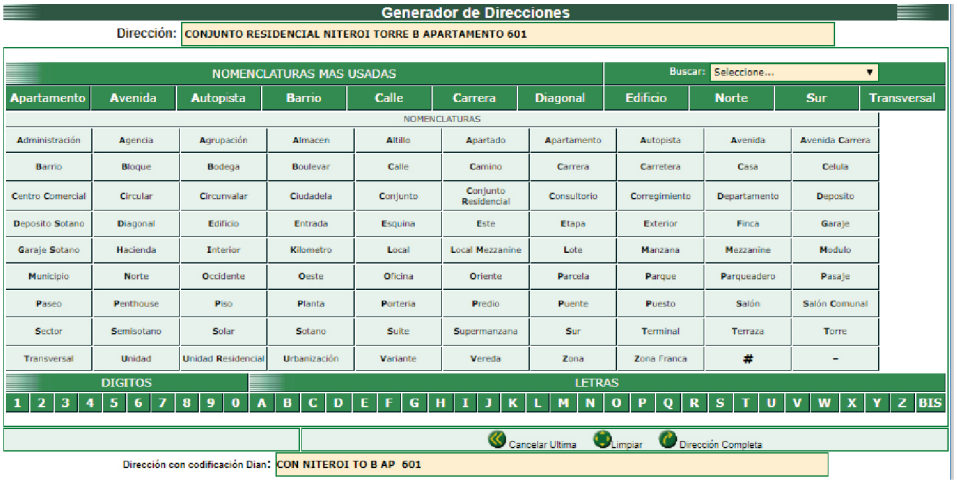


Fig. 2. *Generador de Direcciones Trámite de Actualización RUT*

Se selecciona el TAD del RUT para realizar pruebas de usabilidad en el laboratorio de la Universidad del Quindío y en el Vivelab Quindío. El propósito es evaluar la experiencia de los usuarios al actualizar la información de dicho trámite. La prueba consiste en actualizar la dirección principal del RUT en la plataforma MUISCA de la DIAN, con un usuario y una contraseña esTable.cidos.

La tarea principal de la prueba es que los usuarios actualicen la dirección actual y escriban una nueva que se les instauró y finalmente que generen un archivo Pdf con los datos corregidos. Para realizar la prueba se esTable.cen las siguientes directrices:

1. Intención de la prueba.
2. Ingreso a la plataforma: usuario y contraseña esTable.cidos.
3. Entrega de la nueva dirección.
4. Tiempo estimado para la actualización.
5. Resultado final: generación de archivo pdf.

Esta prueba se realizó a 24 usuarios con las siguientes características: Rango de edades: 25 a 45 años. Buen manejo tecnológico. Género masculino y femenino. Que tuvieran RUT.

De los 24 usuarios que realizaron la prueba solo dos culminaron con éxito la actualización de la dirección y los otros 22 se frustraron y desertaron del proceso por las siguientes razones:

Es confusa la presentación de los datos, es decir, la nomenclatura.

La dirección se lee y se percibe de dos formas diferentes: dirección completa y dirección con codificación Dian.

El usuario no sabe cómo eliminar la información existente.

Es confuso para el usuario el término “apellido” o complementación de la dirección.

El usuario percibe la misma información en espacios diferentes.

El usuario no percibe el resultado final, es decir, no genera el pdf con los datos actualizados.

Fig. 3. *Mapa de calor de la evaluación de la nomenclatura `para la dirección del RUT.*

Luego de realizar el estudio biométrico con usuarios y de analizar los movimientos oculares se comprueba que los usuarios realizan gran cantidad de movimientos erráticos en su intento de comprender las relaciones estructurales que tienen los conceptos presentados en la interfaz de usuario. Algunos usuarios argumentan que no sabían si esta interfaz se debe leer en sentido de arriba abajo o de izquierda derecha, pues reconocen que existen relaciones estructurales entre conceptos que pueden ser construidas en los dos ejes cardinales.

Luego de estudiar este diseño de interfaz de usuario para gestionar direcciones y de ahondar en la complejidad de abordar tal número de niveles de granularidad se decide elaborar una propuesta de marco de referencia para elaborar diseños estructurales para plataformas que soporten este tipo de sistemas.

IV. MARCO DE REFERENCIA

En esta sección se presenta una propuesta preliminar de un marco de referencia para el desarrollo de propuestas de diseño de la presentación de localizaciones geográficas desde la perspectiva del HCI. Este marco de referencia se elabora a partir del estudio de las bases conceptuales que soportan el diseño estructural de la información de las locaciones geográficas tanto a nivel de dominio como de la presentación de las interfaces de usuario. Adicionalmente a esto, se utilizan unos principios básicos relacionados con el manejo de la granularidad que han emergido como resultado de más de una década de estudios en HCI.

4.1 Principio de la granularidad

En esta sección se presentan algunas reflexiones relacionadas con la forma como los seres humanos interpretan la granularidad de las cosas. Los seres humanos están adecuadamente dotados de códigos y mecanismos para comprender y representar el mundo que los rodea y para crear los complementos y accesorios que hacen parte de su rutina diaria.

El número tres aparece como un número muy relevante para el ser humano y en cantidad de estudios se ha demostrado su importancia y utilización por parte del cerebro para codificar el mundo y realizar sus funciones primarias. Múltiples estudios realizados en nuestro laboratorio, en el campo del HCI, que tienen que ver con la forma cómo las personas realizan el particionamiento de los elementos que los rodean nos han llevado a identificar un patrón de comportamiento general.

Sin el ánimo de ser pretenciosos, y aun faltando un poco de trabajo formal, hemos decidido denominar a este comportamiento el principio de la granularidad. El principio de la granularidad representa la forma como los seres humanos realizan el particionamiento de las cosas para abordar su complejidad y comunicarse más fácilmente con otros. Se define como:

[Todo sistema complejo será presentado al mundo en un plano de renderización en el que dicho sistema se encuentra particionado utilizando solo tres niveles de granularidad. En

caso de requerirse varios planos para recorrer la presentación de dicho sistema, será necesario conmutar entre parejas de planos utilizando dos planos denominados: "primer plano" (foreground - punto focal de la acción a realizar) y "segundo plano" (background). La navegación entre planos de un sistema complejo se realizará mediante la conmutación de estos dos planos utilizando secuencias de a dos. En un recorrido hacia adelante, una conmutación se obtendrá cuando el primer plano pase a estar en un segundo plano y el nuevo primer plano será el plano en donde está todo el foco de la siguiente acción a realizar.]

La granularidad es la relación estructural entre dos elementos que denota que uno hace parte del otro en algún plano de referencia, bien sea abstracto o concreto, en el cual estos dos elementos puedan ser correlacionados. Tres elementos del mismo tipo, por tanto, tendrán en común un plano que denota la categoría de ser del mismo tipo, en este contexto, es claro que estos elementos tendrán alguna relación estructural. Sin embargo, la descomposición granular no está supeditada a que los elementos sean del mismo tipo, sino a que puedan pertenecer a un mismo plano en el cual puedan tener una relación estructural del tipo bien sea del tipo visual, semántica, emocional, cognitiva, etc.

Un nivel de granularidad define un plano en el que todos los elementos son del mismo nivel jerárquico (hermanos) y la distancia entre dichos elementos es uno independientemente de la geometría de dicho plano. Elementos en el mismo nivel de granularidad, respecto de un mismo elemento, pero que están en planos distintos tienen una distancia de dos y aunque tienen el mismo nivel jerárquico no son hermanos (primos). No confundir la interpretación del número tres respecto de los niveles de granularidad con la pregunta "¿cuál es el número de elementos que pueden estar en el mismo nivel de granularidad?", para esto existe la regla del 7 +/- 2. Un plano podría verse como una hoja impresa en donde está organizada la información de un sistema en tres niveles de agrupación o granularidad.

Ejemplos de planos basados en el principio de la granularidad son:

- **tiempo** = año[estaciones[meses]]. mes[semanas[días]]. día[franja horaria[hora]]. hora[cuarto de hora[minuto]]. año[estaciones[meses]].
- **geografía** = continente[región[país]]. país[departamento[ciudad]]. ciudad[-zona[sector]]. sector[comuna[barrio]]. barrio[manzana[terreno]].
- **serie** = serie[temporada[episodio]]
- **contenido** = pensum[semestres[asignaturas]]. asignatura[unidades[temas]].

Los seres humanos utilizan información en primer plano y en segundo plano para navegar entre niveles de granularidad. Por ejemplo: una persona que asiste a un congreso en otro país. En primera instancia, la persona debe dirigirse al aeropuerto desde su casa, por tanto, su segundo plano es: barrio[manzana[terreno]] y su primer plano es: barrio[manzana[terreno - aeropuerto]] si el aeropuerto está ubicado en la misma ciudad. Si por el contrario, si el aeropuerto está en otra ciudad, su primer plano es: país[departamento[ciudad]] -> ciudad[zona[sector]] y luego conmuta a: sector[-comuna[barrio]] -> barrio[manzana[terreno]]. Una vez en el aeropuerto, su segundo plano es: país de salida[departamento[ciudad - aeropuerto]] y su primer plano es: país de llegada[departamento[ciudad - aeropuerto]]. Una vez llega al aeropuerto se realizan conmutaciones entre planos de granularidad hasta que llega al sitio. Una vez en el sitio, inicia el proceso de ubicar el edificio, la planta piso y el local.

El diseño de una plataforma que soporte este viaje debe considerar que esta simple tarea requiere la conmutación de múltiples planos de granularidad. El principio de la granularidad indica que el cerebro prefiere ubicar solo tres niveles de granularidad en un mismo plano. Todo esto quiere decir que es necesario realizar un particionamiento de las pantallas para que la interacción con la información sea más natural. En muchos de los casos este particionamiento es más natural, para el ejemplo, algunos de los recorridos se hacen en diferentes medios de transporte. Si se pretende calcular los impuestos asociados al trayecto total es mejor entender los niveles de granularidad para entender cómo se acumulan los distintos impuestos según cada plano de granularidad.

4.2 Patrón de dominio de ubicación geográfica

En esta sección se estudian los patrones de negocio dispuestos en la literatura que definen el contexto del problema y de la solución de la ubicación geográfica. Eriksson y Penker definen en su libro "Business Modeling With UML: Business Patterns at Work" [17] el patrón de ubicación geográfica para evitar que el modelado de direcciones o ubicaciones puedan quedar obsoletas en un corto período de tiempo. La dirección de un lugar (donde se aloja una persona, una empresa, un gobierno, una entidad, etc) es uno de los datos más importantes en cualquier sistema de información, sin embargo, y a pesar de que existan patrones de dominio para tal fin, esto sigue siendo un problema para la mayoría de las empresas.

El concepto ubicación geográfica, para manejo de direcciones, ha sido históricamente Interpretado como una ubicación física geográfica específica. Una oficina de correo postal, por ejemplo, entiende una dirección como el número de una casa o de un edificio, o un nombre de la calle, o una ciudad y un departamento más un código postal y el nombre y código del país.

En las últimas décadas, el crecimiento de las operaciones digitales ha generado una nueva codificación del concepto de dirección geográfica pasando a ser un concepto más abstracto (dirección lógica) dejando de ser concreto (dirección física) por lo que se hace más difícil hacer un mejor procesamiento de esta información, por ejemplo para hacer un seguimiento a un envío.

Lo primero sería definir la relación que existe entre los conceptos dirección, lugar, ubicación y sitio. Se entiende por ubicación como un concepto geográfico que define un área geográfica bien delimitada en algún nivel de granularidad. Lugar es la ubicación, en algún nivel de granularidad (terreno en el primer plano de granularidad - figura 4 (columnas 18-21). Número en el segundo plano de granularidad - figura 4 (columnas 23-30)), en la que se aloja un sitio. Sitio es un concepto lógico utilizado para referirse a una persona, una empresa, un gobierno, una entidad, etc. Una dirección es un concepto lógico utilizado para conectar la ubicación de un sitio con procesos de negocio (logísticos).

Enviar un paquete a una empresa consiste en enviar dicho paquete a una oficina en el segundo piso de un edificio con salida en una calle principal que es en donde se aloja dicha empresa, entonces, geográficamente se puede decir que la empresa es el sitio que está ubicado dentro de un sector de la ciudad. La dirección es la forma como se etiqueta dicho lugar, sitio, empresa para que tenga visibilidad dentro de múltiples sistemas logísticos, por ejemplo para que un sistema de ventas pueda incluirlo dentro de una ruta de reparto de productos. Entender qué parte de una dirección corresponde a la ubicación geográfica y qué parte corresponde al sitio no resulta siempre obvio, pero, es necesario mantener esta separación.

Una dirección que especifica una posición geográfica (física) y un sitio (lógica) será más efectiva comparada con otra que especifica solamente la posición geográfica. Tanto las posiciones geográficas como los sitios están sujetos a composición en varios niveles de granularidad, inclusive mezcladas entre ellas, lo que hace más difícil su manejo.

Primer Componente	Segundo Componente	Tercer Componente	Cuarto Componente			Quinto Componente	Sexto Componente	Séptimo Componente	Octavo Componente		
DPTO	**MPIO**	**ZONA**	**SECTOR**	**COMUNA**	**BARRIO**	**MANZANA O VEREDA**	**TERRENO**	**CONDICIÓN DE PROPIEDAD**	**No. del Edif. o Torre**	**No. del Piso dentro del Edif.**	**No. de Unidad en PH o Mejora**
1 2	3 4 5	6 7	8 9	10 11	12 13	14 15 16 17	18 19 20 21	22	23 24	25 26	27 28 29 30
DANE-División Político administrativa	DANE-División Político administrativa	00 Rural 01 Urbano 02 a 99 Corregimientos	Sector	Comuna	Barrio	Manzana Vereda	Número de Orden de Terreno en la Manzana ó Vereda	0 NPH 9 PH 8 Condominio 7 Parques Cementerios 6 Mejoras por Edificaciones en Terreno Ajeno en PH 5 Mejoras por Edificaciones en Terreno Ajeno de Propiedades no Reglamentadas en PH 4 Vías 3 Bienes de Uso Público Diferentes a las Vías 2 No Ley 14	No. de Torre dentro del Terreno	No. Piso dentro del Edif. ó Torre	Unidad Predial ó Mejora

Fig. 4. *Clasificación en doce niveles de granularidad de una dirección*

En la figura 4 se presenta la estructura taxonómica del IGAC para la definición de una dirección o número predial. En esta clasificación se puede observar un grupo de columnas, que va desde la columna "departamento" hasta la columna "terreno", que se refiere a áreas geográficas que están contenidas unas dentro de otras, ocho niveles de granularidad en total. La columna de "condición de propiedad", seguidamente, no se refiere en sí misma a una región geográfica sino más bien se refiere a un sitio que puede ser identificado mediante una ubicación geográfica y en la mayoría de los casos mediante una dirección. El grupo "número de construcción" comprender las siguientes tres columnas y se refiere a regiones geográficas y sitios pero ambos especifican un cierto orden inferior en granularidad. Tal como se intentó ilustrar en el principio de granularidad, el mundo tal como lo percibe o lo modifica el ser humano conmuta entre planos de granularidad.

Propuesta de estructuración de los niveles de granularidad

La propuesta de estructuración de los niveles de granularidad para la gestión de direcciones se desarrolla en esta sección. Esta propuesta de estructuración debe abordar la complejidad que supone para un individuo el hecho de pensar en doce niveles de granularidad al mismo tiempo. Se propone una estructura taxonómica a partir de la cual se deben realizar los diseños de las plataformas que los ciudadanos usan para ingresar su dirección, todo esto, partiendo de la estructura taxonómica del IGAC y utilizando los fundamentos del principio de la granularidad.

En la figura 5 se presenta una propuesta de organización de las categorías definidas por el IGAC para la gestión de las nomenclaturas de números del predial. Esta propuesta está dividida en tres categorías:

Terreno (figura 5a): es el grupo de categorías que definen el terreno (lugar) en donde está ubicado un sitio, se supone que dicho terreno tiene un medio de acceso.

Punto de referencia (figura 5b): es la ubicación geográfica que denota una condición de propiedad de un sitio al que se puede acceder por algún medio. Frecuentemente, puede ser reconocido por la gente mediante una seña. Aunque puede ser un área geográfica no tiene interés como un área para buscar algo adentro, sino para que sirva como punto de referencia para identificar dicho sitio. En la imagen se denota como una burbuja de ubicación de google maps y al lado se ha puesto el icono de caminante que indica que se puede llegar al sitio, de alguna forma, guiado por el mapa.

Sitio (figura 5c): es la ubicación geográfica en donde se encuentra el sitio. Si este lugar tiene varios niveles de granularidad se hace necesario iterar una conmutación a un nuevo plano de granularidad pasando por la identificación del terreno y punto de referencia hasta llegar al sitio deseado.

La narrativa correcta en un plano de granularidad es: "ubicar el terreno y ubicar el punto de referencia que demarca el sitio". Hacer lo mismo para el siguiente plano de granularidad.

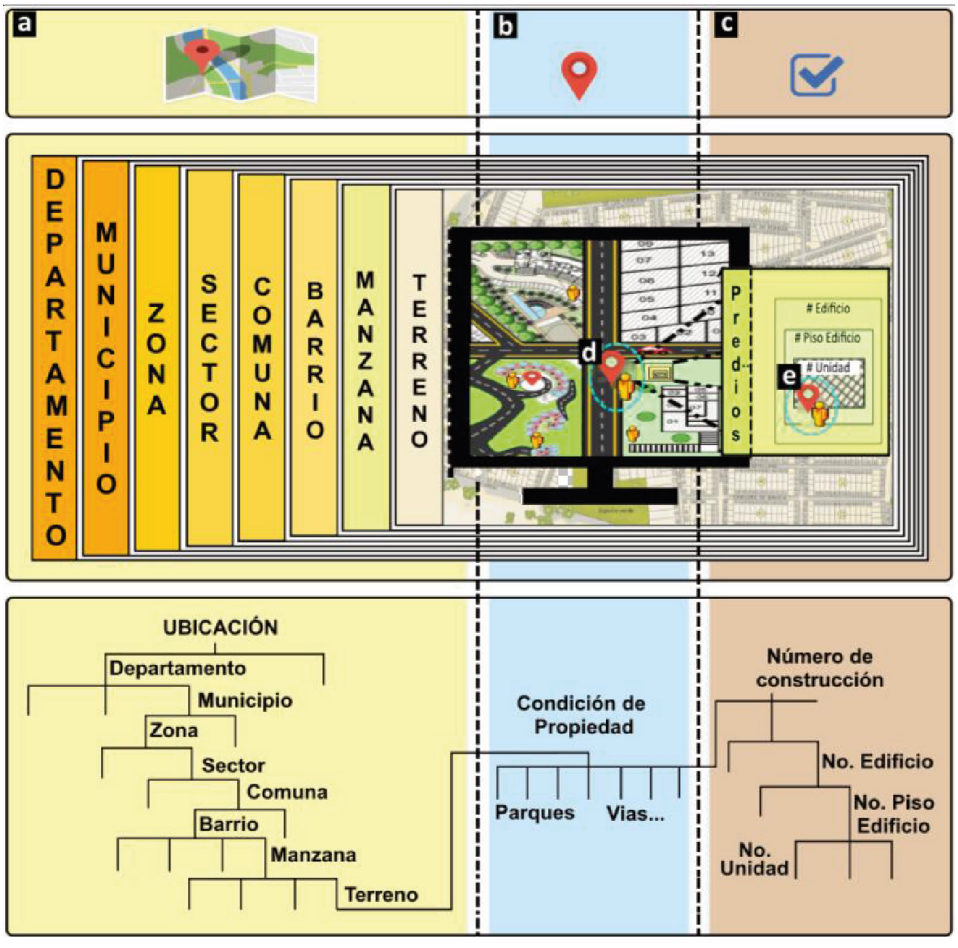


Fig. 5. *Propuesta de mejora nomenclatura dirección con tres niveles de granularidad*

La forma como se utilice esta clasificación va a depender del objetivo que la plataforma tenga respecto de la utilización de la información de las direcciones de los sitios. El particionamiento mínimo recomendado en un plano de granularidad para realizar la captura de una dirección es de tres partes: terreno, punto de referencia y sitio.

V. CASO DE ESTUDIO

En esta sección se presenta como caso de estudio la gestión de la dirección del trámite de actualización de documento (TAD) del RUT en el cual se utiliza la estructura taxonómica para gestionar las direcciones de los ciudadanos contribuyentes.

En la sección de la motivación se explica de dónde surge el interés por abordar esta problemática específica de la modificación de una dirección cuyo resultado debe ser un código de dirección de la DIAN.

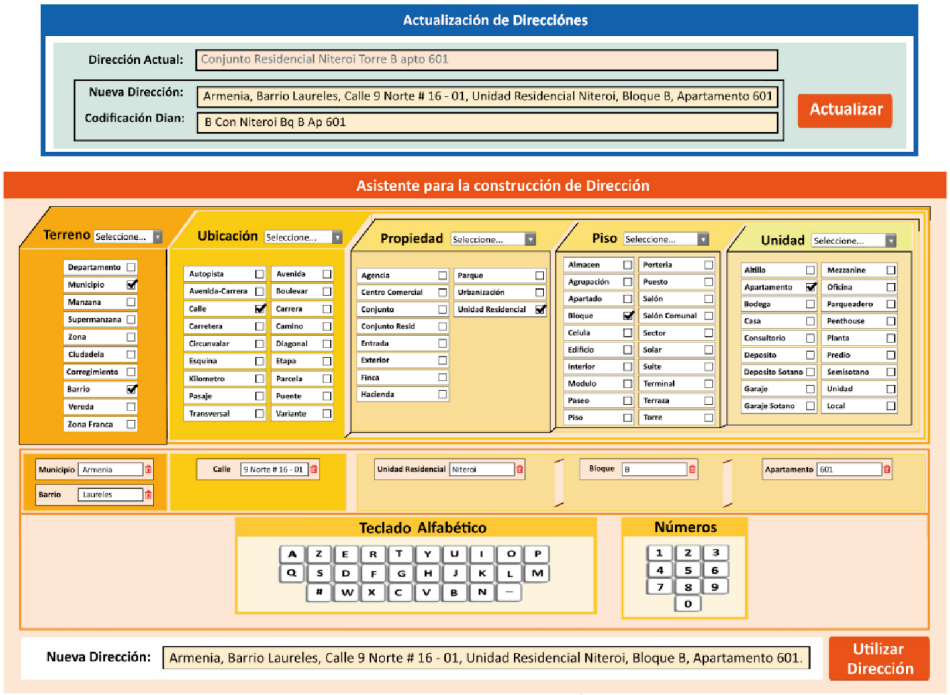


Fig. 6. *Propuesta de mejora de interfaz de usuario para la modificación de la dirección en el TAD del RUT.*

En la figura 6 se presenta la propuesta de mejora de la interfaz de usuario del TAD del RUT. Lo primero que hay que indicar es que el diseño del trámite es para una pantalla web por lo que es imposible visualizar dicho diseño correctamente en el tamaño de esta plantilla en papel.

La tarea de modificar la dirección del contribuyente es en sí un TAD por lo que el diseño debe considerar tanto el valor de la dirección actual como la nueva dirección a ingresar. Se recomienda separar el formulario de actualización de dirección del formulario donde se ubica el asistente para la construcción de una dirección según una norma específica. Esto evitará que el contribuyente realice enorme cantidad de movimientos de los ojos por estar haciendo tareas que no tengan concentrados los lugares en donde realizarlas.

El asistente para la elaboración de las direcciones está particionado en tres bloques: terreno, ubicación y sitio. En el bloque terreno, el usuario puede ir agregando categorías que hacen parte de cualquiera de los ocho niveles de granularidad que este bloque representa. En el bloque ubicación, el usuario indica el lugar en donde se ubica el sitio específicamente. En el bloque sitio, el usuario utiliza categorías de acuerdo a la complejidad del lugar.

Esta propuesta es el resultado de varias iteraciones en el proceso de mejorar esta tarea. En la actualidad se cuenta con una versión para dispositivos móviles que podría servir como un servicio genérico para todo aquel que tenga esta tarea dentro de su sistema de información.

VI. CONCLUSIONES

El diseño de sistemas de información que requieren la gestión de direcciones debe abordar la complejidad que subyace el manejo de múltiples niveles de granularidad. El diseño de la presentación de la interfaz de usuario de estos sistemas no debe simplemente disponer en una misma pantalla todas las categorías a utilizar porque es muy difícil para los ciudadanos realizar las asociaciones correctas.

El principio de la granularidad puede ser utilizado como eje rector del diseño de sistemas que involucran múltiples niveles de granularidad. Suponga el diseño de una plataforma para la venta de apartamentos los cuales están diseñados en varios tipos, distintas planta piso, distintos bloques de apartamento y con distintos enseres. Disponer toda esta información en una misma pantalla no sería práctico ni facilitaría la tarea de compra. Usando el principio de granularidad se puede diseñar la navegación necesaria entre pantallas (planos de granularidad) necesarias para reducir la complejidad y facilitar al usuario la identificación de una opción de compra.

REFERENCIAS

[1] *SIG_MEN, (2009). Propuesta estándar de las direcciones urbanas para los equipamientos del ministerio de educación.*

[2] *D. H. Á. Jaime Andrés Vargas Benjumea. (2013), proceso de geocodificación de direcciones en la ciudad de medellín, una técnica deterministica de georreferenciación de direcciones, USBMED, vol. 4, no. No. 1, p. 16.*

[3] *IGAC. (2013) Numero de predial nacional. "http://www.catastrolatino.org/documentos/foros_tematicos_IGAC/foro%20I/numero_predial_nacional_IGAC.pdf"*

[4] *IGAC. (2013), Estandarización de abreviaturas. "https://www.sistemamatriculas.gov.co/ayuda/direccciones.pdf"*

[5] *Daniel Felipe Rivas Burbano. (2016), propuesta de guía metodológica para el manejo de problemas en la estandarización y calidad de datos de direcciones urbanas en colombia, universidad pontificia bolivariana.*

[6] *DIAN. [Online], (revisado el 22 de octubre de 2018). Available: https://muisca.dian.gov.co/WebArquitectura/DefLogin.faces.*

[7] *Código- postal.co. [Online], (revisado el 22 de enero de 2019) Available: https://codigo-postal.co/colombia/. [9] https://www.dane.gov.co/*

[8] *A. G. V. Darryn Ibañez Ardila. (2015), estandarización de direcciones de la base de datos de estudiantes del itm para realizar consultas georeferenciadas y determinar zonas de influencia dentro del área metropolitana, instituto tecnológico metropolitano.*

[9] *Sistema de información geográfico del sector educativo. (revisado el 22 de enero de 2019) [Online]. Available: https://www.mineducacion.gov.co/1621/article-190610.html.*

Biografía Autor(es)

a. Doctor en Arquitectura y gestión de la información y del conocimiento en sistemas en red. Docente, Universidad del Quindío. E-mail: wjgiraldo@uniquindio.edu.co. Quindío, Colombia.

PARLANA: Sistema Experto para determinar el nivel de inglés de un estudiante según el Marco Común Europeo de Referencia (MCER) para las Lenguas en los estudiantes de la Facultad de Ingeniería de la Universidad de Nariño

Alfonso Joaquín Patiño Galarza (a)
Dario Fernando Criollo Estrada (b)
Jairo Antonio Guerrero García (c)
Sandra Marleni Vallejo Chamorro (d)

Resumen

En este proyecto se llevó a cabo la construcción de un sistema experto orientado a la web que determina el nivel de inglés de los estudiantes de la Facultad de Ingeniería de la Universidad de Nariño según el Marco Común Europeo de Referencia (MCER) para los Idiomas, que es una prueba inteligente que permitió saber cuál es el nivel de inglés del alumno que realizó la prueba.

De esta forma PARLANA ha sido la denominación que ha recibido el sistema experto desarrollado, permitió a los estudiantes realizar una prueba disponible en la web y donde los estudiantes se registraron llenando un formulario en el que crearon una cuenta que una vez realizado un registro exitoso permitió a los estudiantes el acceso a la plataforma y hacer la prueba. Por lo tanto, cada vez que se realizó la prueba para los estudiantes dio un resultado que se guardó en la base de datos de PARLANA; que sirve para crear un historial de cada usuario, con el fin de evidenciar el progreso de cada alumno.

Así con este proyecto se intentó contribuir a mejorar el nivel de inglés de los alumnos de la Facultad de Ingeniería de la Universidad de Nariño para que la apropiación de un segundo idioma no es un problema en el momento de solicitar una beca extranjera y/o realizar un intercambio de estudiantes, también esto mejoró el perfil internacional de la universidad generando un salto de calidad de forma inmediata.

Palabras clave: Bilingüismo, inglés, MVC, MCER, sistemas expertos.

PARLANA: Expert System for determine the level of English of a student according to the Common Framework European Reference (CEFR) for languages in students of the Faculty of Engineering of the University of Nariño

Abstract

In this project carried out the construction of a web-oriented expert system that determines the level of english of the students of the Engineering Faculty of the University of Nariño according to the Common European Framework of Reference (CEFR) for the Languages, which is an intelligent test that allowed to know what is the level of english of the student that performed the test.

In this way PARLANA has been the denomination that has received the expert system developed, allowed to the students to perform a test available in the web and where the students registered filling out a form in which they created an account that once performed a successful registration allowed to the students the access to the platform and to do the test. Thereby each time that the test was performed for the students gave a result that was saved in the PARLANA's database; which serves to create a history for each user, with the purpose to evidence the progress of each student.

Thus with this project tried to contribute to improving the level of english of the students of the Engineering Faculty of the University of Nariño so that the appropriation of a second language do not be a problem in the moment to a apply for a foreign scholarship and/or to make a student exchange, also this upgraded the international profile of the university generating a quality jump immediately.

Keywords: Bilingualism, english, MVC, CEFR, expert systems.

I. INTRODUCCIÓN

Hoy en día el mundo exige que para competir a nivel mundial cada persona debe tener una profesión definida, para esto la Universidad de Nariño brinda la posibilidad que las personas obtengan su título de pregrado. Posicionada como una de las universidades más reconocidas a nivel nacional y que en 2018 fue incluida en el ranking mundial de SCImago (SIR WORLD 2018) [1] que incluye otras 27 universidades colombianas. Ganando esta distinción gracias a la publicación de al menos 100 documentos científicos durante el año concurso. Así, siempre está buscando la mejora constante, mejora que se ve reflejada en los integrantes de la comunidad educativa, en el departamento y también en la nación.

De esta manera los directivos recientemente han logrado que la universidad sea reconocida como una institución certificada de alta calidad y también ya hay en marcha planes para buscar la internacionalización y una posterior certificación, por ello se ha hecho énfasis en la apropiación del idioma inglés en la Facultad de Ingeniería para que de esta manera sea un aporte más a los procesos que se vienen llevando. Se hace énfasis en el inglés debido a que es el idioma mundialmente conocido y es innegable que esto permanecerá así por un largo tiempo.

Con lo anterior y para apoyar los procesos que la universidad está llevando a cabo se ha realizado un sistema experto denominado PARLANA, que permite medir el nivel de inglés de los estudiantes de la Facultad de Ingeniería de la Universidad de Nariño, apoyando en la labor educativa del aprendizaje de una lengua extranjera que en nuestra universidad está enfocada al inglés, para que el estudiante pueda ver su progreso a medida que avanza en los niveles de inglés ofrecidos por la universidad y que una vez finalizados todos los mismos esté en capacidad de aprobar con un alto nivel de eficiencia el examen propuesto, generando confianza para que el inglés ya no sea una barrera.

II. METODOLOGÍA

2.1 Paradigma de la Investigación

Para este proyecto se realizó una investigación social cuantitativa que según Ipes[2] se basa en información cuantificable con el fin de demostrar los fenómenos que se ponen a consideración, teniendo en cuenta que los resultados obtenidos estarán basados en una estructuración lógica. Así el objetivo de una investigación cualitativa es obtener los conocimientos fundamentales que permitan conocer la realidad de una manera imparcial por medio del modelo más adecuado; debido a que se recolectan y se procesan los datos a través de conceptos y variables.

2.2 Enfoque de la Investigación

El enfoque de esta investigación es descriptivo porque según Cazau[3] se considera investigación descriptiva cuando se selecciona un conjunto de variables y se realiza la medición de cada variable por separado. Con lo anterior se puede analizar el

comportamiento de un fenómeno y sus componentes. De esta manera con PARLANA se intenta medir el nivel de inglés de los estudiantes de la Facultad de Ingeniería de la Universidad de Nariño.

2.3 Población y Muestra

2.3.1 Población

La población elegida para esta investigación han sido los estudiantes de la Facultad de Ingeniería; teniendo en cuenta solamente a los estudiantes que se encuentran cursando la carrera en la sede de Pasto, descartando los estudiantes egresados, graduados y que se encuentren cursando el último semestre.

Table. 1. *Estado de los estudiantes de la Facultad de Ingeniería de la Universidad de Nariño - Sede de Pasto.*

Estado del Estudiante	***Cantidad de estudiantes***
Cursando	*356*
Egresado	*102*
Graduado	*20*
TOTAL	***478***

Fuente: Centro de Informática de la Universidad de Nariño.

Como se puede observar en la Tabla 1, se ha decido tomar la población activa del período de B de 2018 porque facilitará la investigación cuando se realicen los test.

Table. 2. *Estudiantes por programa en la Facultad de Ingeniería de la Universidad de Nariño – Sede Pasto.*

Programa	***Cantidad de estudiantes***
Cívil	*113*
Electrónica	*87*
Sistemas	*95*
TOTAL	***295***

Fuente: Centro de Informática de la Universidad de Nariño.

Según la Tabla 2, para el período 2018-B se encuentran matriculados 295 estudiantes en la Facultad de Ingeniería de la Universidad de Nariño repartidos en cada programa de la siguiente manera: Cívil (113), Electrónica (87) y Sistemas (95).

Table. 3. *Estratificación por semestre de los estudiantes del programa de Ingeniería Civil de la Universidad de Nariño – Sede Pasto.*

Semestre	Cantidad de estudiantes
Semestre 2	13
Semestre 4	24
Semestre 6	26
Semestre 8	24
Semestre 10	26
TOTAL	**113**

Fuente: Centro de Informática de la Universidad de Nariño.

Table. 4. *Estratificación por semestre de los estudiantes del programa de Ingeniería Electrónica de la Universidad de Nariño – Sede Pasto.*

Semestre	Cantidad de estudiantes
Semestre 2	13
Semestre 4	9
Semestre 6	19
Semestre 8	20
Semestre 10	26
TOTAL	**87**

Fuente: Centro de Informática de la Universidad de Nariño.

Table. 5. *Estratificación por semestre de los estudiantes del programa de Ingeniería de Sistemas de la Universidad de Nariño – Sede Pasto.*

Semestre	Cantidad de estudiantes
Semestre 2	14
Semestre 4	16
Semestre 6	21
Semestre 8	22
Semestre 10	22
TOTAL	**95**

Fuente: Centro de Informática de la Universidad de Nariño.

2.3.2 *Muestra*

Según Adimen[3] el tamaño de la muestra serán 100 estudiantes; a continuación se muestran los resultados obtenidos y los valores usados para dicho cálculo:

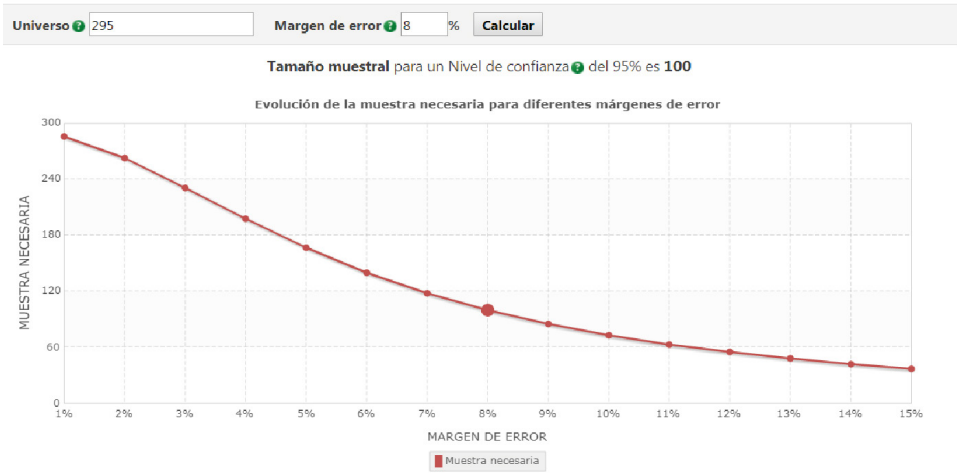


Fig. 1: *Calculadora de muestras. Muestra necesaria. [4]*

Con la muestra obtenida anteriormente se tomó en cuenta la cantidad de estudiantes por cada programa, obteniendo los siguientes porcentajes:

Table. 6. *Porcentaje de estudiantes por programa en la Facultad de Ingeniería de la Universidad de Nariño – Sede Pasto.*

Programa	Cantidad de estudiantes	Porcentaje (%)
Cívil	*113*	*38,31*
Electrónica	*87*	*29,49*
Sistemas	*95*	*32,20*
TOTAL	**295**	**100,0%**

Fuente: Esta investigación.

Con la información de la Tabla 6, se podemos inferir que para obtener unos resultados más precisos de la muestra necesaria (100) se requiere que 38,31% de los estudiantes pertenezcan a Ingeniería Cívil, 29,49% de los estudiantes pertenezcan a Ingeniería Electrónica y que un 32,20% pertenezcan al Ingeniería de Sistemas.

De esta manera la cantidad muestral aproximada cada programa; Cívil, Electrónica y Sistemas respectivamente es de 38, 30 y 32 estudiantes.

Basados en porcentajes obtenidos anteriormente y en una estrategia de estratificación muestral, se obtuvieron los siguientes resultados:

Table. 7. *Cantidad muestral de estudiantes por semestre del programa de Ingeniería Cívil de la Universidad de Nariño – Sede Pasto.*

Semestre	Cantidad de estudiantes	Porcentaje (%)	Cantidad muestral
Semestre 2	13	11,50	4
Semestre 4	24	21,24	8
Semestre 6	26	23,0	9
Semestre 8	24	21,24	8
Semestre 10	26	23,01	9
TOTAL	**113**	**100,00%**	**38**

Fuente: Esta investigación.

Table. 8. *Cantidad muestral de estudiantes por semestre del programa de Ingeniería Electrónica de la Universidad de Nariño – Sede Pasto.*

Semestre	Cantidad de estudiantes	Porcentaje (%)	Cantidad muestral
Semestre 2	13	14,94	4
Semestre 4	9	10,34	3
Semestre 6	19	21,84	7
Semestre 8	20	22,99	7
Semestre 10	26	29,89	9
TOTAL	**87**	**100,00%**	**30**

Fuente: Esta investigación.

Table. 9. *Cantidad muestral de estudiantes por semestre del programa de Ingeniería de Sistemas de la Universidad de Nariño – Sede Pasto.*

Semestre	Cantidad de estudiantes	Porcentaje (%)	Cantidad muestral
Semestre 2	14	14,74	5
Semestre 4	16	16,84	6
Semestre 6	21	22,10	7
Semestre 8	22	23,16	7
Semestre 10	22	23,16	7
TOTAL	**87**	**100,00%**	**32**

Fuente: Esta investigación.

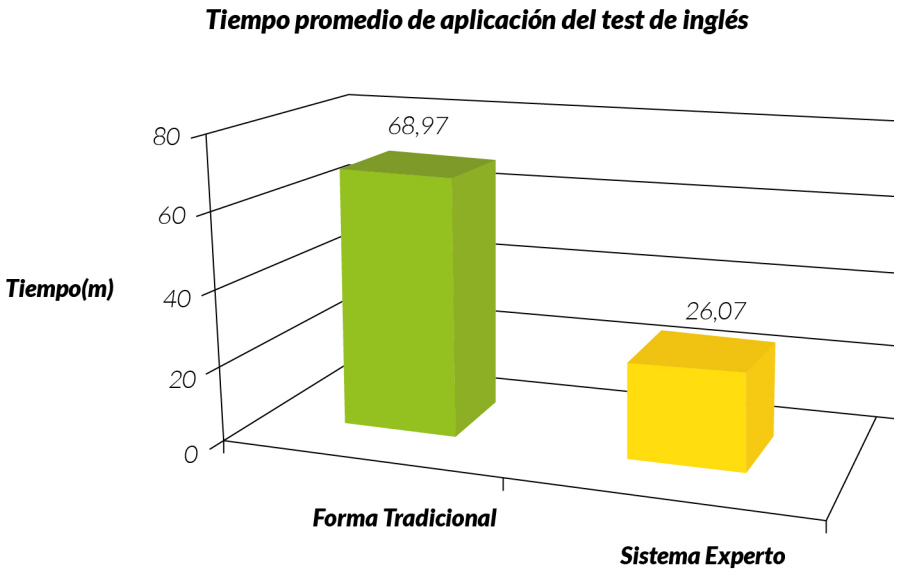
2.4 Instrumentos de Recolección de Información

En esta investigación se aplicó un instrumento denominado test de inglés, el cuál fue proporcionado por el Departamento de Lingüística e Idiomas de la Universidad de Nariño, el cual permitió conocer el nivel de inglés de los estudiantes de la Facultad de Ingeniería de la Universidad de Nariño; permitiendo analizar y documentar los resultados obtenidos. Además, este test fue previamente valorado por expertos en el tema, quienes han trabajado durante bastante tiempo para la obtención del producto; motivo por el cual se convierte en una fuente confiable para esta investigación.

III. RESULTADOS Y DISCUSIÓN

Para realizar la investigación se trabajó con 100 estudiantes de los programas de Civil, Electrónica y Sistemas de la Facultad de Ingeniería de la Universidad de Nariño. Los resultados obtenidos con la muestra se presentan a continuación:

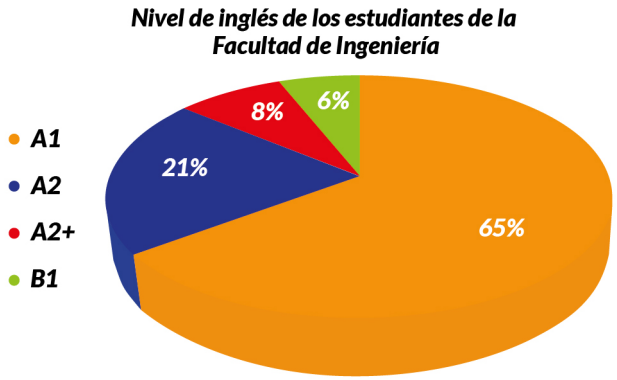
Gráfica 1. *Tiempo promedio de aplicación del test de inglés.*



Fuente: Esta investigación.

En la Gráfica 1, se observa como el tiempo promedio de realización del test usando el Sistema Experto fue de 26,07 minutos mientras que de la Forma Tradicional fue de 68,97 minutos. Por lo cual se comprueba que es más eficiente el sistema experto. Además la Forma Tradicional de aplicación no le proporciona al usuario unos resultados inmediatos, a diferencia del sistema experto que inmediatamente finalizado el test presenta los resultados al usuario.

Gráfica 2. *Nivel de inglés los estudiantes de la Facultad de Ingeniería de la Universidad de Nariño.*



Fuente: Esta investigación.

En la Gráfica 2 se puede apreciar cual es el nivel de inglés de los estudiantes de la Facultad de Ingeniería de esta manera los resultados arrojan qué:

- 65 de cada 100 estudiantes que realizaron el test de inglés usando PARLANA lo cual corresponde al 65% de la muestra tienen un nivel de inglés A1 y que el MCER lo cataloga como un USUARIO BÁSICO (Acceso).
- 21 de cada 100 estudiantes que realizaron el test de inglés usando PARLANA lo cual corresponde al 21% de la muestra tienen un nivel de inglés A2 y que el MCER lo cataloga como un USUARIO BÁSICO (Plataforma).
- 8 de cada 100 estudiantes que realizaron el test de inglés usando PARLANA lo cual corresponde al 8% de la muestra tiene un nivel de inglés A2+ y que el MCER lo cataloga como un USUARIO BÁSICO (Plataforma).
- 6 de cada 100 estudiantes que realizaron el test de inglés usando PARLANA lo cual corresponde al 6% de la muestra tienen un nivel de inglés B1 y que el MCER lo cataloga como un USUARIO INDEPENDIENTE (Umbral).

El análisis anterior fue realizado en base al gráfico propuesto por el MCER para así poder clasificar los resultados obtenidos.

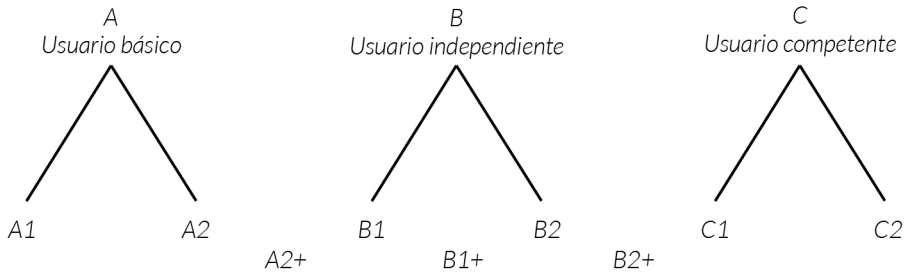


Fig. 2. *Niveles comunes de referencia según el MCER [5].*

De esta manera para comprender con más claridad la es los niveles que el MCER propone, tenemos las definiciones exactas de cada nivel, así:

NIVEL A1 (Acceso): Es considerado con el nivel más bajo en este punto el alumno es capaz de interactuar de la manera más sencilla, planteando y contestando preguntas él, sobre donde vive, sobre las personas que conoce y las cosas que tiene. Su capacidad permite hacer afirmaciones sencillas sobre todo en áreas de necesidad inmediata, temas cotidianos, sabiendo responder dejando de depender de un repertorio. [Ibíd.]

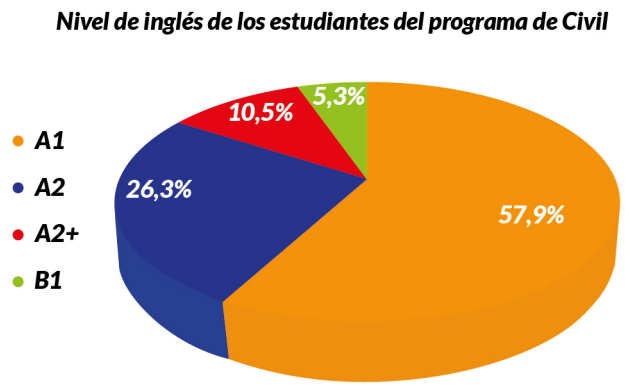
NIVEL A2 (Plataforma): En este nivel se consideran las funciones sociales. Tales como formas habituales de saludar, dirigirse amablemente, saludar personas y desenvolverse en intercambios sociales muy breves. Es capaz de hacer y contestar preguntas sobre su trabajo y su tiempo libre, puede hacer una invitación y también responder a ella, además de discutir a donde ir, arreglar una cita y también realizar un ofrecimiento y aceptarlo. [Ibíd.]

NIVEL A2+ (Plataforma): La diferencia con el usuario de A2 es que en el nivel A2+ un usuario es más activo, tiene ciertas limitaciones que logra sortear con cierta ayuda. Este usuario es capaz, de iniciar y terminar una conversación sin un esfuerzo excesivo, será capaz de realizar interlocuciones con la ayuda adecuada y en situaciones sencillas fáciles de predecir, es propio de este usuario corregir los mensajes e intentar buscar algunas palabras, en fin la interacción resulta fácil. La participación en debates abiertos es limitada pero los monólogos son una capacidad característica de estos usuarios. [Ibíd.]

NIVEL B1 (Umbral): Se basa en 2 características principales: La primera es la capacidad de mantener una interacción y de hacerse entender en diferentes situaciones y la segunda característica es la capacidad de saber cómo enfrentar de forma flexible problemas cotidianos. Este usuario está en la capacidad de entender debates extensos siempre y cuando sean articulados con claridad, su lenguaje es amplio pero sencillo.

En ocasiones resulta entenderle lo que dice, pero por lo general su comunicación es comprensible evidenciando cierta lentitud porque planifica gramaticalmente. La segunda es la capacidad de enfrentar problemas cotidianos de manera flexible y con naturalidad, sobre todo es un usuario que puede enfrentarse a situaciones atípicas, teniendo la habilidad de plantear quejas, tomar la iniciativa en una entrevista, aunque por lo general necesita de un guía que le pueda guiar en la conversación.[Ibíd.]

Gráfica 3. *Nivel de inglés los estudiantes del programa de Civil.*

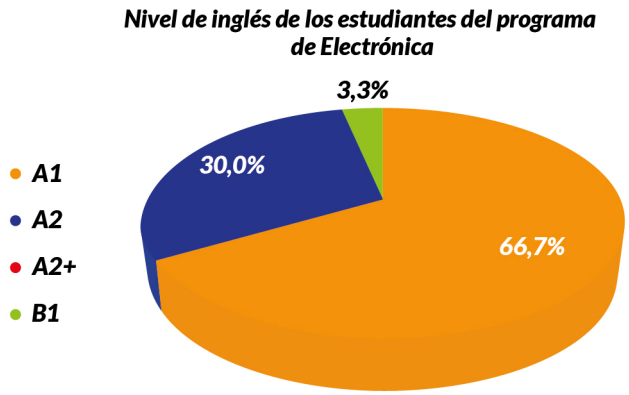


Fuente: Esta investigación.

La gráfica 3 correspondiente a los estudiantes que pertenecen al programa de Ingeniería Civil de la Facultad de Ingeniería de la Universidad de Nariño sede Torobajo, arrojó los siguientes resultados:

- Un 57.9% de estudiantes del programa de Civil tienen un nivel de inglés A1 correspondiente a Usuarios Básicos (Acceso).
- Un 26.3% de estudiantes del programa de Civil tienen un nivel de inglés A2 correspondiente a Usuarios Básicos (Plataforma).
- Un 10.5% de estudiantes del programa de Civil tienen un nivel de inglés A2+ correspondiente a Usuarios Básicos (Plataforma).
- Un 94.7% de estudiantes del programa de Civil no alcanzan el nivel B1 que es el requisito esTable.cido por la universidad, y solamente el 5.3% de los estudiantes cumplen con el requisito.

Gráfica 4. *Nivel de inglés los estudiantes del programa de Electrónica.*

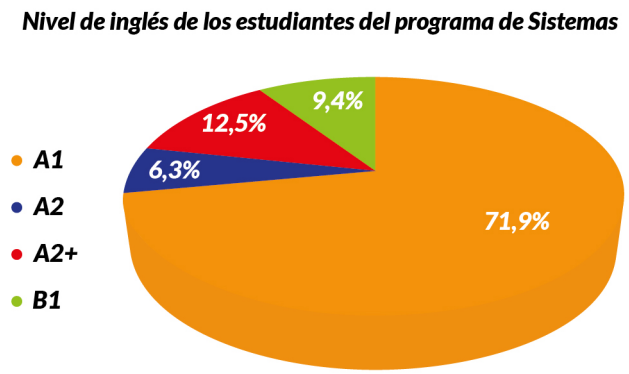


Fuente: Esta investigación.

La gráfica 4 correspondiente a los estudiantes que pertenecen al programa de Ingeniería Electrónica de la Facultad de Ingeniería de la Universidad de Nariño sede Torobajo, arrojó los siguientes resultados:

- Un 66.7% de estudiantes del programa de Electrónica tienen un nivel de inglés A1 correspondiente a Usuarios Básicos (Acceso).
- Un 30.0% de estudiantes del programa de Electrónica tienen un nivel de inglés A2 correspondiente a Usuarios Básicos (Plataforma).
- Ningún estudiante del programa de Electrónica tiene un nivel de inglés A2+ correspondiente a Usuarios Básicos (un grado elevado de Plataforma).
- Un 96.7% de estudiantes del programa de Electrónica no alcanzan el nivel B1 que es el requisito esTable.cido por la universidad, y solamente el 3.3% de los estudiantes cumplen con el requisito.

Gráfica 5. *Nivel de inglés los estudiantes del programa de Sistemas.*



Fuente: Esta investigación.

La gráfica 5 correspondiente a los estudiantes que pertenecen al programa de Ingeniería de Sistemas de la Facultad de Ingeniería de la Universidad de Nariño sede Torobajo, arrojó los siguientes resultados:

- Un 71.9% de estudiantes del programa de Sistemas tienen un nivel de inglés A1 correspondiente a Usuarios Básicos (Acceso).
- Un 6.3% de estudiantes del programa de Sistemas tienen un nivel de inglés A2 correspondiente a Usuarios Básicos (Plataforma).
- Un 12.5% de estudiantes del programa de Sistemas tiene un nivel de inglés A2+ correspondiente a Usuarios Básicos (Plataforma).
- Un 90.6% de estudiantes del programa de Sistemas no alcanzan el nivel B1 que es el requisito esTable.cido por la universidad, y solamente el 9.4% de los estudiantes cumplen con el requisito.

IV. CONCLUSIONES

Luego de reconocer la fundamentación teórica que nos brinda el MCER y de haber realizado las reuniones con el grupo de expertos en idioma inglés fue posible la definición de los parámetros de evaluación, la consolidación de las reglas del Sistema Experto PARLANA y también la comprensión e interpretación de los resultados obtenidos.

La implementación de PARLANA ha sido un éxito en la comunidad universitaria de la Universidad de Nariño porque demostró como los Sistemas Expertos aventajan a la Forma Tradicional de aplicación de un test; entre las ventajas tenemos eficiencia, usabilidad, portabilidad, y estadísticas en tiempo real lo cual demuestra que la herramienta permite mejorar el proceso; haciéndolo más eficiente determinando el nivel de

aprendizaje que tiene el usuario en menor tiempo y con la presentación de los resultados en tiempo real.

PARLANA se consolidó como una herramienta pionera en proyectar procesos educativos multidisciplinarios enfocados en la Internacionalización, dado que en la Universidad de Nariño existen planes de mejoramiento del segundo idioma inglés y PARLANA proporcionó los resultados y su respectivo análisis para atacar las falencias detectadas respecto al manejo del inglés dentro de la Facultad de Ingeniería de la Universidad de Nariño.

PARLANA permite a los usuarios realizar el seguimiento de su nivel de inglés mediante el historial que muestra el resumen de los exámenes que han presentado, motivando a los estudiantes a que estén mejorando continuamente, porque se puede acceder a los resultados en tiempo real añadiendo que son muy detallados y de esta manera el usuario observó cuáles fueron sus debilidades y fortalezas en el manejo del idioma inglés sin tener que esperar a que recibir una calificación por parte del docente.

Este estudio ha demostrado la eficiencia del Sistema Experto PARLANA cuando se compara respecto a las formas tradicionales de realización de un test; así la eficiencia del sistema experto hace que sea posible obtener los resultados en tiempo real acompañados de unas estadísticas detalladas del resultado obtenido, lo cual convierte a PARLANA en una herramienta que contribuye al mejoramiento del nivel de inglés.

PARLANA ha resultado muy atractivo para la comunidad educativa de la Universidad de Nariño por lo cual fue seleccionado y presentado en varios eventos científicos en la región (ver Anexo F), los cuales se nombran a continuación:

- Convocatoria de Investigación Docente, Estudiantil y Trabajos de Grado 2017. Aprobado mediante Resolución No. 079 del 27 de Junio del 2017.
- III Congreso Andino de Computación, Informática y Educación CADIED 2017.
- 1er Encuentro de Jóvenes Investigadores GALERAS 2017.
- 2do Encuentro de Jóvenes Investigadores GALERAS 2018.
- POLICARPA, Radar de Ideas Inspiradoras, Buenas Prácticas e Innovación Educativa en la Universidad de Nariño.

Dada la proyección de PARLANA como herramienta para ser utilizada por toda la comunidad universitaria de la Universidad de Nariño fue aceptada para ser parte de las aplicaciones de la Universidad de Nariño, convirtiéndose en un proyecto institucional que busca dar cumplimiento a algunos de los objetivos del plan de mejoramiento institucional que se encuentra en marcha.

Además, se tomó una capacitación impartida por parte del Centro de Informática de la Universidad de Nariño acerca de Angular, Jasson, NodeJS Express y PostgresSQL para la posterior migración de la aplicación a las tecnologías usadas por parte del Centro de Informática.

REFERENCIAS

[1] *BENAVIDES, Nathalie. Universidad de Nariño se encuentra incluida en el Ranking Mundial de SCImago (SIR WORLD 2018), Pasto, Nariño, Colombia, Junio 21 de 2018. http://www.udenar.edu.co/recursos/wp-content/uploads/2018/06/COMUNICADO-032.pdf*

[2] *IPES, Investigación Cuantitativa, Montevideo, Uruguay, Instituto de Perfeccionamiento y Estudios Superiores "Prof. Juan E. Pivel Devoto", p. 3-4. http://ipes.anep.edu.uy/documentos/investigacion/materiales/inv_cuanti.pdf.*

[3] *CAZAU, Pablo. Introducción a la investigación en Ciencias Sociales (3ª edición), Buenos Aires, Argentina, Marzo de 2006, p. 27 http://alcazaba.unex.es/asg/400758/MATERIALES/INTRODUCCI%C3%93N%20A%20LA%20INVESTIGACI%C3%93N%20EN%20CC.SS.pdf.*

[4] *ADIMEN, Investigación, S.L. Grupo de Análisis e Investigación. Calculadora de Muestras, Bilbao, España, 2018. http://www.adimen.es/calculadora-muestras.aspx*

[5] *Instituto Cervantes, Marco Común Europeo de Referencia para las Lenguas: Aprendizaje, Enseñanza, Evaluación, Madrid, España, 2002, p. 47. https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf*

Biografía Autor(es)

Correspondencia.

Sandra Marleni Vallejo Chamorro . Magister en Gestión de Tecnología Educativa. Docente, Universidad de Nariño. E-mail: savach210@gmail.com. Pasto, Colombia.

Procedimiento para el tratamiento de datos climáticos utilizando la herramienta Pentaho

Ronald Mauricio Gelvez Ruiz (a)
Luz Marina Santos Jaimes (b)

Resumen

En la actualidad la gran cantidad de datos en diferentes áreas crece en forma continua a una alta velocidad, con gran variedad en sus datos y en donde se tiene que conservar la integridad de los mismos para que su valor no se pierda. Aquí herramientas de Big Data juegan un papel fundamental para almacenar, procesar y descubrir información valiosa que contribuya a la solución de diversos problemas. Este trabajo contiene la descripción detallada de las actividades que deben seguirse para llevar a cabo un procedimiento general de tratamiento y procesamiento de grandes cantidades de información aplicado a un conjunto de datos climáticos. El trabajo se desarrolló principalmente en la herramienta ***Pentaho*** como entorno de trabajo central, en la cual se investigó los módulos para tratamiento de datos y las demás funciones que se puedan utilizar. Se tomó como referencia un modelo ETL *(Extract*, *Transform*, *Load)* para la extracción, transformación y carga de datos en bodegas de datos (Data Warehouse) o bases de datos NoSQL con el fin de automatizar y mejorar el proceso de consulta y análisis. Los datos de prueba fueron adquiridos del IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales) para el desarrollo y validación del procedimiento propuesto.

Palabras clave: Análisis de datos, big data, datos meteorológicos, Pentaho, procedimiento.

Procedure for the treatment of climatic data using the Pentaho tool

Abstract

Currently the large amount of data in different areas grows continuously at a high speed, with great variety of data and where the integrity of the same has to be kept so that their value is not lost. Here Big Data tools play a fundamental role in storing, processing and discovering valuable information that contributes to the solution of various problems. This work contains a detailed description of the activities that must be followed to carry out a general procedure for the treatment and processing of large amounts of information applied to a set of climatic data. The work was developed mainly in the Pentaho tool as the central work environment, in which the modules for data processing and other functions were investigated. An ETL (Extract, Transform, Load) model was taken as reference for the extraction, transformation and loading of data into data warehouses or NoSQL databases in order to automate and improve the consultation and analysis process. The test data were acquired from IDEAM (Institute of Hydrology, Meteorology and Environmental Studies) for the development and validation of the proposed procedure.

Keywords: Data analysis, big data, meteorological data, Pentaho, procedure.

I. INTRODUCCIÓN

Las repercusiones generadas por la falta de predicciones correctas y oportunas influyen críticamente en finanzas, seguridad alimentaria y prevención de desastres en nuestra región y el país. Según *Weather Analytics*, una compañía que proporciona datos climáticos, estima que el 33% del PIB mundial se ve afectado por el clima [1]. Este problema se debe a que el estudio para llevar a cabo una predicción es un proceso arduo y que necesita el análisis de una gran cantidad de datos recolectados a través de los años.

Para hacer una correcta predicción del clima o cualquier estudio climático existen grandes volúmenes de datos ambientales, generados por estaciones de monitoreo ambiental, estos datos se utilizan para: describir las variables en el tiempo según la localidad de la toma de datos, conocer el comportamiento de las medidas, descubrir la relación entre los datos y la dinámica de los fenómenos climáticos. Las herramientas que se usan actualmente en la meteorología no son suficientes para analizar la gran cantidad de datos acumulados a través de los años, esto actualmente representa un problema que requiere el uso de herramientas informáticas más avanzadas para su almacenamiento, tratamiento, análisis y extracción de los datos que será relevante para mejorar la predicción del tiempo.

Mediante el uso de Big Data, la información histórica resulta tan valiosa como la información más reciente. Permite mapear distintas tendencias y patrones que se pueden utilizar para predecir mejor qué ocurrirá en un futuro. Y conocer lo que está por venir implica la existencia de soluciones más viables para gestionar potenciales problemas [2]. El trabajo en [3] introduce el concepto de Big Data para la manipulación y análisis de datos climatológicos, utilizando para su representación y análisis sistemas de información geográfica, y para su gestión y control las EIRD (Estrategias Internacionales de Reducción de Desastres).

1.1 Big Data

No existe unanimidad en el concepto de Big Data, aunque si un cierto consenso en la fuerza disruptiva que supone los grandes volúmenes de datos y su necesidad de captura, almacenamiento y análisis. Una primera definición, "Big Data excede el alcance de los entornos de hardware de uso común y herramientas de software para capturar, gestionar y procesar los datos dentro de un tiempo transcurrido tolerable para su población de usuarios" [4].

Entre otras definiciones se encuentra la de la consultora Garner, "Big Data son los grandes conjuntos de datos que tiene tres características principales: volumen (cantidad), velocidad (velocidad de creación y utilización) y variedad (tipos de fuentes de datos no estructurados, tales como la interacción social, video, audio, cualquier cosa que se pueda clasificar en una base de datos)" [5]. Según Doug Laney (considerado padre del modelo de las 3v para Big Data) existen unas características esenciales que encajan en todas las definiciones y él las describió de esta manera en el blog *Gartner Blog Network* [6]. Además de las 3V clásicas con las que todas las fuentes coinciden, aparecen otras dos características: Veracidad y Valor [7].

Expertos coinciden en que lo importante es aprovechar el potencial y descubrir nuevas oportunidades mediante el análisis de grandes volúmenes de datos. La consultora IDC *(International Data Corporation)*, realiza un estudio anual sobre estadísticas referentes

al crecimiento de los datos, dicho estudio llamado "Estudio del Universo Digital" estima que en el 2013 habían 4.4 Zettabytes (ZB) de datos digitales en todo el mundo, y para el 2020 habrán 44 ZB, es decir, desde el año 2013 al 2020 los datos globales crecerán en un factor de 10, por lo que se está duplicando su tamaño cada dos años [8]. Según [9], un artículo escrito en la página de IBM cada día se crean 2,5 quintillones de bytes de datos, de forma que el 90% de los datos del mundo actual se han creado en los últimos dos años.

En Big Data se consideran diferentes tipos de datos según su organización y estructura, estos pueden ser estructurados, semiestructurados y no estructurados. Según, la mayoría de las fuentes de datos tradicionales son datos estructurados, datos con formato o esquema fijo que poseen campos fijos y que conforma las bases de datos relacionales. Los datos semiestructurados tienen un flujo lógico y un formato que puede ser definido, pero no es fácil su comprensión por el usuario. La lectura de datos semiestructurados requiere el uso de reglas complejas que determinan cómo proceder después de la lectura de cada pieza de información [10]. Un ejemplo típico de datos semiestructurados son los registros Web logs de las conexiones a Internet. Los datos no estructurados son datos sin tipos predefinidos, se almacenan como "documentos" u "objetos" sin estructura uniforme, y se tiene poco o ningún control sobre ellos. Ejemplos son: datos de texto, audio, video, fotografías, documentos impresos, cartas, hojas electrónicas, imágenes digitales, formularios especiales, mensajes de correo electrónico y de texto, formatos de texto libre como correos electrónicos, mensajes instantáneos, artículos, libros, entre otros.

1.2 ETL (Extract, Transform and Load)

Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos o Data Warehouse [11]. La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros planos, pero pueden incluir bases de datos no relacionales u otras estructuras diferentes. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

La fase de transformación de un proceso de ETL aplica una serie de reglas o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos. No obstante, en otros casos pueden ser necesarias aplicar operaciones más complejas. En la fase de carga los datos resultantes de la fase de transformación son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. Los Data Warehouse mantienen un historial de los registros de manera que se pueda hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

El artículo [12] plantea un modelo genérico que aplica las etapas de ETL y permite realizar seguimiento del proceso al mantener un registro histórico de errores filtrados y calcular indicadores para identificar la calidad en el procesamiento. La validación del modelo fue realizada sobre un caso de estudio con datos ambientales. El modelo demostró obtener resultados satisfactorios.

1.3 Pentaho Data Integration (PDI)

Pentaho es una herramienta de código abierta de las que se denominan ETL, es decir, una herramienta de extracción de datos de una fuente, transformación de esos datos, y carga de esos datos en otro sitio. PDI admite una amplia gama de formatos de entrada y salida, incluidos archivos de texto, hojas de datos y motores de bases de datos comerciales y gratuitos. Además, las capacidades de transformación de PDI le permiten manipular datos con muy pocas limitaciones. PDI es fácil de usar, cada proceso se crea con una herramienta gráfica donde se especifica qué hacer sin escribir código para indicar cómo hacerlo. La transformación es el elemento básico de diseño de los procesos ETL en Pentaho. Una transformación se compone de pasos, que están enlazados entre sí a través de saltos [13].

Spoon es el entorno gráfico estándar de Pentaho, mediante esta interface gráfica se puede diseñar todos los KTR (archivos generados de Pentaho) basados en una tecnología *Rapid Application Development* (RAD). Las tareas son modeladas tipo *Workflow* ó flujo de trabajo para coordinar recursos, ejecución y dependencias de actividades ETL [14].

II. METODOLOGÍA

Para la ejecución del procedimiento se plantea una metodología en tres ciclos, cada ciclo está asociado a un objetivo caracterizado en una serie de actividades.

2.1 Ciclo 1. Revisión y refinamiento del marco conceptual y preparación de los datos ambientales

En el primer ciclo se estudia las tecnologías asociadas a Big Data para la construcción de modelos de análisis de datos. Se tienen las siguientes actividades:

- Capacitación en conceptos básicos de Big Data mediante un curso online.
- Exploración de las herramientas.
- Instalación de las herramientas a utilizar.
- Elaboración del marco conceptual.
- Recolección de los datos

2.2 Ciclo2. Búsqueda y estudio del modelo ETL para el procesamiento de datos ambientales

El objetivo del Segundo ciclo es examinar un modelo de predicción de comportamiento climático a partir de fuentes de datos ambientales. Se tienen las siguientes actividades:

- Exploración de modelos ETL.
- Selección del modelo.
- Apropiación del modelo.
- Revisión de conceptos básicos acerca del modelo.

2.3 Ciclo3. Aplicación de las fases del modelo y descripción de las actividades realizadas

El ciclo 3 busca desarrollar un procedimiento de aplicación de Big Data en el caso de estudio de análisis de cambio climático. Se tienen las siguientes actividades:

- Elaboración de la fase de prerrequisitos o extracción de los datos.
- Construcción de la fase principal
- Analizar el proceso, mostrar resultados y concluir.

III. RESULTADOS Y DISCUSIÓN

En esta sección se expone el desarrollo de un procedimiento general siguiendo las etapas las etapas ETL, y en segundo lugar el caso de estudio aplicando el procedimiento a un conjunto de datos ambientales.

3.1 Procedimiento general para el tratamiento de datos

El procedimiento general plantea las actividades a desarrollar en las etapas de prerrequisitos, general y de carga; también se visualiza las herramientas de apoyo (Fig. 1).

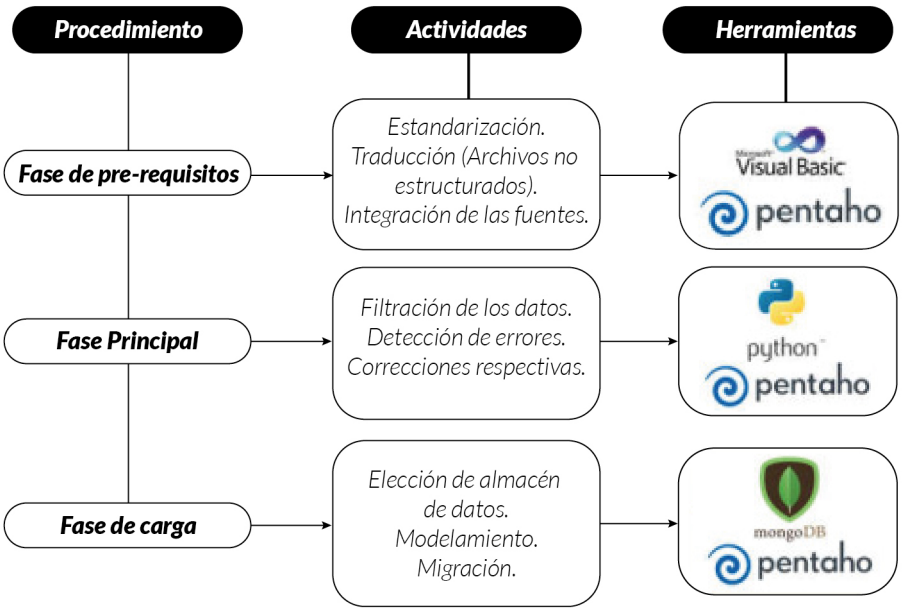


Fig. 1. *Procedimiento general.*

Pre-requisitos. Los datos pueden ser entregados en varias formas de almacenamiento, como por ejemplo archivos planos y repositorios estructurados (bases de datos), los cuales pueden presentar problemas y errores en su adquisición. Por lo anterior es necesario contar con un proceso llamado 'traducción'; este proceso consiste en la toma y estandarización de la estructura de los datos, para luego proceder a la carga de los datos en la herramienta ***Pentaho***. Este proceso es lento debido a que hay operaciones que se hacen de forma manual o utilizando herramientas no optimizadas para manejar tal cantidad de datos como Visual Basic. Los pasos de extracción de ***Pentaho*** incluye una variedad de tipos, siendo una de las más empleadas los archivos .CSV (ver Fig. 2).

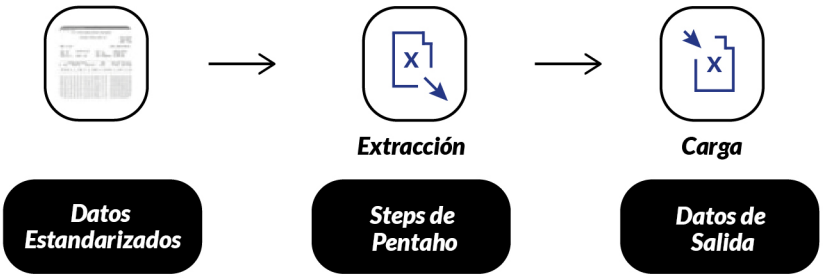


Fig. 2. *Pasos de extracción*

Fase Principal. Consta de detección y corrección de errores, es la fase encargada de recibir los datos organizados en una estructura estándar, construida en la fase de prerrequisitos, y a partir de esto realizar las transformaciones necesarias para que esta información tenga un valor agregado.

Detección de fallas. En esta actividad se detectan los errores presentes en las mediciones de cada una de las variables; es la encargada de detectar los valores atípicos, valores faltantes e inconsistencias en los datos fuente. Muchas veces los datos están incompletos debido a diferentes causas de falla, siendo posible completarlos mediante métodos numéricos, regresiones o algoritmos de inteligencia artificial.

Corrección de fallas: En esta actividad se reciben los datos con los errores detectados y se sigue un conjunto de transformaciones para su corrección (Ver Fig. 3). Los filtros de corrección deben ser investigados en el ámbito del área de aplicación del procedimiento. Por ejemplo, sí se encuentran valores negativos o nulos en alguna variable se debe investigar si esos valores están en el rango permitido, si no, se procede a la corrección.

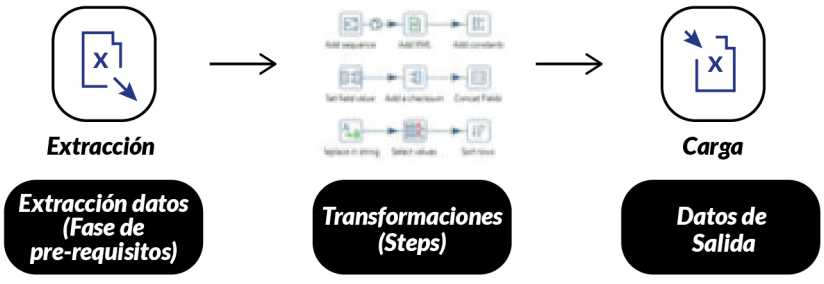


Fig. 3. *Pasos de transformación.*

Las transformaciones incluyen reemplazo de campos, concatenación de campos, entre otros. En caso necesario de completar datos, ***Python*** provee una serie de librerías matemáticas y de inteligencia artificial a utilizar dependiendo de la complejidad de los datos.

Fase de carga al almacén de datos. Esta fase es una de las más importantes del procedimiento, se encarga de llevar los datos a un almacén de datos, éste será diseñado para optimizar el trabajo de consultas de análisis según el área de los datos. Hay dos posibilidades de hacer un almacén de datos para Big Data, la primera es utilizar una base de datos no relacional, estas bases de datos son las mejores al momento de manejar un gran volumen de datos, ya que se tienen respuestas muy rápidas al momento de hacer consultas. Dentro de las bases de datos no relacionales hay varias alternativas que se evalúan según la que más beneficia el análisis. La segunda opción es más compleja y está relacionada a sistemas que no quieren hacer cambios bruscos a su centro de datos, y es usar una base de datos relacional, pero usando alguna de las metodologías para que funcione como Data Warehouse.

La metodología Ralph Kimbal es conocida como arquitectura Data Warehouse de datos dimensional y la arquitectura de Bill Inmon es conocida como fábrica de información corporativa [15]. Estas metodologías se basan en la desnormalización y en la no normalización. La desnormalización es el proceso de procurar optimizar el funcionamiento de una base de datos por medio de agregar datos redundantes, y la no normalización es el modelamiento de una base de datos evitando niveles normales. Esto con la finalidad de obtener modelos estrella o copo de nieve. En [16] presenta detalles y características relevantes de la aplicación de estas metodologías.

3.2 Caso de estudio

Los datos analizados fueron recibidos del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), pertenecen a la estación 16015010 de la ciudad de Cúcuta – Norte de Santander, los datos han sido recolectados desde el año 1989 hasta el 2003. Los datos cuentan con varias variables relacionadas a temperatura, velocidad de vientos y precipitación. Para la realización del procedimiento sólo se utilizó las 24 variables referentes a temperatura como: temperatura máxima, temperatura mínima, amplitud, termómetros secos 07, 13, y 19, termómetro seco media, termómetros húmedos 07, 13 y 19, termómetro húmedo media, humedades relativas 07, 13 y 19, humedad relativa media, tensiones de vapor en milibares 07, 13 y 19, tensión de vapor media, puntos de rocío 07, 13 y 19, punto de rocío medio.

Pre-requisitos. Las fuentes de los datos pueden ser distintas, aquí se cumple con la tercera "gran V" que caracteriza Big Data del resto de tecnologías, esto es la variedad. La herramienta principal usada para esta fase es ***Pentaho Input***, que cuenta con muchas opciones para cargar los datos al entorno de trabajo. El proceso de traducción se encarga de estandarizar los datos en el caso en que se tenga archivos de la forma no estructurada, para este particular caso de estudio es completamente necesario ya que se recibieron archivos planos que contienen tablas, pero mucho contenido sobrante que no serán comprendidos por ninguna base de datos o herramienta. El procedimiento de traducción se hizo en tres pasos:

1. Se utilizó la opción *Search and Replace* para buscar el texto completo del encabezado que tenía cada mes y se reemplazó por filas vacías, cabe resaltar que en el encabezado había un dato importante que era el mes y el año al que pertenecían los datos, este se conservó.

2. Se eliminaron las filas vacías con el paso *(Non Empty)*.

3. Se usó la herramienta macro de Visual Basic para recorrer todo el documento, tomar el año y mes perteneciente al registro y añadírselo a la columna 25. Cabe resaltar que la ejecución del macro tarda un tiempo ya que son muchos registros y se obrescribe uno por uno y la herramienta no está optimizada para tal cantidad de datos (Ver Fig. 3)

```
Sub fechas()
Dim N As Integer
Dim texto As String
Dim valor As String
'Nos situamos en la celda con el primer dato
Range("U1").Select
'Bajamos hasta la última fila adyacente, que contenga datos
Selection.End(xlDown).Select
'Mostramos el número de líneas
N = ActiveCell.Row
respuesta = MsgBox("Esta hola de cálculo tiene " & N & " filas con datos continuos.")
For i = 1 To N
    Cells(i, 21).Select
    texto = ActiveCell.Text
    celda = InStr(" ", texto)
    If celda <> 0 Then
    'Si la celda esta vacia no hacemos nada
    Else
    texto = ActiveCell.Text
    valor = "ENE-FEB-MAR-ABR-MAY-JUN-JUL-AGO-SEP-OCT-NOV-DIC"
    celda = InStr(valor, texto)
    If celda <> 0 Then

    For j = 1 To 31
    Cells(i + j, 25).Value = texto + Cells(i, 22).Value

    Next
    Else

    End If
    End If

Next
End Sub
```

Fig. 4. *Macro en Visual Basic empleada en la traducción.*

En la fase de traducción se hizo una limpieza con el fin de cargar los datos por medio del paso *"Microsoft Excel Input"* al entorno de trabajo de ***Pentaho***, una vez ahí se tienen varias opciones de cargarlo según sea el caso:

1. Cargar mes por mes: Esta opción es viable y muy efectiva cuando el propósito de uso de los datos es el análisis mensual, pero en un rango corto de años, ya que se necesitan conocer las coordenadas de cada tabla mensual para poderla montar.

2. Cargar todos los datos en un paso: Esta opción es la más recomendada cuando el análisis se va a hacer en todo el rango de años que se tienen como es el caso de estudio que tiene 14 años de muestras (Ver Fig. 5).

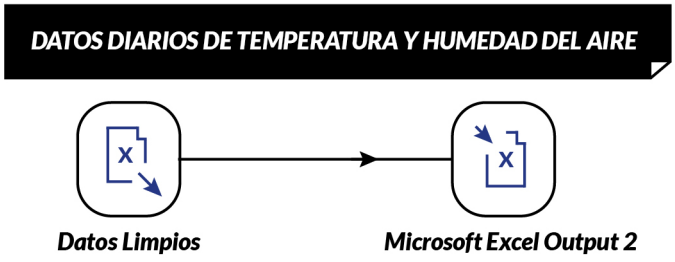


Fig. 5. *Extracción general.*

Fase principal. En este caso de estudio se tiene una pérdida de datos desde mayo hasta octubre de 1989, también se pueden conseguir datos faltantes en pequeñas cantidades o individuales. Para suplir los datos faltantes de 1989 se entrenó una red neuronal del tipo perceptron multicapa utilizando como datos de entrenamientos los años 1990 y 1991, otros trabajos como referentes empleando redes neuronales [17, 18]. A continuación, se detallan los pasos utilizados en la herramienta ***Pentaho***, en el *workflow* 1 se busca solución a los errores encontrados, en los *workflow* 2 y 3 se extraen los archivos de entrenamiento para la red neuronal que busca completar los datos faltantes (Ver Fig. 6).

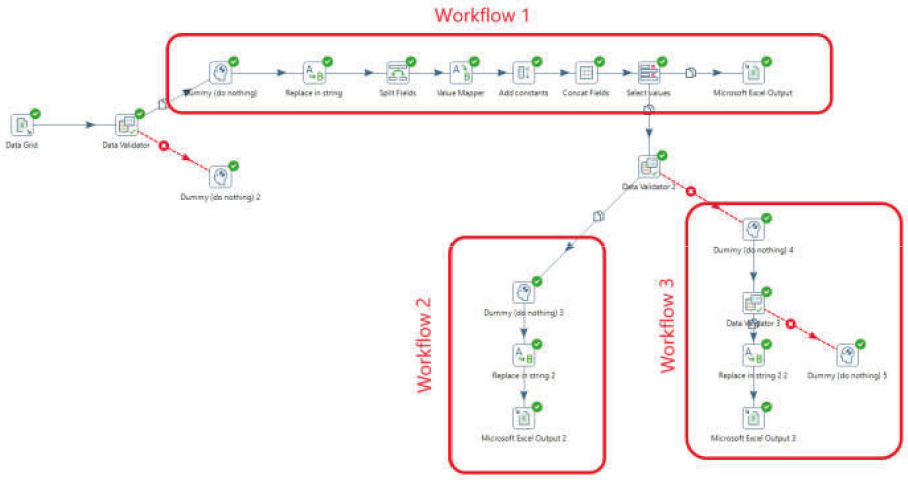


Fig. 6 *Workflows en Pentaho de la Fase Principal.*

WorkFlow 1

1. *Microsoft Excel Input:* Se cargan los datos desde el archivo de salida que resultó del proceso de la traducción.

2. *Data Validator:* Se usa para poner un rango de 1-31 en la columna de días y así eliminar todos aquellos registros que no pertenecen a un mes

3. *Dummy (1):* Recibe los datos validados del paso anterior.

4. *Dummy (2):* Recibe los datos no validados del paso anterior.

5. *Replace In String (1):* Se busca (19 y 20) en la columna de la fecha que se tiene y se reemplaza por (19- y 20-) esto con el fin de tener un delimitador.

6. *Split-Fields:* Se dividen campos con el delimitador puesto anteriormente (-) y se forman dos nuevas columnas mes y año.

7. *Value Mapper:* Con este paso se le da el formato al mes ya que presenta la forma: ENE, FEB, MAR, ABR, etc. Después del paso quedará de la siguiente manera: 01, 02, 03, 04, etc.

8. *Select Values:* Se selecciona las columnas que se quieren en el documento de salida y se agregan los encabezados correspondientes.

9. *Microsoft Excel Output:* Se le da la salida a la fase principal del procedimiento.

Workflow 2

1. *Data validator (2):* Se usa para validar que los registros sean del año 1990 (un año después de la perdida de datos)

2. *Dummy (3):* Recibe los datos validados del paso 1.

3. *Dummy (4):* Recibe los datos no validados del paso 1.

4. *Replace in String (2):* Si se pasó por el paso 2, se busca 1990 en la columna fecha y se reemplaza por 1989

5. *Microsoft Excel Output:* Se cargan los datos en un archivo Excel de nombre "entrenamiento1".

Workflow 3

1. *Data validator (2):* Se usa para validar que los registros sean del año 1991 (dos años después de la perdida de datos)

2. *Dummy (5):* Recibe los datos no validados del paso anterior.

3. *Replace in String:* Si pasó por el paso 1, se busca 1991 en la columna fecha y se reemplaza por 1989

4. *Microsoft Excel Output:* Se cargan los datos en un archivo Excel de nombre "entrenamiento2"

Para completar los datos faltantes se realizaron los siguientes pasos:

1. Se creó un *dataframe* con pandas para la manipulación de los datos de entrenamiento y la tabla objetivo de predicción.

2. Se entrenó una red neuronal de tipo perceptron multicapa de la librería sknn de python, para este caso se usó 800 neuronas, una tasa de aprendizaje de 0,00001 y 8000 iteraciones. Los datos predichos estaban en el rango según el sitio Weather Spark, esto es de 22 °C a 33 °C [19].

3. Se creó un arreglo con los datos resultantes de la predicción para completar los datos faltantes del año 1989.

4. Los datos predichos se pasan a un documento en Excel y se agregan los días del año a los cuales pertenecen. Luego se verifica la región de datos faltantes en el archivo del IDEAM y se pegan los datos de la red neuronal en las fechas correspondientes.

5. Para llenar los datos individuales, se promedió los valores cercanos antes y después de la perdida, ya sea horas antes y horas después como también valores a la misma hora, pero en días cercanos a la falla. Se elaboró un macro en Visual Basic que buscaba los registros nulos, y llenaba dichos registros con el promedio de los registros vecinos.

Fase de carga al almacén de datos. Esta es la última fase del procedimiento, en ella se hará la migración de los datos a un entorno en donde se puedan consultarlos a una mayor velocidad. Esta se realiza hacia MongoDb desde un archivo de Excel (Ver Fig. 7). Se extrae el archivo de salida de la fase anterior y se relaciona con el paso "MongoDB Output".

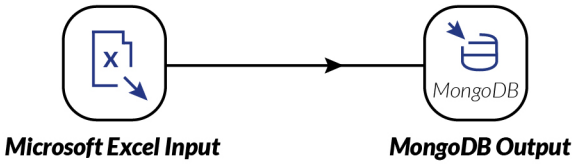


Fig. 7. *Carga desde Excel hacia Mongo.*

La ventana de MongoDB Output presenta las opciones para configuración de la conexión, salida, documento e índices (Ver Fig. 8). En la pestaña de configuración de conexión de Mongo se dejan los valores por defecto (hostname y puerto), a menos de que se tenga un usuario y contraseña esTable.cidos. En opciones de salida aparecerán las opciones de la carga hacia la base de datos (Database y Collection), cabe resaltar que si no hay una base de datos creada con el nombre registrado, está se creará, igual sucede con la colección.

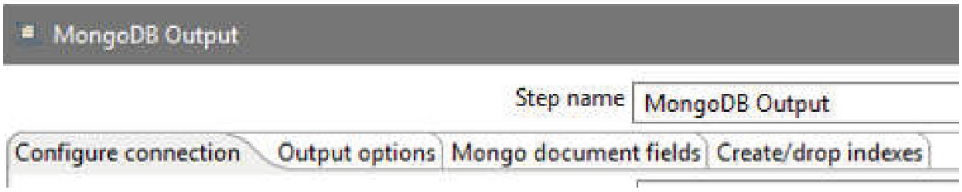


Fig. 8. *Opciones de configuración de Mongo.*

En la pestaña *"Mongo document fields"* aparece la opción en el pie de la ventana *"Get fields"* está hará que se trasladen las cabeceras de la tabla para hacer los campos en el documento de mongo. Por último, se mira la estructura del documento con la opción *"Preview document structure"* (Ver Fig. 9).

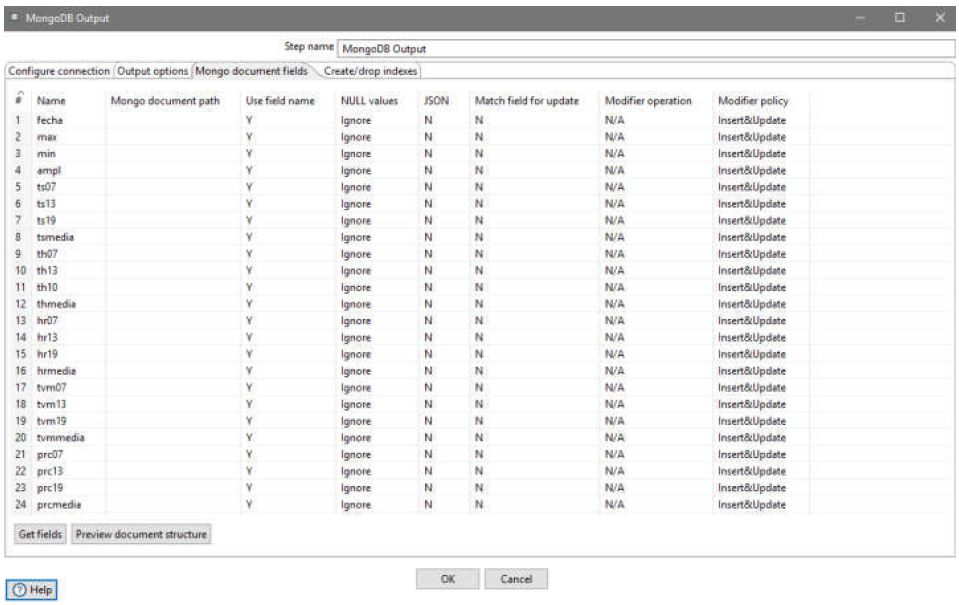


Fig. 9. *Carga de los encabezados a documento Mongo.*

Al correr la transformación se tiene que verificar que efectivamente los datos si se cargaron hacia mongo. Para esto se abre símbolo del sistema y se escribe "mongo", luego se escribe *"show dbs"* para consultar las bases de datos (Ver Fig. 10).

Fig. 10. *Bases de datos Mongo consultada desde terminal.*

Una vez comprobada que existe la data, se escribe el comando "use xxxxxxx" reemplazando las "x" por el nombre de la base de datos, en este caso "salidaproyecto". Hay que tener en cuenta que el comando "use" puede crear una base de datos, así que, si no se escribe bien el nombre, se crea una data con el nombre erróneamente escrito. Luego, de esto se tiene que consultar las colecciones con el comando "show collections" (Ver Fig. 11).

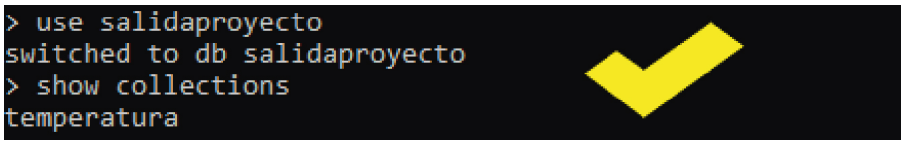


Fig. 11. *Colección consultada desde terminal.*

En este caso aparece la colección "temperatura" que es el nombre que se le dio en el entorno de ***Pentaho***, aquí ya se pueden consultar los datos que están en esa colección con el comando "db.xxxxxxx.find().pretty()" reemplazando las "x" por el nombre de la colección, el agregado "pretty()" es para que la información se vea de una manera más organizada en el símbolo del Sistema (Ver Fig. 12).

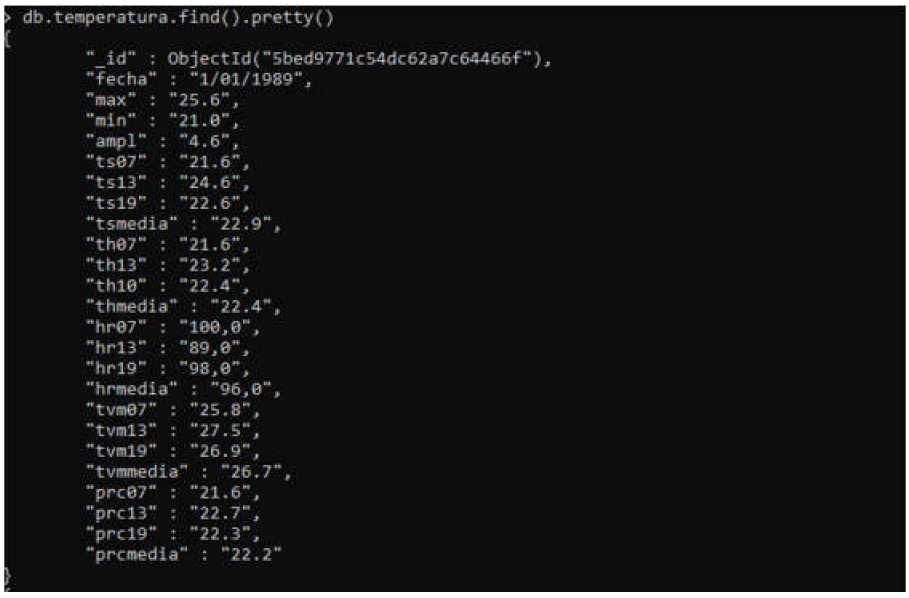


Fig. 12. *Consulta de datos en Mongo.*

IV. CONCLUSIONES

El procedimiento que se formalizó en el artículo demuestra que la tecnología Big Data es un campo de estudio en donde prima el manejo de grandes cantidades de datos priorizando la integridad y la calidad de la información. Debido a los problemas generados en la meteorología en el último siglo por el crecimiento de datos queda claro que se necesitan mejorar las tecnologías de recolección y tratamiento de los datos. Herramientas ETL como ***Pentaho*** permiten manipular grandes cantidades de información y llevarlas mediante diferentes transformaciones a un formato estándar requerido para su fácil análisis y consulta. El apoyo de mecanismos como las redes neuronales ayuda a completar datos faltantes y mejorar la calidad de los datos.

REFERENCIAS

[1] *Fernández, C. Cómo el Big Data puede ayudar a luchar contra el cambio climático. Business Insider. Homepage, https://www.businessinsider.es/como-big-data-puede-ayudar-lucharcambio-climatico-183600, accesado 2018/12/06.*

[2] *Nathan Sykes. Big Data y la lucha contra el cambio climático. Homepage, https://www.bbvaopenmind.com/big-data-y-la-lucha-contra-el-cambio-climatico/, accesado 2018/12/06.*

[3] *Fernando, R. Big Data en el comportamiento de datos climatológicos y estrategias internacionales de reducción de desastres para la gestión de riesgo ambiental. Universidad Mayor de San Andrés. Homepage, https://repositorio.umsa.bo/handle/123456789/10689, accesado 2018/12/06.*

[4] *Merv, A. Big Data. Homepage, http://www.nxtbook.com/nxtbooks/mspcomm/teradata_2011q1/index.php?startid=8#/40, accesado 2018/12/06.*

[5] *Hung LeHong. Reporte de Gartner analiza "big data" alrededor de tecnología de datos. Homepage, https://searchdatacenter.techtarget.com/es/noticias/2240171952/Reporte-deGartner-analiza-big-data-alrededor-de-tecnologia-de-datos, accesado 2018/12/06.*

[6] *Laney, D. Deja VVVu: Others Claiming Gartner's Construct for Big Data, https://blogs.gartner.com. Homepage, https://blogs.gartner.com/doug-laney/deja-vvvueothers-claiming-gartners-volume-velocity-variety-construct-for-big-data/, accesado 2018/12/06.*

[7] *Prometeus Global Solutions. Las cinco dimensiones del Big Data. Homepage, https://prometeusgs.com/volumen-variedad-velocidad-veracidad-y-valor-las-5-dimensiones-del-big-data-la/, accesado 2019/02/20.*

[8] *Mesa, A. R. BIG DATA: La evolución de los datos. openwebinars.net. Homepage, https://openwebinars.net/blog/big-data-la-evolucion-de-los-datos/?cat=big-data,*

[9] *Fragoso, R. B. ¿Qué es Big Data? IBM. Homepage, https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html, accesado 2018/12/06.*

[10] *Aguilar, L. J. Big Data - Analisis de grandes volumenes de datos en organizaciones. 1ra edn. Alfaomega Grupo Editor, México. (2013).*

[11] *Pinto, J. M. C. (2017). ¿Que es una ETL?. Homepage, https://www.linkedin.com/pulse/quees-una-etl-juan-manuel-castillo-pinto, accesado 2018/12/06.*

[12] *Duque Méndez, N. D., Hernández Leal, E. J., Pérez Zapata, Á. M., Arroyave Tabares, A. F., & Espinosa Gómez, D. A. Modelo para el proceso de extracción, transformación y carga en bodegas de datos. Una aplicación con datos ambientales. Ciencia e Ingeniería Neogranadina, 26(2), 95–109 (2016). https://doi.org/10.18359/rcin.1799*

[13] *Espinosa, R. Construccion procesos ETL utilizando Kettle (Pentaho Data Integration). El Rincon del BI. Homepage, https://churriwifi.wordpress.com/2010/05/10/16-3-construccionprocesos-etl-utilizando-kettle-pentaho-data-integration/, accesado 2018/12/06.*

[14] *Leninmhs. Manual para la Herramienta PDI – KETTLE de Pentaho. Homepage, https://tubasededatoslibre.org.ve/manual-pdi-kettle-pentaho/, accesado 2018/12/06.*

[15] *ZenTut. Kimball vs. Inmon Data Warehouse Architectures. Homepage, http://www.zentut.com/data-warehouse/kimball-and-inmon-data-warehouse-architectures/, accesado 2019/02/20.*

[16] *Roberto Espinosa. Kimball vs Inmon. Ampliación de conceptos del Modelado Dimensional. Homepage, https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-delmodelado-dimensional/?fbclid=IwAR0aiuPac6Z709UYa6YcJNDgsEHHStiWs7zeBfrZH2mRATZgFaCb3QOFRc, accesado 2018/12/06.*

[17] *Moreno J, Serrano J and Corredor J. Forecasting meteorology models implemented by the feed-forward architecture of neural networks. The 2013 World Congress in Computer Science Computer Engineering and Applied Computing. (2013)*

[18] *Montoya, S. Cómo rellenar datos faltantes de precipitacion con redes neuronales en python. http://gidahatari.com/ih-es/como-rellenar-datos-faltantes-de-precipitacion-de-con-redesneuronales-en-python, accesado 2018/12/06.*

Biografía Autor(es)

a. Ingeniero de Sistemas. Docente, Universidad de Pamplona, Facultad de Ingeniería y Arquitectura. E-mail: ronaldwar88@gmail.com. Pamplona, Colombia.

b. Ingeniera de Sistemas. Docente, Universidad de Pamplona, Facultad de Ingeniería y Arquitectura. E-mail: lsantos@unipamplona.edu.co. Pamplona, Colombia.

Propuesta de un conjunto de Heurísticas para el diseño de portales Web para niños

Mallory Julieth Muñoz Gutiérrez (a)
Liceth Marcela Domínguez Pinzón (b)
Daniela Pérez Oliveros (c)
Gabriel Elías Chanchí Golondrino (d)
Katerine Márceles Villalba (e)

Resumen

El concepto de usabilidad ha ganado fuerza en los últimos años con el aumento en el número de aplicaciones desplegadas en la nube y en el número de usuarios que acceden a ellas. En este sentido, la usabilidad se ha convertido en un atributo fundamental para el desarrollo de aplicaciones interactivas y competitivas dentro de la industria del software, lo que permite que los sitios web cumplan con el propósito para el que fueron creados, además toma en cuenta la población a la que se dirigen. Sin embargo, a pesar de los aportes de la usabilidad en el desarrollo de aplicaciones, se encuentra que los usuarios en edad infantil no han sido considerados como actores principales en el uso de los recursos tecnológicos, provocando que su interacción con ellos sea compleja y deficiente. En respuesta a la importancia de los sitios web para niños y la falta de pautas específicas para el diseño de estos portales, el propósito de este artículo es proponer un conjunto de usabilidades heurísticas para los sitios web infantiles, así como una herramienta para la inspección de las propuestas heurísticas mediante la evaluación de criterios asociados a cada uno de ellos. Estas contribuciones están destinadas a beneficiar a desarrolladores, diseñadores y otros actores involucrados en la creación de sitios web para niños.

Palabras clave: Heuristica, interaccion, sitios Web de usabilidad para niños.

Proposal for a set of Heuristics for the design of Web portals for children

Abstract

The concept of usability has gained strength in the last years with the increase in the number of applications deployed in the cloud and in the number of users that access them. In this sense, the usability has become a fundamental attribute for the development of interactive and competitive applications within the software industry, which allows websites to fulfill the purpose for which they were created, in addition it takes into account the population they are targeting. However, despite the contributions of usability in the development of applications, it is found that the users of child-age have not been considered as major players in the use of technological resources, causing their interaction with them to be complex and deficient. In response to the importance of websites for children and the lack of specific guidelines for the design of these portals, the purpose of this article is to proposes a set of seven heuristic usabilities for children's websites, as well as a tool for inspection of the proposed heuristics by the evaluation of criteria associated with each of them. These contributions are intended to benefit developers, designers and other actors involved in the creation of websites for children.

Keywords: Heuristics, interaction, usability Websites for children.

I. INTRODUCCIÓN

Actualmente está demostrado que la aplicación de nuevas tecnologías y la incorporación de un sitio web, está correlacionado de manera positiva con el incremento de las ventas, la productividad y el valor del mercado de las empresas ya que el uso de este tipo de portales son la forma más sencilla que tienen las personas para dar a conocer información sobre un servicio [1]. En relación a esto es importante que los sitios web cumplan con los requisitos para lo cual fueron creados y sean acordes al tipo de población a la cual se dirigen, de tal modo que estos dos aspectos logren captar la atención del usuario sin dejar de un lado la idea principal para la cual se creó el sitio web. Es por esto, que el concepto de usabilidad ha tomado fuerza con el objetivo de que los usuarios puedan cumplir sus expectativas de uso de manera adecuada [2].

Dentro de las ventajas y aportes que brinda la usabilidad al campo de desarrollo del software se encuentran: es un atributo fundamental en el desarrollo de las aplicaciones interactivas, ya que permite tener un desarrollo competitivo dentro de la industria del software al considerar a las personas como parte fundamental del éxito o fracaso de un producto software [3]. De este modo, el concepto de usabilidad agrupa aquellos aspectos que deberían tenerse en cuenta a la hora de crear un aplicativo software que cumpla con las finalidades de su uso al momento de ser utilizado por un usuario [4]. A pesar de las grandes ventajas que provee la usabilidad en el desarrollo de aplicaciones, se ha podido detectar la poca relevancia que se les ha dado a usuarios en edad infantil como actores principales en el uso de recursos tecnológicos, razón por la cual muchos niños abandonan la mayoría de sitios web debido a que la interacción con estos se torna demasiado compleja a la hora de navegar [5].

Al observar la importancia que tienen los sitios web para niños, la poca relevancia que se le da al diseño de estos y la falta de un conjunto puntual de heurísticas que permitan guiar el diseño y construcción de páginas web dirigidas a este tipo de población, en este artículo se presenta como aporte un conjunto de heurísticas de usabilidad para sitios web dirigidos a la población infantil, a partir de las cuales se pretende beneficiar a los desarrolladores web, diseñadores, tecnólogos y a todas las personas que hagan parte de esta áreas, para que mediante estas recomendaciones puedan tener una pauta que les permita guiar el proceso de diseño de sitios web para niños. A partir de las heurísticas propuestas se propone también una herramienta para la inspección de usabilidad para aplicaciones web, la cual fue desarrollada en ellenguaje Java. Esta herramienta permite evaluar el cumplimiento de un conjunto de criterios asociados a las siete heurísticas del test (Estructura y navegación, diseño de la página, facilidad en la interacción, control y retroalimentación, elementos multimedia, ayuda y búsqueda) obteniendo un reporte gráfico sobre el porcentaje de cumplimiento y el promedio de valoración obtenido por cada principio. La herramienta pretende ser de apoyo para conducir inspecciones de usabilidad sobre portales web para niños, de tal modo que puedan identificar diferentes elementos a mejorar de usabilidad en un portal web.

El resto del artículo está organizado de la siguiente forma: en la sección 2 se presentan las fases de la metodología empleada para la presente investigación, en la sección 3 se presentan un conjunto de conceptos que se tuvieron en cuenta para el desarrollo de este trabajo, en la sección 4 se describen las heurísticas propuestas para portales web dirigidos a la población infantil, en la sección 5 se presenta la estructura funcional de la herramienta de inspección, en la sección 6 se presenta un caso de estudio realizado sobre el portal web de la Secretaría de Educación del Cauca y finalmente en la sección 7 se presentan las conclusiones y trabajos futuros.

II. METODOLOGÍA

Para la obtención de las heurísticas de usabilidad dirigidas a portales web infantiles y el desarrollo y la validación de la herramienta para la inspección de usabilidad, se generaron las siguientes fases a saber: obtención de heurísticas de usabilidad, definición de los criterios asociados a las heurísticas, construcción de la herramienta de inspección de usabilidad y por último caso de estudio (ver figura 1).

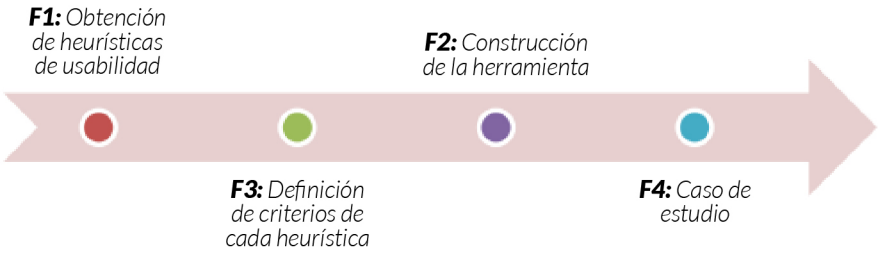


Fig. 1. *Metodología propuesta.*

Fase 1 – Obtención de heurísticas de usabilidad: A partir de la revisión de diferentes tipos de principios y recomendaciones de usabilidad para aplicaciones de escritorio y de propósito general, se obtuvieron un conjunto de heurísticas para guiar el proceso de diseño y construcción de aplicaciones web enfocadas a la población infantil.

Fase 2 – Definición de criterios de cada heurística: Partiendo de las heurísticas definidas en la fase 1, se diseñaron un conjunto de criterios en términos de preguntas para cada heurística. Estos criterios permiten verificar la medida en la que un portal web es adecuado para la población infantil desde la perspectiva de la usabilidad.

Fase 3 - Construcción de la herramienta: Partiendo de los criterios obtenidos en la fase 2, se procedió a diseñar y construir una herramienta para la inspección de usabilidad, la cual está organizada por pestañas, donde cada pestaña corresponde a una de las heurísticas propuestas. Así mismo, se incluyen unas pestañas adicionales para el análisis de los resultados.

Fase 4: Caso de estudio: Una vez generada la herramienta para evaluar los sitios web dirigidos a la población infantil, se procedió escoger el portal web al cual aplicar el estudio, tomando como elección el portal de la Secretaría de Educación del Cauca, puesto que al ser un portal web de tipo gubernamental, requiere garantizar las condiciones necesarias debido a la cantidad de visitantes ya que este portal cuenta con el portal para niños.

III. MARCO CONCEPTUAL

En esta sección se incluyen un conjunto de conceptos relevantes que se tuvieron en cuenta para el desarrollo del presente trabajo. Dentro de los cuales se encuentran: heurística, método de inspección y usabilidad.

3.1 Heurística

La palabra heurística procede etimológicamente de la palabra griega "euriskien" que procede de "eureka", un vocablo que significa hallar o encontrar. El diccionario de la real academia española define a la palabra heurística como:

- Técnica de la indagación y descubrimiento.
- Búsqueda o investigación de documentos en la historia.
- En algunas ciencias, manera de buscar la solución de un problema mediante métodos no rigurosos, como por tanteo, reglas empíricas, etc. [6].

3.2 Método de inspección

La Inspección de usabilidad es el nombre genérico para un conjunto de formas eficaces de evaluación de las interfaces de usuario para encontrar problemas de usabilidad, son muy informales y fáciles de usar. Este método consiste en la conformación de un grupo de expertos en usabilidad que analizan o inspeccionan la aplicación considerada. Estos realizan un informe comentando sobre distintos aspectos de usabilidad de la aplicación, basándose en su experiencia en el área y teniendo en cuenta un conjunto de principios previamente definidos. Este informe es utilizado para realizar los cambios o ajustes necesarios en la aplicación, para resolver los problemas indicados. Dos de los métodos más utilizados de este tipo son: evaluación heurística y recorrido cognitivo [7]. En el presente artículo se hizo una inspección de usabilidad sobre el portal de la Secretaria de Educación del Cauca, teniendo en cuenta las heurísticas propuestas en la sección 4.

3.3 Usabilidad

La usabilidad es entendida según la ISO 9241-11 como la medida en que un sistema, producto o servicio puede ser utilizado por usuarios específicos para lograr objetivos específicos con efectividad, eficiencia y satisfacción en un contexto específico de uso, del mismo modo, puede ser definida como la disciplina que estudia la forma de diseñar sitios web para que los usuarios puedan interactuar con ellos de la forma más fácil, cómoda e intuitiva posible [8].

IV. HEURÍSTICAS DE USABILIDAD

En la tabla 1 se presentan el conjunto de heurísticas de usabilidad propuestas para el diseño y construcción de sitios web dirigidos a la población infantil, los cuales se obtuvieron a partir de diferentes fuentes tales como [9] y [10]. Para diseñar estas heurísticas se consideró un rango de edad entre 3 y 12 años.

Table. 1. *Heurísticas Torres Burriel*

Heurística	Descripción
H1: *Estructura y navegación*	*Lo criterios de estructura y navegación hacen referencia a la organización, estructura y navegación adecuadas del sitio web para niños.*
H2: *Diseño de la página*	*Los aspectos de diseño de la página están relacionados con el contenido de la interfaz, tales como distribución espacial, colores, contenidos y elementos de navegación pertinentes para un entorno infantil.*
H3: *Facilidad de interacción*	*Se refiere a los criterios que posibilitan la comunicación con los usuarios de la página para niños; por ejemplo, el uso de lenguaje claro, concreto, sencillo, sin acudir a tecnicismos.*
H4: *Control y retroalimentación*	*Son aspectos que reflejan el control del usuario en la interfaz, para realizar acciones de navegación y deshacerlas libremente. Además, implican información constante del estado del sitio web para niños mediante información de errores, tiempo de respuesta y ventanas de notificación.*
H5: *Elementos multimedia*	*Los elementos multimedia están relacionados con el uso pertinente de recursos visuales y auditivos para la página web infantil.*
H6: *Ayuda*	*Aspectos relacionados con facilidad de acceso y efectividad del sistema para encontrar secciones o contenidos de manera sencilla y oportuna dentro del sitio web infantil.*
H7: *Búsqueda*	*Aspectos relacionados con facilidad de acceso y efectividad del sistema para encontrar secciones o contenidos de manera sencilla y oportuna dentro del sitio web infantil.*

En concordancia con lo anterior, en la tabla 2 se presentan la cantidad de criterios que está asociado a cada una de las heurísticas presentadas en la tabla 1. Tal como se mencionó previamente dichos criterios son presentados al evaluador en términos de preguntas.

Table. 2. *Criterios por heurística del test heurístico de Torres-Burriel*

Heurística	Número de criterios
H1: *Estructura y navegación*	*4*
H2: *Diseño de la página*	*15*
H3: *Facilidad de interacción*	*19*
H4: *Control y retroalimentación*	*11*
H5: *Elementos multimedia*	*7*
H6: *Ayuda*	*5*
H7: *Búsqueda*	*4*

A modo de ejemplo, a continuación en la tabla 3 se presentan dos criterios por cada una de las heurísticas propuestas en este artículo.

Table. 3. *Ejemplo de criterios definidos*

Heurística	***Descripción***
H1: *Estructura y navegación*	*¿Los elementos utilizados dentro del portal web para niños tienen un tamaño grande y son llamativos para el usuario?* *¿En qué medida se hace uso de zonas de alta jerarquía dentro del portal web para niños?*
H2: *Diseño de la página*	*¿Qué tan adecuado es el tamaño y el tipo de fuente en el portal web para niños?* *¿El tamaño de la fuente es lo suficientemente grande como para no dificultar la correcta legibilidad del texto?*
H3: *Facilidad de interacción*	*¿El portal web permite crear una conexión entre los niños y el sitio?* *¿En qué medida el portal ha tenido en cuenta en el diseño pautas básicas de usabilidad (Nielsen, Schneiderman)?*
H4: *Control y retroalimentación*	*¿Existe un glosario de términos dentro del portal web para niños?* *¿Existe algún mecanismo dentro del portal que contribuya a entender términos complejos?*
H5: *Elementos multimedia*	*¿El portal web para niños incorpora música, animaciones, videos y personajes?* *¿En qué medida los contenidos del portal web para niños tienen una alternativa de audio al contenido multimedia?*
H6: *Ayuda*	*¿El portal web ofrece una sección de ayuda al usuario?* *¿La sección de ayuda maneja un lenguaje claro y conciso?*
H7: *Búsqueda*	*¿En qué medida las páginas de resultados de búsqueda en el portal web para niños incluyen sugerencias claras?* *¿En qué grado el portal web para niños permite el uso de términos familiares en la consulta, incluyendo errores de ortografía y/o digitación?*

V. HERRAMIENTA PROPUESTA

La herramienta propuesta en este trabajo partió de un conjunto de criterios de usabilidad asociados a cada una de las 7 heurísticas descritas en la sección 4. Estos criterios son presentados en términos de preguntas y pueden ser evaluados en un rango de 0 a 5, siendo 0 la ausencia del criterio en el portal web y 5 el máximo cumplimiento del criterio dentro del portal. Del mismo modo, en caso que el criterio no aplique para el portal web en cuestión, la herramienta brinda la posibilidad al evaluador de escoger la opción (N/A).A nivel de diseño, la herramienta está basada en pestañas, cada una de las cuales comprende una heurística y sus criterios asociados.

De acuerdo a lo anterior, la herramienta propuesta está constituida por diez pestañas organizadas de la siguiente forma: de la pestaña uno a la pestaña siete se incluyen cada una de las heurísticas de usabilidad y sus criterios asociados, en la pestaña ocho se presenta una tabla con el promedio obtenido en cada una de las heurísticas evaluadas y el promedio general de las heurísticas, en las pestaña nueve se incluye una gráfica con los promedios de cada principio, y finalmente en la pestaña diez se presenta una gráfica con el porcentaje de cumplimiento de cada una de las heurísticas de usabilidad para el diseño y construcción de sitios web dirigidos a la población infantil.

Un diagrama de flujo que ilustra de manera más clara el funcionamiento de la herramienta se presenta en la figura 2. Tal como se puede apreciar en la figura 2, la herramienta de permite al usuario evaluar uno a uno los diferentes criterios asociados a los principios de usabilidad propuesta, para posteriormente generar estadísticas del valor promedio por

heurística de usabilidad y del porcentaje de cumplimiento de cada heurística dentro del portal web evaluado. En resumen, de acuerdo a la figura 2 es posible apreciar cómo las funcionalidad de las pestañas se pueden agrupar en dos: en primera instancia la valoración de los criterios de las once heurísticas (pestaña uno a la pestaña siete) y en segunda instancia lo relacionado con el análisis de los resultados estadísticos y gráficos de la evaluación realizada. Mediante los resultados obtenidos se pretende guiar al evaluador sobre aquellas heurísticas que el portal web evaluado no está cumpliendo a cabalidad.

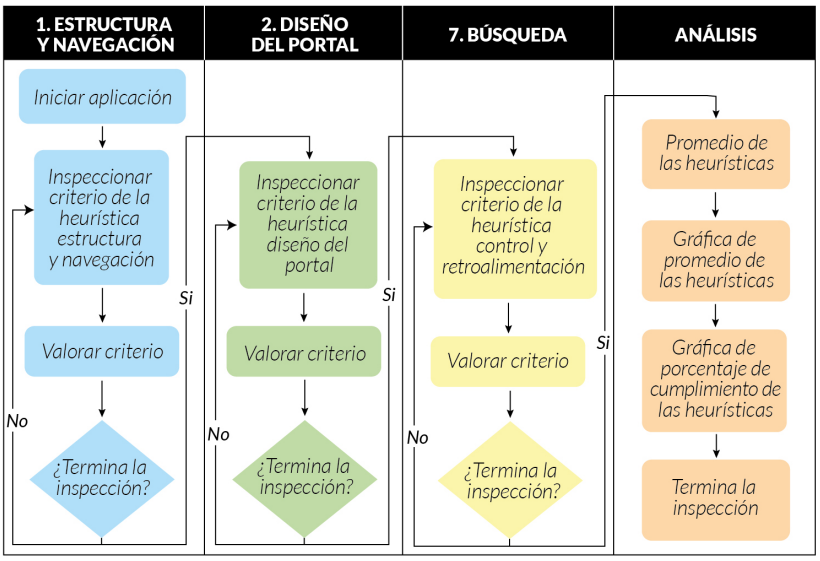


Fig. 2. *Diagrama funcional de la herramienta propuesta*

Para dar cumplimiento al proceso presentado en la figura 2, se construyó en el lenguaje Java una herramienta por pestañas y con diferentes tablas por pestaña, haciendo uso de los componentes JTabbedPane (clase Swing) y JTable. (ver figura 3). En cuanto a la generación de las gráficas de análisis la herramienta hace uso de la librería JFreeChart [11] la cual en este caso permite la obtención de gráficas de barras sobre el promedio de las valoraciones por heurística y el porcentaje de cumplimiento de las mismas en el portal web a evaluar.

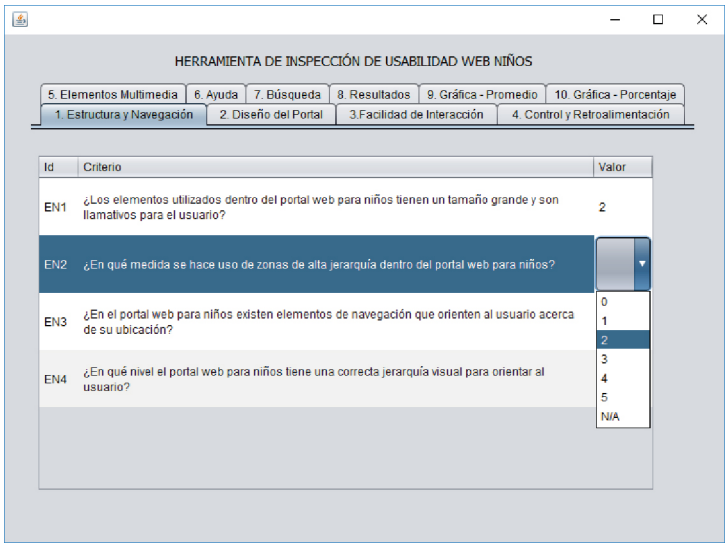


Fig. 3. *Pestañas de evaluación de heurísticas*

Tal como se aprecia en la figura 3, cada una de las siete primeras pestañas tiene asociada una tabla con tres columnas: Id, Criterio y Valor. La columna "Id" permite distinguir los diferentes criterios de cada heurística, por lo cual se usa la letra inicial de la heurística y un número, en ese orden de ideas el primer criterio de la heurística "Estructura y Navegación" tiene por código EN1. En lo referente a la columna "Criterio", se incluye una pregunta que permite al evaluador verificar el cumplimiento del criterio de una determinada heurística. Finalmente la columna "Valor" comprende un conjunto de valores que el evaluador puede asignar a cada criterio, dependiendo el nivel de cumplimiento del mismo en el portal web. La escala de valores varía entre 0 y 5, incluyendo la opción N/A, la cual puede ser usada en caso que el criterio no aplique para el portal web en cuestión. El valor 0 en la escala corresponde a la no aplicación del criterio en el portal web, mientras que 5 corresponde a la aplicación completa del criterio en el portal web.

Una vez terminada la inspección por las diferentes heurísticas y criterios de usabilidad presentados en la sección 4, es posible visualizar en la pestaña "Resultados" el promedio obtenido en cada una de las heurísticas y el promedio general obtenido en la inspección de usabilidad (ver figura 4). El promedio obtenido dentro de cada una de las heurísticas permite al evaluador identificar los aspectos de la aplicación web que no cumplen de manera adecuada los criterios de usabilidad.

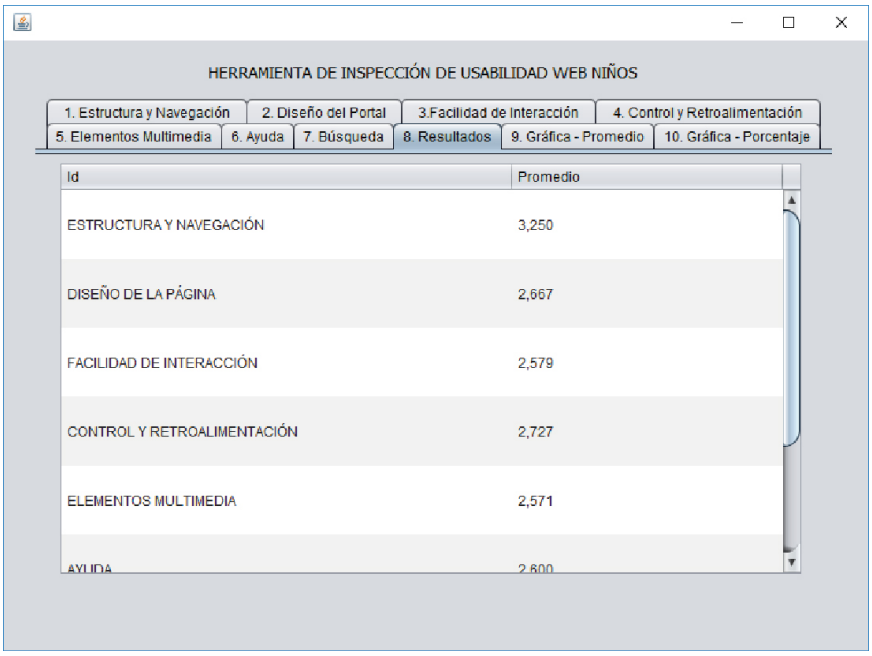


Fig. 4. *Pestaña de resultados de heurísticas*

Una vez generados los promedios de las heurísticas y el promedio general de la evaluación, en las pestañas "Gráfica – Promedio" es posible visualizar gráficamente el promedio obtenido por cada una de las heurísticas de usabilidad dirigidas a portales web infantiles (ver figura 5).

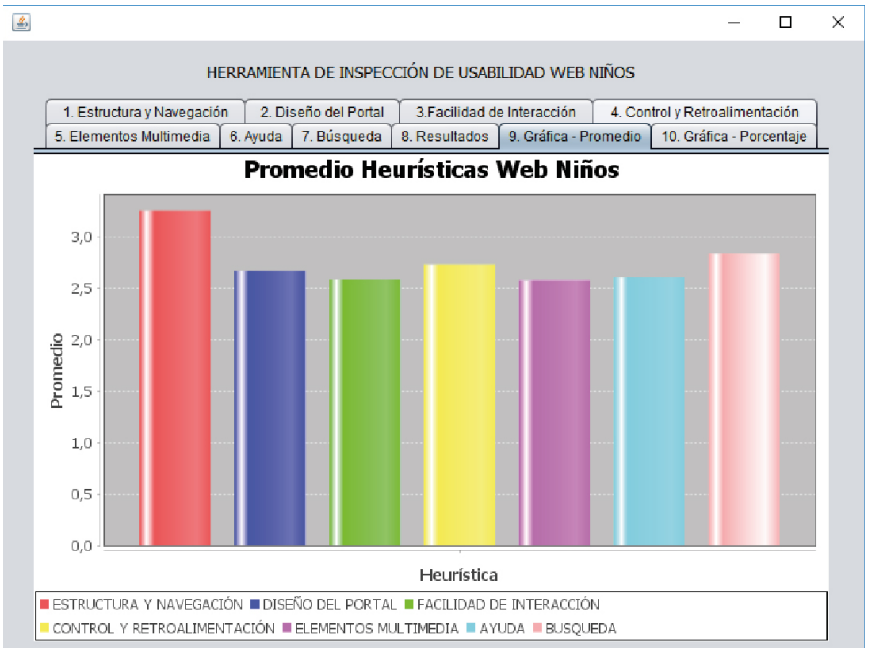


Fig. 5. *Pestaña de "Gráfica - Promedio"*

Del mismo modo, en la pestaña "Gráfica – Porcentaje", es posible visualizar la gráfica con el porcentaje de cumplimiento de cada uno de las heurísticas de usabilidad para sitios web presentados en la sección 4 (ver figura 6).

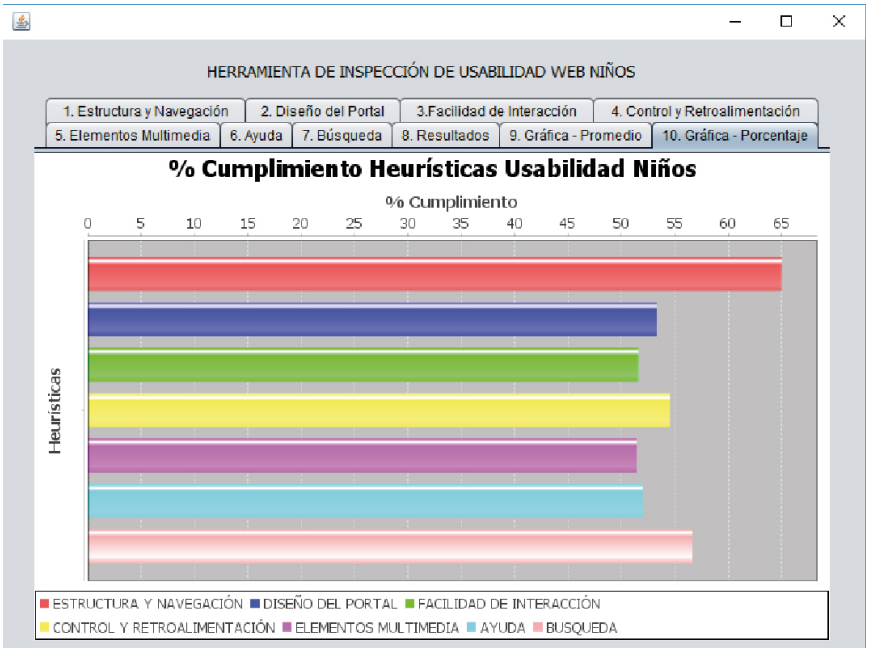


Fig. 6. *Pestaña de "Gráfica - Porcentaje"*

VI. CASO DE ESTUDIO

En este artículo además de presentar un conjunto de heurísticas de usabilidad dirigidas a la población infantil, también se propuso una herramienta de inspección con el fin de facilitar la aplicación y análisis de los resultados del test heurístico para sitios web dirigido a la población infantil, la cual plantea en sus diferentes pestañas un conjunto de criterios asociados a las siete heurísticas que deben cumplir los portales y aplicativos web para ser usables por los niños. La herramienta de inspección fue aplicada al portal web para niños de la Secretaría de Educación del Cauca en su portal infantil (https://sedcauca.gov.co/webinfantil/), con el fin de evaluar que tan adecuado resulta para la población infantil, a nivel de usabilidad. La Secretaria de Educación del Cauca busca administrar con eficiencia y equidad los recursos para garantizar la cobertura y calidad de la educación a través de la implementación de modelos administrativos y pedagógicos pertenecientes al contexto de los niños, niñas, adolescentes y jóvenes del Departamento del Cauca (Ver figura 7). En este sentido, la Secretaría de Educación propuso un portal web alternativo enfocado a la población infantil, el cual busca incluir a los niños como usuarios finales, así como dar cumplimiento a lo planteado en la Ley de Transparencia [12].

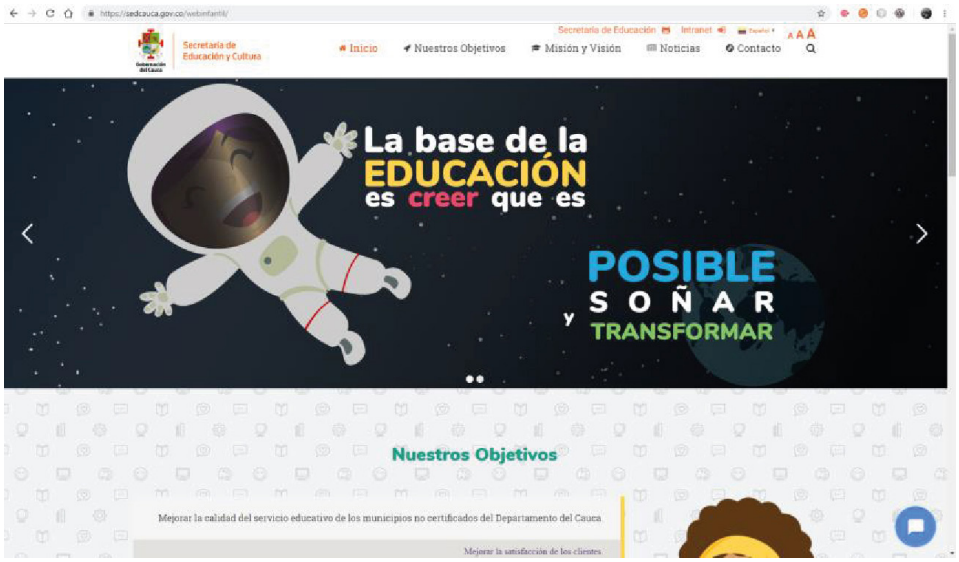


Fig. 7. *Portal de la Secretaria de Educación de Puebla*

6.1 Resultados de la inspección

Durante la realización de la inspección se tuvieron en cuenta cada uno de los criterios de usabilidad consideradas por la herramienta de evaluación de portales web para niños, las cuales a su vez están asociados a las diferentes pautas de las siete heurísticas propuestas. La evaluación fue realizada por un grupo de expertos sobre el portal de la Secretaría de Educación del Cauca en su portal infantil haciendo una inspección de cada heurística, tal como se muestra en la figura 8. En el caso de la figura 8, se presenta la evaluación asignada por los expertos a los criterios de la heurística "Elementos Multimedia".

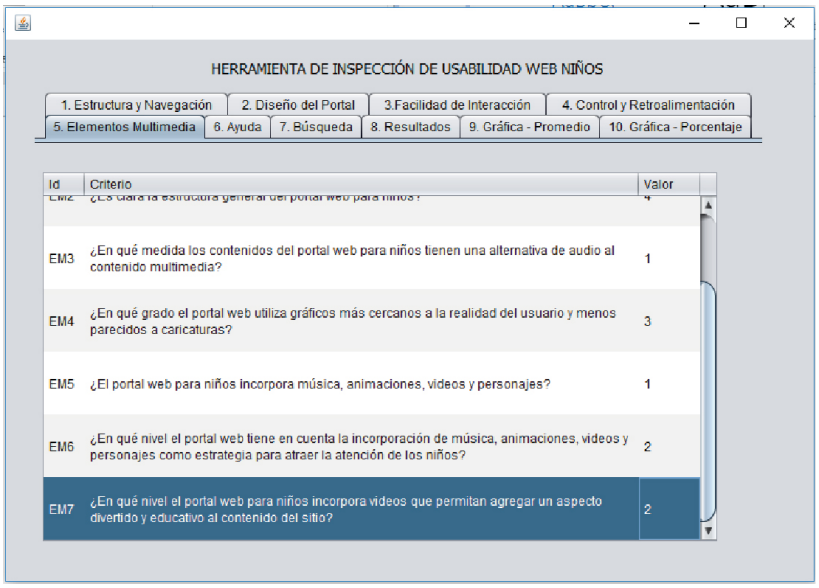


Fig. 8. *Evaluación de los expertos*

Una vez culminada la inspección sobre la página de la Secretaría de Educación del Cauca en su portal infantil y teniendo en cuenta los diferentes criterios obtenidos, en la figura 9 se presenta uno de los resultados gráficos obtenidos por la herramienta de inspección para páginas web centrada en la población infantil, el cual muestra el promedio de las valoraciones obtenidas por cada una de las siete heurísticas evaluadas.

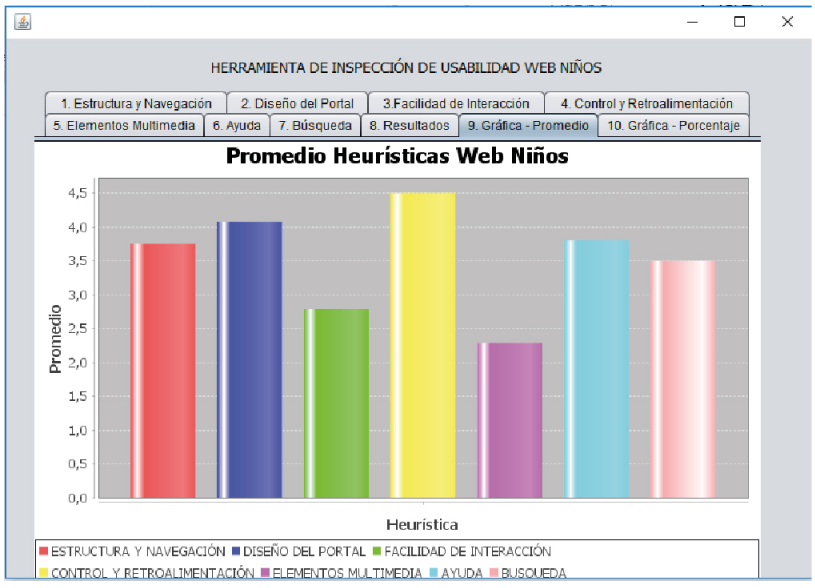


Fig. 9. *Promedio gráfico de las heurísticas*

Se puede observar que el portal es medianamente usable teniendo en cuenta los resultados de la inspección, ya que solo dos de sus heurísticas ("Diseño del Portal" y "Control y Retroalimentación") alcanzan un promedio de 4 en una escala de 1 a 5, quedando por debajo de 4 el resto de las heurísticas evaluadas, lo que en términos

matemáticos quiere decir que solo dos de las heurísticas alcanzan un valor mayor o igual al 80 % de cumplimiento. Del mismo, es posible analizar que 4 de las 7 heurísticas evaluadas tienen un porcentaje de cumplimiento superior al 70%, mientras que 2 de las heurísticas tienen un porcentaje de cumplimiento por debajo del 60%. La heurísticas que tiene un menor porcentaje de cumplimiento son "Facilidad de Interacción" con un porcentaje de 55% y "Elementos Multimedia" con un porcentaje de 45% valor considerado bajo, lo anterior considerando que aunque el portal puede resultar llamativo para un niño, no cuenta con elementos multimedia como videos o audios que presenten el contenido del portal e invitan a interactuar de manera más activa al usuario.

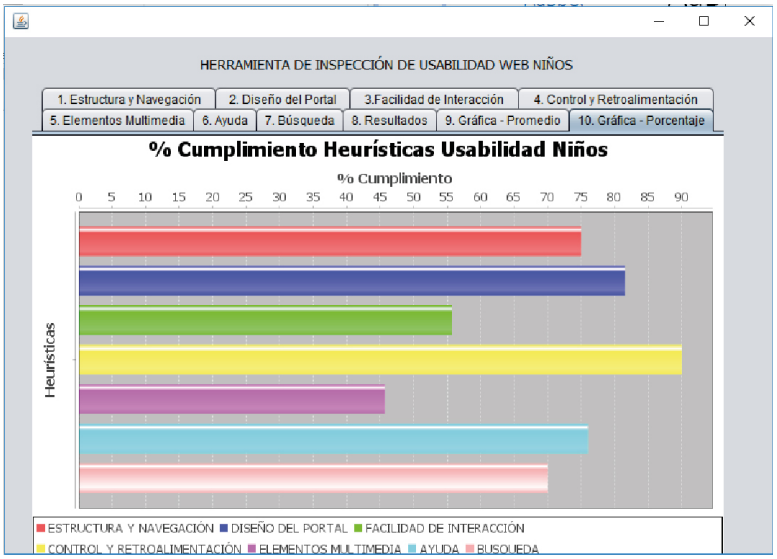


Fig. 10. *Cumplimiento de heurísticas de usabilidad web*

Finalmente el promedio general del portal web para niños es de 3.52, lo que corresponde a un porcentaje de cumplimiento de 70.4% (ver tabla 3). Conviene que se tengan en cuenta los diferentes criterios con el fin de que el portal tenga un valor superior al 80%, el cual puede considerarse acepTable.. De esta manera la inspección sobre el portal infantil de la Secretaría de Educación del Cauca ayuda a generar un conjunto de recomendaciones teniendo en cuenta los promedios por heurísticas (ver tabla 4), las cuales pretenden servir de ayuda para poder así generar posibles mejoras, teniendo en cuenta el número de usuarios potenciales que pueden acceder el sitio.

Table. 3. *Promedio por heurística*

Heurística	Promedio
H1: *Estructura y navegación*	*3.75*
H2: *Diseño del portal*	*4.07*
H3: *Facilidad de interacción*	*2.78*
H4: *Control y retroalimentación*	*4.50*
H5: *Elementos multimedia*	*2.86*
H6: *Ayuda*	*3.80*
H7: *Búsqueda*	*3.50*
Promedio General	***3.52***

6.2 Recomendaciones

Como resultado de la inspección realizada al portal de la Secretaria de Educación de Puebla, en esta sección se presentan un conjunto de recomendaciones (ver tabla 4) para las once heurísticas evaluadas y teniendo en cuenta los criterios que por valoración de los evaluadores obtuvieron el más bajo valor.

Table. 4. *Recomendaciones de usabilidad generadas*

Heurística	*Recomendación*
H2: *Estructura y navegación*	• *Se recomienda estimular a los niños con elementos reconocibles por ellos. Debido a que las experiencias de los niños en la vida son limitadas, algunas de las cosas con las que están más familiarizados se encuentran en la naturaleza. Es por eso importante utilizar elementos familiares y que hagan alusión a la naturaleza, tales como: agua, árboles, nieve y animales.* • *Se recomienda incluir formas extruidas como sombras, paisajes, efectos de biselado, gradientes brillantes u objetos flotantes que permitan a los niños estimular su imaginación.*
H3: *Facilidad de interacción*	• *Se recomienda incluir juegos que permitan ilustrar el contenido del portal, o que proporcionen un valor agregado a una información específica.*
H4: *Control y retroalimentación*	• *Se recomienda proporcionar un mecanismo para identificar las definiciones específicas de palabras usadas de modo inusual dentro del sitio web y que contribuya a entender términos complejos.*
H5: *Elementos multimedia*	• *Se recomienda el uso de contenidos de audio para ambientar e ilustrar de mejor forma el contenido del sitio web.* • *Se recomienda incorporar música, animaciones, videos y personajes divertidos dentro del sitio web, con el fin de hacer más entretenida la experiencia del usuario en el portal.*
H6: *Ayuda*	• *Se recomienda incluir ayuda contextual en tareas complejas puesto que los niños son más propensos a leer las instrucciones de uso que los adultos.*
H7: *Búsqueda*	• *Se recomienda generar una opción de búsqueda avanzada que asista al usuario en caso de no poder ofrecer resultados para una consulta específica.*

VII. CONCLUSIONES

En este trabajo se propuso como aporte un conjunto de heurísticas de usabilidad dirigidas a portales web infantiles, teniendo en cuenta que no se ha evidenciado la existencia de principios y pautas puntuales asociados a este tema. Del mismo modo, se generaron un conjunto de criterios asociados a cada una de las heurísticas propuestas, los cuales están expresados en términos de preguntas, y pretenden guiar al evaluador de cara a verificar si el portal web es adecuado para la población infantil, desde la perspectiva de la usabilidad.

El principal aporte de la herramienta propuesta en este trabajo es la posibilidad de evaluar cada uno de los criterios asociados a las heurísticas de usabilidad para sitios web dirigidos a la población infantil, así como generar de manera automática la gráfica de promedios y la gráfica de porcentaje de cumplimiento de las heurísticas propuestas. De este modo, se pretenden obtener indicadores claros a los evaluadores, de cara a mejorar la usabilidad de un sitio web.

El lenguaje Java demostró ser adecuado para la implementación de herramientas de inspección usabilidad para portales web infantiles, teniendo en cuenta el componente de manejo de pestañas (JTabbedPane), el componente de manejo de tablas (JTable.) y la API para la generación de gráficos estadísticos JFreeChart.

A modo de reflexión, es importante resaltar que dado que gran parte de los portales web gubernamentales han sido creados a través del uso de gestores de contenido, los sitios web construidos a través de estos, traen consigo diversos problemas de usabilidad que no necesariamente dependen del diseñador de la página, ya que estos gestores en su mayoría fueron concebidos sin tener en cuenta criterios de usabilidad y mucho menos criterios de usabilidad dirigidos a la población infantil.

Es importante resaltar que aunque desde la página web de la Secretaría de Educación del Cauca, se ha hecho un esfuerzo por construir un portal enfocado a la población infantil, aún existen ciertas heurísticas de usabilidad que no se cumplen a cabalidad, tales como "Elementos Multimedia" y "Facilidad de Interacción". Lo anterior considerando que aunque el portal puede resultar llamativo para un niño, no cuenta con elementos multimedia como videos o audios que presenten el contenido del portal e inviten a interactuar de manera más activa al usuario.

Como trabajo futuro derivado de la presente investigación, se pretende incluir dentro de la herramienta propuesta en este trabajo, la posibilidad de generar recomendaciones automáticas derivadas de los resultados estadísticos y gráficos obtenidos en la prueba.

REFERENCIAS

[1] *M. Virtual, «Importancia de un sitio web,» 25 Junio 2016. https://medium.com/@marcavirtual/importancia-de-un-sitio-web-d1e1479cc416.*

[2] *M. J. Muñoz Gutierrez, K. Márceles Villalba y G. E. Chanchi Golondrino, «Recomendaciones de usabilidad para el diseño de sitios web dirigidos a la población infantil,» de Ingeniería, tecnología y automatización, Colombia, CIMTED, 2018, pp. 38-53.*

[3] *G. M. d. l. Teja, «Usabilidad y accesibilidad web,» 2009. http://www.semac.org.mx/archivos/6-11.pdf.*

[4] *D. Moner Cano y J. Sabaté Alsina, «Usabilidad,» 2009. http://www.unsj.edu.ar/unsjVirtual/comunicacion/seminarionuevastecnologias/wpcontent/uploads/2015/04/usabilidad_para_web.pdf.*

[5] *W. Sanchez, «La usabilidad en ingeniería de software,» 2012. https://es.scribd.com/document/338137562/2-La-Usabilidad-en-Ingenieria-de Software-Definicion-y-Caracteristicas.*

[6] *M. P. González, A. Pascual y J. Lorés , «Evaluación Heurística,» 2008. http://jrgonzalez.es/wp-content/files/HE.pdf.*

[7] *J. G. Enriquez y S. I. Casas, «usabilidad en aplicaaciones móviles,» 2013. https://dialnet.unirioja.es/servlet/articulo?codigo=5123524.*

[8] *M. Bolaños Pizarro , A. Vidal, C. Navarro Molina, J. C. Valderrama y R. Benavent, «Usabilidad: concepto y aplicaciones en las páginas web,» 2007. http://sedom.es/wpcontent/themes/sedom/pdf/4cbd571ba0446art_3.pdf.*

[9] *J. Nielsen, «Sitios web para niños,» 13 Septiembre 2010. https://www.nngroup.com/articles/childrens-websites-usability-issues/.*

[10] *L. García, A. Pernnett y J. Cano, «Estudio exploratorio de usabilidad para niños de Colombia,» 26 Enero 2017. https://dialnet.unirioja.es/descarga/articulo/6416720.pdf.*

[11] *A. Viklund, «JFreeChart,» 2005.: http://www.jfree.org/jfreechart/.*

[12] *A. g. d. l. nación, «TRANSPARENCIA Y ACCESO A INFORMACIÓN PÚBLICA,» 2014, http://www.archivogeneral.gov.co/transparencia.*

Biografía Autor(es)

a. Ingeniera Informática. Docente, Institución Universitaria Colegio Mayor del Cauca. E- mail: juliethgutierrez@unimayor.edu.co. Cauca, Colombia.

b. Ingeniera Informática. Docente, Institución Universitaria Colegio Mayor del Cauca. E- mail: lmdominguez@unimayor.edu.co. Cauca, Colombia.

c. Ingeniera Informática. Docente, Institución Universitaria Colegio Mayor del Cauca. E- mail: danielaperez.edu.co. Cauca, Colombia.

d. Doctor en Ingeniería Telemática. Docente, Institución Universitaria Colegio Mayor del Cauca. E- mail: gchanchi@unimayor.edu.co. Cauca, Colombia.

e. Magister en Seguridad Informática y de Sistemas. Docente, Institución Universitaria Colegio Mayor del Cauca. E- mail: kmarceles@unimayor.edu.co. Cauca, Colombia.

Sistema de gestión del seguimiento del progreso académico

Cesar Augusto Castelblanco Celis (a)
William Joseph Orozco (b)

Resumen

El seguimiento al progreso académico de los estudiantes es de vital importancia para mejorar la atención del estudiante y dar frente a las deficiencias que éste pueda presentar en su proceso de formación. Este seguimiento puede mejorar los indicadores de deserción, retención, y repetición por parte de los estudiantes. Así mismo, podría mejorar la evaluación de la Institución por parte del Ministerio de Educación Nacional (MEN), por lo anterior se propone un Modelo de Progreso Académico (MPA) y un Modelo de Gestión Académica (MGA). El MPA se apoya en los indicadores de rendimiento académico, en cuatro dimensiones, que permiten realizar un ciclo de la mejora continua (PHVA). El MGA se encarga de las acciones sobre los resultados del progreso académico del estudiante, este modelo cuenta con un dashboard que permite el seguimiento y control visual del progreso académico de cada estudiante y apoya a la toma de decisiones a partir de un sistema de gestión de reportes centrado en los modelos mentales de los usuarios.

Palabra Clave: Gestión académica, programa académico, rendimiento académico.

Academic progress monitoring management system

Abstract

Monitoring the academic progress of students is of vital importance to improve student attention and address any deficiencies that the student may present in their training process. This monitoring can improve the indicators of dropout, retention, and repetition by students. Likewise, it could improve the evaluation of the Institution by the National Ministry of Education (MEN), therefore, an Academic Progress Model (MPA) and an Academic Management Model (MGA) are proposed. The MPA is based on academic performance indicators, in four dimensions, which allow a cycle of continuous improvement (PHVA) to be carried out. The MGA is responsible for the actions on the results of the students academic progress; this model has a dashboard that allows the visual monitoring and control of the academic progress of each student and supports decision-making based on a management system of reports focused on the users mental model.

Keywords: Academic management,academic progress,academic performance.

I. INTRODUCCIÓN

Para la Institución Educativa (IE) poder contar con herramientas que apoyen su gestión educativa, en especial el proceso de seguimiento académico a los estudiantes se convierte en un factor de gran valor en el que se ven reflejado los índices de calidad educativa no solo de la institución sino de los estudiantes, es por medio de este seguimiento que la Gestión Académica (GA) puede encontrar de manera oportuna fortalezas o debilidades en su proceso formativo, permitiendo potencializar fortalezas o esTable. cer las acciones de mejoramiento necesarias tanto para la IE como para el estudiante.

Esta trabajo se centra en el estudio de la medición del progreso académico del estudiante de la formación básica y media, porque es a partir de este seguimiento que la institución educativa en sus diferentes niveles puede definir las acciones necesarias para el mejoramiento en la formación del estudiante y de otros indicadores que afecten la evaluación de la institución no solo por parte de organismos nacionales como el Ministerio de Educación Nacional sino por parte de los mismos integrantes de la comunidad educativa.

Esta necesidad de realizar el seguimiento al progreso académico del estudiante de manera eficiente genera la oportunidad de desarrollar una metodología y un instrumento de apoyo a la toma de decisiones por parte de la persona encargada del proceso de la GA, aportando de esta manera a la mejora del estudiante como de la IE; para lograr con este objetivo se requieren entonces de estrategias y herramientas acordes a las necesidades de la institución, que le aporten a los responsables de cada proceso información integral y oportuna sobre la realidad de la práctica pedagógica en cada uno de los ambientes de aprendizaje de cada curso por parte de docentes y estudiantes, estas estrategias o herramientas deberán además permitir el seguimiento no solo en los aspectos estáticos del proceso de formación sino de aspectos dinámicos en el estudiante como es el conjunto de sus competencias blandas el cual es complejo y diverso en cada estudiante.

Se pretende resolver la dificultad que se tiene en realizar el seguimiento al progreso académico del estudiante por parte de docentes y directivos docentes en la IE, en especial la pública la cual presenta un número significativo de dificultades para su medición y seguimiento, en donde la GA se ve enfrentada a factores como la falta de recursos tecnológicos, el gran número de estudiantes por curso, la diversidad de la realidad social, familiar y económica de cada uno de los estudiantes y a una cantidad considerable de carga laboral que impiden realizar un seguimiento al progreso académico de manera oportuna y que verdaderamente permita esTable.cer acciones para apoyar al estudiante en la superación de sus debilidades y en la consecución de sus logros.

A continuación se presenta el estado del arte de la cuestión, seguido se presenta un modelo de rendimiento académico que soporta el progreso académico. Seguido se presenta una propuesta de seguimiento académico y finalmente se presenta la propuesta de dashboard y la plataforma de apoyo.

II. ESTADO DEL ARTE

La Institución Educativa (IE) en los últimos años ha tenido la necesidad de adoptar metodologías de otras áreas del conocimiento como lo afirma Botero en su artículo sobre las cinco tendencias de la educación, “El objeto de la gestión educativa como disciplina, es el estudio de la organización del trabajo en el campo de la educación, por tanto, está influenciada por teorías de la administración, pero además, existen otras

disciplinas que han permitido enriquecer el análisis, como son: la administración, la filosofía, las ciencias sociales, la psicología, la sociología y la antropología"[1], estas áreas del conocimiento han permitido a la IE adaptarse y cumplir con las nuevas exigencias que le imponen los gobiernos en sus procesos de descentralización, procesos que sean venido implementado de manera más relevante desde la constitución de 1991 y en especial de la Ley 115 de 1994 conocida como Ley General de Educación, estos cambios en la manera de dirigir ha requerido por parte de los directivos docentes adquirir nuevas competencias gerenciales que le permitan ver a la IE como una organización donde confluyen procesos más allá de lo netamente académico, como los procesos financieros, conTable.s, talento humano y la gestión de los recursos físicos entre otros.

El MEN definió cuatro áreas de gestión sobre las cuales las IE deben dar cuenta, estas áreas de gestión son: la gestión directiva, las gestión administrativa y financiera, la gestión a la comunidad y la gestión académica, es esta última el eje central de este trabajo en especial en su proceso de seguimiento académico, por otra parte esta área de gestión la define Alejandra Dalila Rico Molano en su artículo sobre la gestión educativa "como un proceso organizado y orientado a la optimización de procesos y proyectos internos de las instituciones, con el objetivo de perfeccionar los procedimientos pedagógicos, directivos, comunitarios y administrativos que en ella se movilizan"[2], la autora cita también a Romero (1982: 14), el cual define la gestión educativa "como un proceso concebido para integrar recursos y coordinar actividades especializadas e interdependientes, en busca del cumplimiento de los objetivos propuestos"[2], por tal motivo el desarrollar un modelo de progreso académico junto con una herramienta que permita el control visual es de gran apoyo no solo para el seguimiento académico sino para todos los procesos e integrantes de la comunidad educativa, a continuación se realiza un acercamiento conceptual de los principales elementos que fueron tenidos en cuenta para el desarrollo del modelo.

2.1 Rendimiento Académico (RA)

Inicialmente lograr definir conceptualmente el RA se hace complejo por la gran variedad de factores internos y externos que lo afectan, como lo afirma Gloria Pérez "El concepto de rendimiento académico es complejo, dado que en él inciden multitud de variables"[3], Héctor Lamas afirma que "la complejidad del rendimiento académico se inicia desde su conceptualización. En ocasiones se le denomina aptitud escolar, desempeño académico o rendimiento escolar, pero, generalmente, las diferencias de concepto sólo se explican por cuestiones semánticas ya que se utilizan como sinónimos"[4], Rubén Edel Navarro plantea que "la complejidad del rendimiento académico inicia desde su conceptualización, en ocasiones se le denomina como aptitud escolar, desempeño académico o rendimiento escolar, pero generalmente las diferencias de concepto sólo se explican por cuestiones semánticas, ya que generalmente, en los textos, la vida escolar y la experiencia docente, son utilizadas como sinónimos"[4], alineados con lo anteriormente expuesto y con fin de unificar criterios para el desarrollo de la tesis siempre haremos referencia a rendimiento académico.

Entonces el RA según Oscar Erazo en su artículo titulado "El Rendimiento Académico, un Fenómeno de Múltiples Relaciones y Complejidades", cita a (Jiménez, 2000; citado por Navarro, 2003; y Paba, 2008; citado por Zapata, De Los Reyes, Lewis & Barceló, 2009) y donde se concreta que el " Rendimiento Académico (R.A), es entendido como el sistema que mide los logros y la construcción de conocimientos en los estudiantes, los cuales se crean por la intervención de didácticas educativas que son evaluadas a

través de métodos cualitativos y cuantitativos en una materia"[4], Héctor Lamas en su artículo sobre rendimiento académico cita a varios autores entre los cuales destacamos a Martínez-Otero (2007) el cual define el RA desde un enfoque humanista como "el producto que da el alumnado en los centros de enseñanza y que habitualmente se expresa a través de las calificaciones escolares" (p. 34)"[4].

En esencia podemos afirmar entonces que el RA se centra en la medición de las capacidades y de los conocimientos adquiridos por el estudiante en un proceso formativo en la IE en un determinado momento, sin embargo "el propósito del rendimiento académico es alcanzar una meta educativa, un aprendizaje, sin embargo este puede verse afectado por diversas variables de índole personal y social, ahora bien, está medición del rendimiento académico por medio de indicadores se convierte en el insumo para definir el progreso académico del estudiante durante su proceso de formación.

2.2 Progreso Académico

El progreso académico es uno de los elementos con mayor relevancia hoy en día dentro de las instituciones educativas por ser éste el componente que muestra los resultados obtenidos en la práctica pedagógica y el cual es evaluado de manera periódica por diferentes actores internos y externos a la IE, en general el Progreso Académico (PA) se evalúa a partir de los resultados de los indicadores de rendimiento académico obtenidos en las evaluaciones aplicadas a cada uno de los estudiantes durante un periodo de tiempo determinado y habiendo cumplido con el desarrollo de los contenidos temáticos definidos en el plan de estudios en un determinado ambiente de aprendizaje.

De igual manera Loaiza y Toloza citan a (Carballo, 1990) plantea que "La evaluación es un proceso integral del progreso académico del educando: informa sobre conocimientos, habilidades, intereses, actividades, hábitos de estudio, etc. Es un método que permite obtener y procesar las evidencias para mejorar el aprendizaje y la enseñanza"[5], del anterior planteamiento se puede esTable.cer entonces que la evaluación como tal es solo una parte del progreso académico y que es a partir de esta que se obtiene la información necesaria para medir el progreso académico del estudiante en las diferentes dimensiones durante su proceso de formación.

2.3 Seguimiento Académico

El seguimiento académico se puede considerar como la actividad que realiza el docente o la institución al estudiante en diferentes momentos de su proceso educativo, este seguimiento se puede comprender según Pérez (2007) citado por Paredes y Rodríguez en la tesis de Gerencia Educativa, "como la concreción de los resultados que se plantean desde la elaboración del PEI y plan de acción individual de cada centro educativo durante cada periodo del año, resultados que según el MEN (2008) se evidencian en aspectos como: asistencia de los estudiantes, calificaciones, pertinencia de la formación recibida, promoción y recuperación de problemas de aprendizaje, entre otros"[6], de acuerdo a la anterior definición se puede conocer que el seguimiento académico no centra exclusivamente en términos de resultados académicos sino que abarcan otros factores relacionados con el actuar del estudiante dentro del proceso de enseñanza-aprendizaje en la IE.

De acuerdo a los anteriores planteamientos, el seguimiento académico es definido en este trabajo como el proceso mediante el cual la IE y el docente realizan observación sobre el cumplimiento de los objetivos planteados en el Proyecto Educativo Institucional y en el desarrollo de los contenidos curriculares en cada uno de los módulos de formación; por otra parte, para el desarrollo de los modelos de Progreso Académico (MPA) y de Gestión Académica (MGA), fue necesario delimitar los elementos de seguimiento académico ya que estos modelos se centran en el seguimiento de los resultados académicos y en las actividades de apoyo académico.

2.4 Dashboard

Los cuadros de mando integral (CMI) o Dashboard son una herramienta administrativa que apoyan a la alta dirección en la toma de decisiones a partir del seguimiento visual de variados indicadores que se relacionan con los objetivos estratégicos de la organización, a pesar de que estos Table.ros se vienen utilizando con mayor fuerza desde principios de la década de los 90's en diferentes sectores de la industria, es desde hace muy pocos años que su uso se viene realizando en el sector educativo.

En el artículo sobre el "diseño de un cuadro de mando integral realizado" por Adriana Avalos Robles y Daniel Eduardo Paz Pérez, definen el cuadro de mando integral como "una herramienta informática que le permite a cualquier empresa contar con una fuente de información oportuna sobre diferentes aspectos de su operación que son vitales para una adecuada toma de decisiones, ya que se conocen las metas de la misma y se pueden controlar los resultados al estar monitoreando constantemente cualquier desviación de los indicadores esTable.cidos"[7], por otra parte (Kerzner,2013) citado por Martínez Robalino plantea que "los dashboard son mecanismos de representación visual utilizados en un sistema de medición operativo de rendimiento, que mide el desempeño contra objetivos y umbrales usando datos de tiempo adecuado"[8], de igual manera en este mismo artículo se define que los dashboard son herramientas "que ayudan a las personas a identificar visualmente tendencias, patrones y anomalías, razonan sobre lo que ven y ayudan a guiarlos hacia decisiones eficaces"[8], en esta misma línea para la IE es de gran importancia contar con una herramienta de este tipo que le permita el seguimiento y control visual del progreso académico a cada uno de los estudiantes.

Por otra parte el diseño y construcción del dashboard se debe de caracterizar por su sencillez y "simplicidad", como afirma Martínez Robalino "El diseño del dashboard no tiene que ser un proceso doloroso, cuando se crea un dashboard, se debe alcanzar la visualización correcta de los datos y llevar esas métricas a las personas adecuadas, es pensar en el usuario final y eliminar las barreras cognitivas, el resultado será un dashboard fácil de leer, que faculta a su equipo para tomar medidas, ya que por otro lado sería un simple conjunto de datos"[8].

III. MODELO DE PROGRESO ACADÉMICO (MPA)

El Modelo Progreso Académico (MPA) le permite a la IE desde la Gestión Académica realizar el seguimiento académico al estudiante durante cada uno de los periodos que conforman el año escolar, este seguimiento se compone de dos elementos, el primero que hace referencia al RA y el segundo al PA en la siguiente gráfica se pueden observar cada uno de estos elementos con sus respectivos componentes.

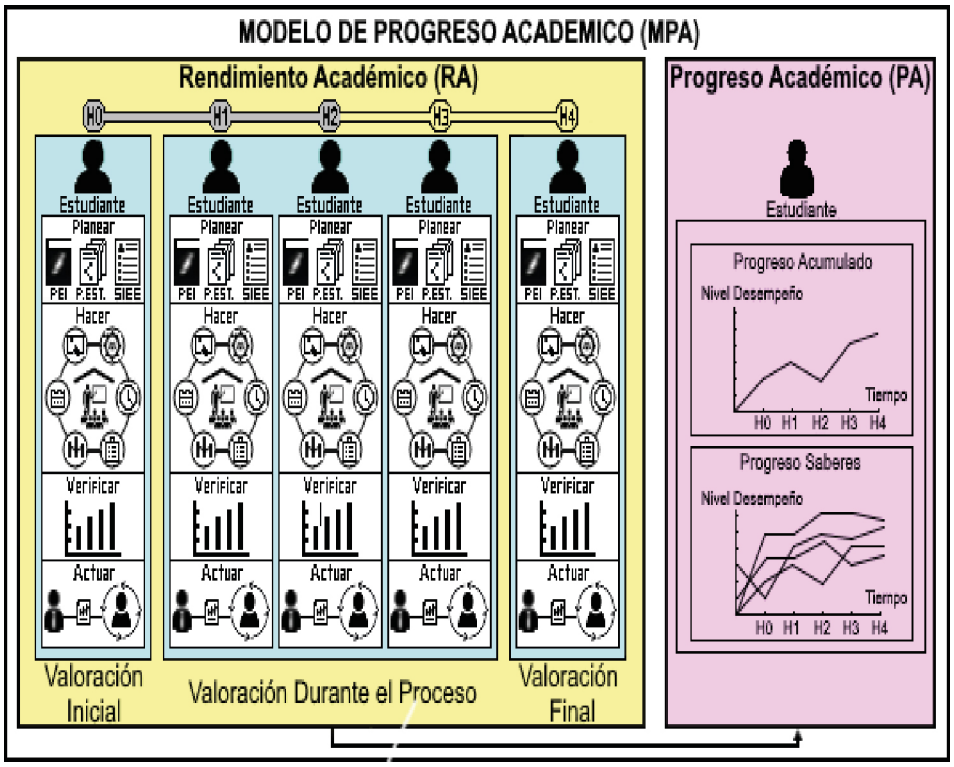


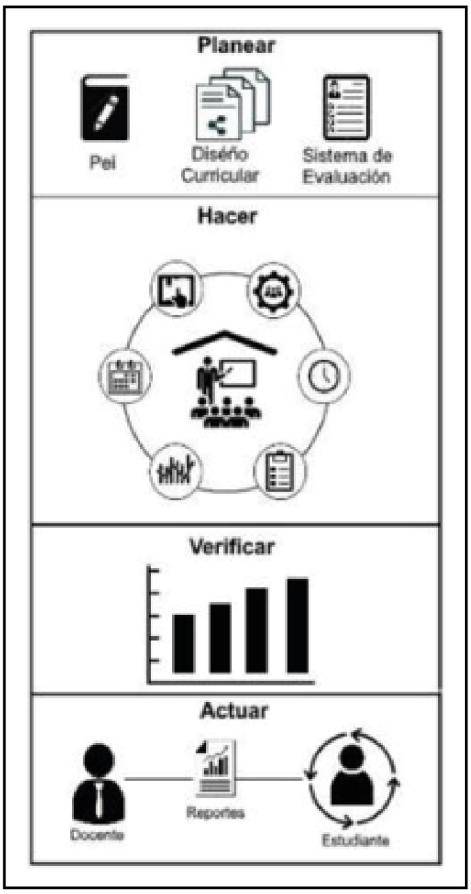
Fig. 1. *Modelo de Progreso Académico*

3.1 Elementos del Modelo de Progreso Académico (MPA)

A continuación, se realiza la descripción de cada uno de los elementos que conforman cada una de las partes del MPA.

Ciclo de Calidad del Hito Académico

El ciclo de Calidad del Hito Académico contiene los elementos que hacen parte de la actividad pedagógica y de la relación docente-estudiante en el ambiente de aprendizaje, estos elementos fueron agrupados según su función en el ciclo de calidad PHVA (planear, hacer, verificar y actuar), por otra parte, el poder utilizar el ciclo de la mejora continua ayuda a organizar cada uno de los hitos académicos de tal manera que se pueda realizar un mejor seguimiento y control a cada uno de quienes participan, estudiantes, docentes y directivos docentes, en la siguiente gráfica se definen cada uno de los elementos que hacen parte de cada una de las etapas del PHVA.



Planear:

En esta etapa la Institución define su Proyecto Educativo Institucional, sus contenidos y su sistema de evaluación, como elementos esenciales que marcan la ruta a seguir en la práctica pedagógica.

Hacer:

Es quehacer diario del estudiante mediado por la didáctica en escenarios concretos de aprendizaje.

Verificar:

Esta etapa realiza la valoración de cada uno de los indicadores de rendimiento académico, en términos del cumplimiento de los criterios de desempeño definidos en los contenidos.

Actuar:

Se refiere a los planes de acción en los que el estudiante replantea y refuerza sus logros.

Fig. 2. *Etapas del PHVA.*

Hitos del MPA

Sergio Tobón plantea que la valoración del estudiante "se debe de aplicar en tres momentos al inicio, durante el proceso formativo y al final de este"[9], es decir que los cinco hitos del RA fueron clasificados de la siguiente manera, el hito 0 hace referencia a la valoración diagnóstica del estudiante una vez inicia su proceso de formación, el objetivo de este hito es medir los indicadores de RA en cuatro saberes (saber-saber, saber-hacer, saber-ser y saber-institucional), resultados con los cuales se define una línea base para la medición del PA y esTable.cer de ser necesario un plan de mejoramiento para el estudiante, es importante destacar que este plan de mejoramiento se podrá ejecutar al inicio de su formación o de manera paralela durante el desarrollo de la primera etapa y hasta antes de valoración del hito 1, con esto se busca la nivelación del estudiante con los mínimos exigidos en cada curso.

Los hitos del uno al tres son definidos como hitos del proceso y se valoran cada 10 semanas, el cuarto hito se define como final, el periodo de medición para cada hito se definió de acuerdo a la Ley 115 (Ley General de Educación)[10], donde esTable.ce que el calendario académico se compone de 40 semanas de trabajo con estudiantes, las cuales se dividen en dos semestres de 20 semanas, es decir que se contará con

dos hitos académicos por semestre, sin embargo esto puede variar de acuerdo a las políticas de cada IE.

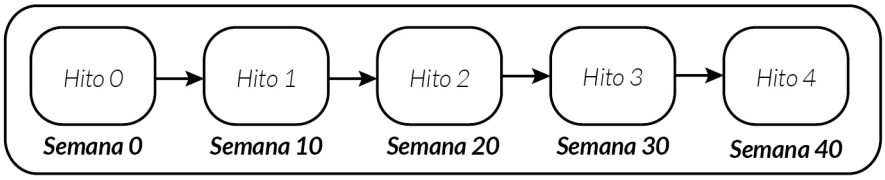


Fig. 3. *Hitos y tiempos*

La medición del RA que se realiza en cada hito académico se puede realizar con diferentes instrumentos de valoración, sin embargo se recomienda se realice con la lista de cotejo, debido a que esta permite una mejor valoración en cada uno de los saberes, al tener la valoración de cada uno de los hitos se pueden generar tres situaciones, primero que el estudiante cumpla con los niveles mínimos esTable.cidos en la escala de valoración por lo que podrá continuar con su proceso formativo, segundo cuando el estudiante no cumple con los niveles mínimos de la escala, lo que lo obliga a presentar un plan de mejoramiento con el fin de suplir las deficiencias y volver a presentarse a la valoración en las mismas condiciones hasta que logre demostrar la superación de las deficiencias y tercero cuando un estudiante desde un proceso de autoevaluación y a pesar que haya superado el hito decida presentar el plan de mejoramiento con el fin de mejorar sus indicadores de rendimiento académico.

Escala de Valoración

La escala de valoración propuesta para el RA se define desde el enfoque de la formación basada en competencias, es decir que la "evaluación" se centra más en el desempeño del estudiante en escenarios que son contextualizados, en este enfoque Sergio Tobón en su libro Formación Basada en Competencias plantea que la VALORACIÓN "es un procedimiento para generar valor (reconocimiento) a lo que las personas aprenden, basado en la complejidad, puesto que tiene en cuenta las múltiples dimensiones y relaciones entre estudiantes, empresas y docentes"[9], Sergio Tobón también plantea que "la valoración, aunque constituye un juicio de valor, se regula con base en una serie de criterios previamente acordados con los estudiantes"[9], de acuerdo a estos planteamientos concluimos entonces que el concepto de valoración se centra en la emisión de juicios de valor en contexto con el fin de entregar oportunidades de mejora y generar retroalimentación durante el proceso de formación de cada estudiante.

Es ahora necesario definir los parámetros para la ponderación y el puntaje que permita el análisis cuantitativo de cada competencia, en lo descrito en el libro "Evaluación de la Competencias como Proceso de Valoración", de Sergio Tobón, Julio H. Pimienta Prieto y Juan Antonio García Fraile definen que la ponderación "consiste en asignarle un valor cuantitativo a los criterios e indicadores respecto a su grado de contribución para valorar la competencia…"[11], además estos mismo autores manifiestan que por lo general la escala que se utiliza va de 0% al 100%, recalcando que se puede utilizar cualquier otra escala, para el MPA se define que la escala para los indicadores de RA que miden cada uno de los cuatro saberes se debe realizar de 0% al 100% y donde la valoración mínima deberá ser del 70% para ser aprobada.

Teniendo en cuenta que el grado de complejidad y de exigencia al estudiante debe incrementarse en la medida que transcurre cada periodo académico se define un peso porcentual para cada hito académico y que de manera acumulativa le permita llegar al

100% o como mínimo al 70% para aprobar, en la siguiente gráfica se observa el peso en la valoración según el hito académico y la etapa en que se encuentre.

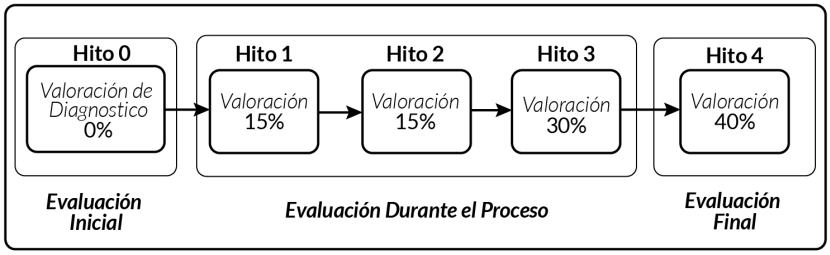


Fig. 4. *Ponderación del Progreso Académico del Estudiante*

Niveles de Desempeño

El Ministerio de Educación Nacional (MEN) en el Decreto 1290 de 2009 entregó autonomía a la IE para esTable.cer su propio Sistema Institucional de Evaluación de Estudiantes (SIEE), y define la escala de desempeño que permitiera la movilidad de estudiantes por diferentes IE y la cual está conformado por los siguientes niveles: "Desempeño Superior, Desempeño Alto, Desempeño Básico, Desempeño Bajo"[12], con esto lo que la IE educativa debe hacer es dar equivalencia de la escala de valoración del RA contra cada nivel de desempeño, para el caso del MPA se define de la siguiente manera:

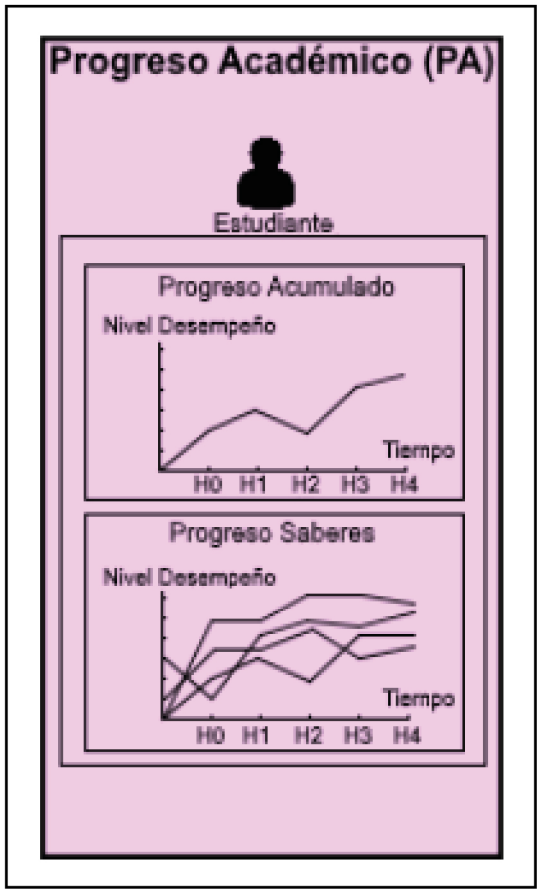
Table. 1. *Equivalencia Valoración Institucional - Nacional*

Valoración Nacional	*Valoración SIEE*
Desempeño Superior	*90% – 100%*
Desempeño Alto	*80% – 89%*
Desempeño Básico	*70% – 79%*
Desempeño Bajo	*0.0% – 69%*

De acuerdo con la anterior se esTable.ce un código de colores que permitiera de manera gráfica identificar el nivel de desempeño académico del estudiante en cada uno de los hitos académicos, a continuación, se muestra la conformación de este código.

Table. 2. *Semaforización Indicador de RA*

Nivel de Desempeño	*Color*	*Valoración SIEE*
Desempeño Superior		*90% – 100%*
Desempeño Alto		*80% – 89%*
Desempeño Básico		*70% – 79%*
Desempeño Bajo		*0.0% – 69%*



Progreso Académico (PA)

El PA es el segundo elemento del MPA y es el que nos permite realizar el seguimiento al estudiante en un determinado periodo de tiempo a partir de comparaciones entre los resultados obtenidos en los diferentes hitos, este seguimiento permite a la Gestión Académica el poder identificar de manera oportuna las oportunidades para ser potencializadas en laformación, o por el contrario las deficiencias las cuales deberán de ser tratadas por medio de planes de mejoramiento por parte de cada estudiante, esto se puede realizar gracias a que el PA se divide en dos partes, la primera muestra el acumulado del PA de acuerdo a los resultados obtenidos en cada hito académico y la segunda nos muestra el comportamiento de cada uno de los saberes o dimensiones en un determinado periodo de tiempo.

Fig.5. *Progreso Académico (PA)*

Plan de Mejoramiento

El plan de mejoramiento es la herramienta con la que cuenta el docente para apoyar al estudiante en la superación de las dificultades evidenciadas en la valoración de cada hito académico, este plan de mejoramiento se debe de caracterizar por identificar de manera focalizada las falencias del estudiante en cada una de las dimensiones valoradas y de esTable.cer las acciones necesarias en común acuerdo con el estudiante para luego ser valorado nuevamente en las mismas condiciones.

Los planes de mejoramiento en el MPA hacen parte del HACER en el ciclo de calidad sobre los cuales se soporte el funcionamiento de un hito académico, de tal manera que es necesario que el docente no solo se centre en el mínimo alcanzado en la valoración por parte de estudiante, sino que se debe de realizar un análisis de causa raíz que permita encontrar las verdaderas causas de los resultados obtenidos en cada uno de las dimensiones valoradas, para el MPA se selecciona el método de los cinco ¿por qué?, está técnica se basa en la formulación de ¿por qué? al menos cinco veces, al final una vez sea más difícil responder al ¿por qué?, la causa más probable que ocasiona el bajo rendimiento del estudiante habrá sido encontrada, los planes de mejoramiento junto con sus respectivas evidencias en la ejecución deberán de quedar consignadas en formatos específicos que la institución designe e ingresados al sistema de información para posteriores consultas y análisis.

IV. MODELO DE GESTIÓN ACADÉMICA (MGA)

Una de las cuatro áreas de gestión que define el MEN es la Gestión Académica, la cual se encuentra a cargo de los procesos de diseño pedagógico o curricular, las prácticas pedagógicas, la gestión del aula y el seguimiento académico, el MGA que se desarrolló en esta tesis hace énfasis en el seguimiento académico de los estudiantes a partir de la medición de su progreso académico y de los resultados obtenidos en el RA.

El MGA se compone de dos elementos, el primero se encuentra integrado por el triángulo de la gestión académica, el Consejo Directivo, el Consejo Académico, el Docente y el padre de familia o acudiente, este elemento se centra en la gestión realizada en el triángulo académico por el Coordinador Académico quien se encarga de realizar el seguimiento al RA y al PA del estudiante con el objetivo de entregar diversos tipos de reportes al órgano encargado teniendo en cuenta sus funciones, el Consejo Directivo quienes son los encargados de la revisión y aprobación de nuevos contenidos curriculares, reformas educativas o pedagógicas, el Consejo Académico quienes son los responsables de tratar todos los temas relacionados con lo académico tanto de la IE como de los estudiantes y padres de familia, al docente quien es el encargo de dirigir el acto pedagógico, y al estudiante y padres de familia o acudientes como actores principales del proceso de formación.

El segundo elemento se compone de un sistema de almacenamiento de información, de un dashboard y de una plataforma para la generación de reportes centrados en el usuario. Para el almacenamiento de la información se desarrolla una base de datos que permite mantener la información generada en cada uno de los elementos y componentes de los modelos de PA y de GA de cada estudiante, en referencia al dashboard o cuadro de mando integral (CMI) su principal función es la de poderse realizar el control visual del RA y del PA del estudiante por parte del Coordinador Académico, fortaleciendo de esta manera la toma de decisiones y el esTable.cimiento de las acciones de mejora necesarias, ya sea para la IE o para los estudiantes, los reportes centrados en el usuario son generados de acuerdo a la necesidad o requerimientos de los integrantes de la IE utilizando la Plataforma para el Soporte del Ciclo de Vida del Reporte Centrado en el Usuario (PGR)[13], proyecto de tesis que fue desarrollado en la Universidad del Quindío en el grupo de investigación SINFOCI y el cual permite la generación de reportes de manera automática de acuerdo a las necesidades o requerimientos del usuario, en la siguiente gráfica se puede observar cada uno de los elementos del MGA.

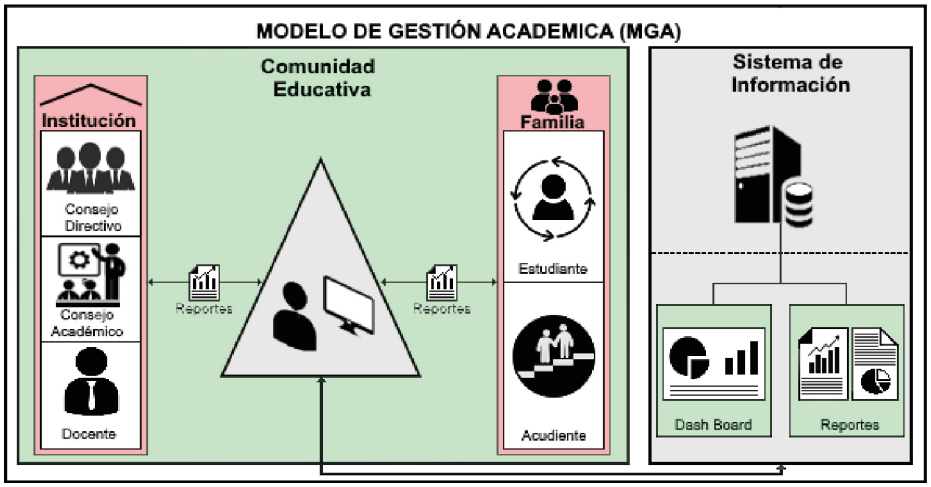


Fig. 6. *Modelo de Gestión Académica (MGA)*

V. CONCLUSIONES

El Ministerio de Educación Nacional (MEN) a partir del decreto 1075 de 2015, también conocido como Decreto Único Reglamentario del Sector de la Educación, esTable.ce que el índice de calidad para la institución educativa estará conformado por el Índice Sintético de Calidad Educativa (ISCE) y el índice de Gestión para la Calidad Educativa (IGCE), es decir que el ISCE en esencia es el instrumento por medio del cual el MEN realiza el seguimiento al progreso de cada una de las instituciones educativas del país sean estas públicas o privadas, en este contexto los MPA y MGA propuestos en este articulo son de gran valor para la gestión de la IE y el logro de la metas propuestas en especial el progreso y rendimiento de cada estudiante.

El seguimiento al progreso académico de los estudiantes por parte de padres de familia, docentes y directivos docentes de la IE educativa hoy en día es más actual y relevante que nunca, pues es a partir de este seguimiento y de la calidad del mismo en términos de oportunidad y eficiencia que se logra mejorar la formación del estudiante no solo en las dimensiones específicas de su nivel o grado, si no también en sus competencias blandas logrando una verdadera formación integral.

La metodología de seguimiento y medición del progreso académico de la estudiante implementada en el MPA le permite a la IE identificar de manera oportuna al estudiante que presenta bajo nivel de PA en una o en varias de las dimensiones sobre las cuales es valorado, y así poder esTable.cer los planes de mejoramiento necesarios para que el estudiante logre superar las deficiencias presentadas, mejorando no solo el rendimiento y progreso del estudiante sino de la IE.

La implementación de herramientas de control visual como el dashboard en la IE le aportan a la gestión académica elementos de valor en el seguimiento del RA y del PA de los estudiantes en cada asignatura o módulo de formación, y a pesar que estas herramientas tradicionalmente han sido implementadas en la industria u otro tipo de organizaciones diferentes a la IE, es importante el resultado que se puede obtener en términos de la mejora en el rendimiento de los estudiantes, la disminución de la tasa de deserción y de pérdida o repitencia de asignaturas por parte de los estudiantes.

La gestión académica se caracteriza por ser dinámica en el día a día, es decir, cada día de labores académicas es diferente uno del otro y esto se debe principalmente que quienes intervienen en el proceso de formación a diferencia de cualquier otro sector son personas y esto exige de los docentes y de los directivos docentes poder generar variedad de reportes e informes de manera oportuna y eficiente, tareas que en el MGA la realiza el PGR y el cual permite gran flexibilidad para la generación de reportes contando con el mínimo de conocimiento en informática, esta característica le aporta gran valor a la gestión académica.

BIBLIOGRAFÍA

[1] *C. A. Botero, "Cinco tendencias de la gestión educativa," Rev. Politécnica, vol. 5, no. gestión educativa, pp. 17–31, 2007.*

[2] *A. Dalila and R. Molano, "La gestión educativa: Hacia la optimización de la formación docente en la educación superior en Colombia," pp. 55–70, 2018.*

[3] *W. Meneses, S. Morillo, G. Navia, and M. Grisales, "Factores que afectan el rendimiento escolar en la institución educativa rural Las mercedes desde la perspectiva de los actores institucionales," Respositorio Inst. Univ. Manizales, pp. 433–452, 2013.*

[4] *P. Lamas, Héctor(Academia Peruana de Psicologia, Lima, "Sobre el rendimiento escolar," Propósitos y Represent., vol. 3, no. 1, pp. 313–386, 2015.*

[5] *R. Esther and L. Velandia, "Propuesta Pedagógica Direccionada desde la Gestión Académica para Mejorar el Desempeño Escolar de los Estudiantes de la Institución Educativa Distrital Agustín Fernández," pp. 1–152, 2016.*

[6] *J. Alexander and R. Gómez, "Acciones para el Seguimiento Académico: Explorando las Prácticas Docentes desde la Gestión Educativa," pp. 1–101, 2016.*

[7] *S. Cárdena and I. Tito, "Diseño de un Cuadro de Mando Integral (Parte I)," Perspectivas, vol. 23, no. 1, pp. 101–114, 2009.*

[8] *L. Rioja and S. Inform, "Metodología para el diseño de Dashboards orientado hacia el registro de evidencias en el proceso de evaluaciones institucionales," 2017.*

[9] *S. Tobon-2da-Edicion, "Formacion basada en competencias." .*

[10] *Ministerio de Educación Nacional, "Decreto 1290 de," Congr. la república Colomb., p. 50, 1994.*

[11] *S. T. Tobón, J. H. P. Prieto, and J. A. G. Fraile, El Modelo de Competencias: Cambios en la Didáctica y la Evaluación desde la Socioformación. 2010.*

[12] *M. de E. N. Presidencia de la Republica, "Decreto 1290 de 2009," no. 1290, 2001.*

[13] *W. Giraldo O, J. C. García Sabogal, F. S. Oquendo Soler, and C. D. Palechor Sepúlveda, "Plataforma para el soporte del ciclo de vida del reporte centrado en el usuario," 2017.*

Biografía Autor(es)

a. Magister en Ingeniería. Docente, Universidad del Quindio. E-mail: cesarcastelblanco@hotmail.com. Quindío, Colombia.

b. Phd Arquitecturas de Sistemas en Red y Gestión del Conocimiento. Docente Universidad del Quindío. E-mail: wjgiraldo@uniquindio.edu.co. Quindío, Colombia.

Sistema Software para la ejecución de pruebas de usabilidad bajo el enfoque de mouse tracking

David Alejandro Albornoz
Sebastián Alejandro Moncayo
Samir Ruano Hoyos
Gabriel Elías Chanchí Golondrino
Katerine Marceles

Resumen

La usabilidad se ha convertido en un atributo fundamental que define la calidad del software. De acuerdo con ISO 9241-11, el software puede utilizarse si permite al usuario alcanzar los objetivos en un contexto específico de uso con eficiencia, eficacia y satisfacción. Con respecto el atributo de eficiencia, uno de los factores que incide es la disposición de elementos en la pantalla, por lo que las pruebas de usabilidad con el seguimiento del mouse el enfoque puede proporcionar información relevante para optimizar la interacción del usuario con el sistema. En este artículo proponemos un sistema de software para soportar pruebas de usabilidad bajo el enfoque de seguimiento del mouse, que tiene como principal contribución el análisis por áreas de pantalla y por tareas de una prueba de usuario. Además, el sistema de software permite grabar un vídeo de la sesión del usuario, de forma que se realice un análisis detallado de las diferentes acciones que el usuario puede realizar durante la prueba. De este modo, el sistema de software propuesto tiene como objetivo identificar en las diferentes tareas de una prueba de usabilidad, la configuración adecuada de los elementos en pantalla, con el fin de hacer la interacción más eficiente.

Palabras clave: Seguimiento del mouse, sistema de software, usabilidad, prueba de usabilidad.

Software system for running tests usability under the approach mouse tracking

Abstract

Usability has become a fundamental attribute that defines software quality. According to ISO 9241-11, software is usable if it allows a user to achieve objectives in a specific context of use with efficiency, effectiveness and satisfaction. With respect to the efficiency attribute, one of the factors that affects is the disposition of the elements on the screen, which is why the usability tests under the mouse tracking approach can provide relevant information in order to optimize the user's interaction with the system. In this article we propose a software system to support usability tests under the mouse tracking approach, which has as its main contribution the analysis by screen zones and by tasks of a user test. Additionally, the software system allows recording a video of the user's session, in such a way that a detailed analysis of the different actions performed by the user during the test can be performed. In this way, the proposed software system aims to identify in the different tasks of a usability test, the appropriate configuration of the elements on the screen, in order to make the interaction more efficient.

Keywords: Mouse Tracking, Sofware System, Usability, Usability Test.

I. INTRODUCCIÓN

El auge de la tecnología ha traído consigo propuestas novedosas que se crean a partir de nuevos campos donde el desarrollo de software ha evidenciado avances significativos como la visión e inteligencia artificial, la robótica, el internet de las cosas, entre otras, que han permitido que la demanda de estas tecnologías aplicadas crezca exponencialmente y con ello la exigencia en la calidad del software de estos aplicativos venga en aumento. Debido a lo anterior, el concepto de usabilidad [1] se ha convertido en un asunto fundamental para para el desarrollo de la informática, puesto que define la productividad del usuario y de las empresas con respecto a un aplicativo software.

De acuerdo a la norma ISO 9241-11, la usabilidad es definida como: "el grado en que un producto puede ser usado por determinados usuarios para lograr sus propósitos con eficacia, eficiencia y satisfacción en un contexto de uso específico" [2]. Esta definición describe los tres atributos (eficiencia, eficacia y satisfacción) que se deben considerar en cuenta para estimar el grado de usabilidad de un producto software [3]. En una prueba de usuario, realizada dentro del contexto técnico de un laboratorio de usabilidad, la eficacia puede ser obtenida a partir del porcentaje de tareas ejecutadas cumplidamente por el usuario de prueba escogido. En lo que respecta a la eficiencia puede ser medida por medio del tiempo empleado para realizar las tareas esTable.cidas por el evaluador de la prueba [4]. Por su parte, la satisfacción al ser definida según la ISO 9241-11 como "ausencia de incomodidad y existencia de actitudes positivas hacia la utilización del producto" [2], está relacionada directamente con la emocionalidad del usuario durante una prueba de usabilidad [5].

Del mismo modo, uno de los indicadores que pueden ser utilizados para determinar el atributo eficiencia puede ser el número de clicks o la traza de mouse que un usuario realiza durante una determinada tarea de una prueba de usabilidad [4]. En este orden de ideas las pruebas de seguimiento de mouse o mejor conocidas como pruebas de mouse tracking permiten obtener indicadores importantes de cara a disponer los diferentes elementos en pantalla de manera más adecuada, con el fin de hacer más eficiente la interacción del usuario con el sistema software. Es así como a partir de la traza de mouse es posible detectar los sectores de la interfaz de un determinado software que pueden ser rediseñados o movidos a zonas de mayor jerarquía [6][7].

Aunque se ha evidenciado la existencia de algunas herramientas académicas para la ejecución de pruebas de seguimiento de mouse, estas en su mayoría solo permiten la generación de la traza del mouse en pantalla, pero no posibilitan el análisis discriminado por zonas de pantalla, o el análisis de la traza en una determinada tarea de un test de usuario desarrollado en un laboratorio de usabilidad. De esta manera en este artículo se presenta como aporte un sistema software para la ejecución de pruebas de usabilidad bajo el enfoque de mouse tracking, el cual además de realizar la captura de la traza de mouse en pantalla, permite el análisis de diferentes zonas de interés durante las diferentes tareas de una prueba de usabilidad, así como la grabación en formato de video de la sesión. En este orden de ideas, a través del análisis por tareas y zonas el sistema software propuesto pretende ser de ayuda para los coodinadores de la prueba en cuanto a la detección de opciones de frecuente uso que están ubicadas en zonas de baja jerarquía. Es así como con el uso del sistema software propuesto, se esperan obtener recomendaciones puntuales en cuanto a la mejor disposición de los elementos en pantalla, con el fin de incrementar la eficiencia del software evaluado.

El sistema software propuesto fue desarrollado en el lenguaje Java, haciendo uso de las APIs EasyCapture, OpenCSV y JFreeChart, las cuales permiten respectivamente

la grabación de la sesión de video, el almacenamiento de las coordenadas y el tiempo en un archivo CSV, y finalmente la generación de gráficos estadísticos a partir de los datos capturados en la prueba de usabilidad. El resto del artículo está organizado de la siguiente forma: en la sección 2 se presentan conceptos relevantes que se tuvieron en cuenta en la presente investigación. En la sección tres, se describen un conjunto de herramientas que se consideraron para el diseño del sistema software propuesto. En la sección cuatro se muestra la estructura funcional del sistema software construido y las interfaces finales del mismo. Finalmente, en la sección cinco se presentan las conclusiones y trabajos futuros derivadas de esta investigación.

II. MARCO CONCEPTUAL

En esta sección se presentan los principales conceptos que se tuvieron en cuenta para el desarrollo de este trabajo. Dentro de estos se encuentran: laboratorio de usabilidad, prueba de usabilidad y atributo eficiencia.

2.1 Laboratorio de usabilidad.

Un laboratorio de usabilidad es un espacio físico compuesto generalmente de dos áreas, una para el usuario de prueba y otro para los evaluadores, en el cual se realizan evaluaciones de usabilidad y de accesibilidad a productos, sistemas y dispositivos, por medio de software y equipos hardware especializados en el monitoreo y registro delaccionar del usuario con el producto a evaluar. Además, cuenta con herramientas que permiten obtener indicadores de los atributos de eficiencia, eficacia y satisfacción [5][8].

2.2 Prueba de usabilidad

Es una técnica en el diseño y desarrollo de aplicaciones software o productos, por medio de la cual se evalúa el grado de uso de estos, mediante la inspección in situ de la interacción del usuario de prueba con estos aplicativos. Como características principales se tienen que los resultados de esta prueba se van evidenciado de manera inmediata, el uso de espacios físicos aislados y el monitoreo constante con equipos técnicos especializados [9]. Dentro de las pruebas de usabilidad, se destaca el enfoque de seguimiento de mouse o mouse tracking, a través del cual es posible determinar elementos de la interfaz que no están distribuidos de manera adecuada y cuya reubicación podría contribuir a mejorar la eficiencia con la que un usuario realiza una determinada tarea dentro del software.

2.3 Atributo eficiencia

La norma ISO 9241-11 [2], define el atributo eficiencia como: "relación entre los recursos empleados y precisión y grado de consecusión con que los usuarios logran

objetivos específicos". Dentro de algunas de las métricas empleadas para determinar el atributo satisfacción se destacan: tiempo empleado en completar una tarea, número de teclas presionadas por tarea, tiempo transcurrido en cada pantalla, eficiencia relativa en comparación con un usuario experto [4]. En este orden de ideas, una prueba de seguimiento de mouse o mouse tracking tiene como fin aportar en mejorar los indicadores del atributo satisfacción, en el sentido que a partir de la traza de mouse capturada de la interacción del usuario con un software determinado, permite identificar posibles componentes que no están distribuidos en pantalla de manera adecuada y que hacen que el usuario tarde más tiempo en ejecutar una tarea determinada.

III. HERRAMIENTAS EXPLORADAS

En esta sección se presentan un conjunto de herramientas que permiten dar soporte a una prueba de usabilidad bajo el enfoque de mouse tracking. Dentro de estas herramientas se destacan: Mouseflow, UserTrack, IoGraph.

3.1 Mouseflow

Es una herramienta para el análisis de la traza de mouse en aplicaciones web, la cual actua en segundo plano, permitiendo hacer grabaciones de pantalla, grabaciones del movimiento del mouse y generando estadísticas sobre los puntos donde el usuario centra su atención al navegar por una portal web. Lo anterior se realiza con el fin de redistribuir o cambiar diferentes elementos de un portal web como pueden ser: el color, el texto, los componentes multimedia, la ubicación de las diferentes opciones, en aras de mejorar la experiencia de un usuario en el sitio y hacer más eficiente la interacción. La herramienta incluye cinco componentes principales: repetición de sesión (grabaciones), mapas de calor (clic, movimiento, desplazamiento, atención y geografía), embudos, análisis de forma y comentarios de los usuarios. Aunque mouse flow permite hacer seguimiento sobre los puntos de interés en los que centra el usuario la atención durante la interacción, la herramienta no incluye el análisis de zonas en pantalla, ni la discriminación de la traza del mouse por tareas [10].

3.2 UserTrack

Es una herramienta analítica que se encuentra actualmente en desarrollo activo, la cual está diseñada para contribuir con la mejora en el diseño de portales web, al definir a partir de la traza del ratón posibles componentes de la interfaz que deben ser reubicados. En este orden de ideas, esta herramienta además de posibilitar el control de los visitantes a un sitio web, permite el registro mediante mapas de calor la traza del mouse de un usuario durante la interacción con un sitio web. De igual modo, esta herramienta se encarga de generar videos sobre la interacción de un usuario con la aplicación web. Los mapas de calor son generados haciendo uso de tres colores: verde, amarillo y rojo, dependiendo de la traza y del tiempo que el mouse permanece en una determinada ubicación de la pantalla. El color verde es usado cuando el tiempo que

permanece el mouse en una posición es poco, el color rojo es usado cuando el ratón permanece mayor tiempo en una posición, mientras que el color amarillo corresponde a un tiempo intermedio. Aunque UserTrack arroja datos de interés sobre la traza del mouse y los lugares donde el usuario centra su atención, no permite el análisis de la traza del mouse por tareas o zonas por tareas de una prueba de usabilidad [11].

3.3 IoGraph

IOGraph es una herramienta gratuita de escritorio desarrollada en Java, la cual se ejecuta en segundo plano y permite la captura y representación gráfica de la traza del mouse durante una prueba de usabilidad. Esta herramienta permite adicionalmente capturar una imagen del fondo y sobreponer sobre esta la traza de mouse capturada, para posteriormente posibilitar los resultados de la prueba en formatos de imagen. Esta herramienta también incluye la funcionalidad de dibujar puntos de interés en los cuales el mouse queda suspendido. Aunque esta herramienta es intuitiva y fácil de manejar, no permite el análisis de la traza del mouse por zonas de interés, ni tampoco el análisis por tareas de un test de usuario [12].

3.4 ClickTale

Esta es una herramienta propietaria que permite registrar las diferentes interacciones de un usuario con una aplicación web, mediante la captura de la traza de mouse. Para realizar lo anterior, esta herramienta permite la generación de mapas de calor y el registro de la interacción del usuario a través de videos. Adicionalmente, esta herramienta tiene asociadas funciones de análsis, las cuales generan un informe acerca de los posibles componentes que pueden generar que la interacción del usuario con el portal web no sea adecuada. Aunque la herramienta permite la generación de videos con la interacción, no posibilita el estudio automático de la traza de mouse durante las diferentes tareas de una prueba de usabilidad. Así mismo las herramientas abordadas no permiten el análisis de la traza del mouse por zonas de interés en pantalla. Esos elementos, son importantes de cara al diseño e implementación del sistema software para el seguimiento de mouse propuesto [13].

IV. SISTEMA SOFTWARE PROPUESTO

En esta sección se describe el diagrama de flujo del sistema, el diagrama de bloques del sistema y las interfaces finales del mismo.

4.1 Diagrama de flujo del sistema

A partir de las diferentes características de un test de usuario desarrollado en un laboratorio de usabilidad, en la figura 1 se presenta el diagrama de flujo del

sistema software construido para el seguimiento de mouse. En primera instancia una vez ejecutado el sistema software, el coordinador de la prueba debe tomar la decisión si registrar una nueva sesión o analizar datos que previamente han sido almacenados en sesiones anteriores. En caso que escoja la opción de registrar una nueva sesión, ingresa los datos de la prueba tales como, el nombre de la sesión, la fecha, el nombre de los usuarios entre otros datos básicos. A continuación, una vez registrados los datos de la sesión y de manera paralela al inicio de la prueba de usabilidad, el sistema software procede a almacenar de manera continua las coordenadas x y y del puntero del mouse, así como el tiempo trascurrido desde el inicio de la prueba, haciendo uso para ello de estructuras de datos temporales como los provistos por la clase Vector de Java. Al tiempo que las coordenadas del puntero y el tiempo van siendo registrados temporalmente, el sistema software se encarga de iniciar la grabación del video que captura las diferentes interacciones del usuario en cuestión con el software evaluado, usando para ello la librería EasyCapture. Una vez termina el test de usuario por parte del usuario en cuestión, el sistema se encarga de volcar los datos almacenados temporalmente en un archivo de texto plano tipo CSV. Se escogió este tipo de formato con el fin de dar la posibilidad también al coordinador de la prueba de revisar los datos en una hoja de cálculo.

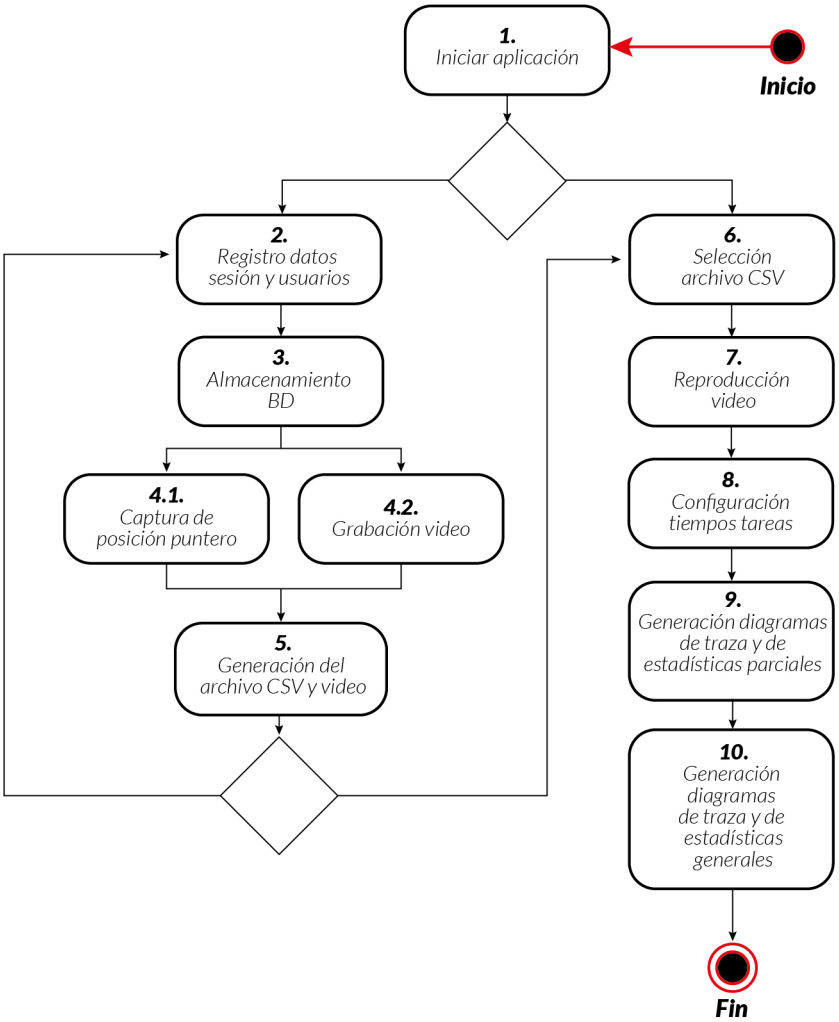


Fig. 1. *Diagrama de flujo del sistema software.*

Una vez almacenados los datos en el archivo de texto plano tipo CSV, se puede proceder a la grabación de una nueva sesión de usuario (repetición de la fase anterior) o se puede dar continuidad a la fase de análisis de los datos capturados y almacenados en el archivo de texto plano CSV. Del mismo modo el coordinador de la prueba puede proceder con la reproducción del video generado en la sesión con cada uno de los usuarios participantes. Para el análisis de las zonas de interés en las diferentes tareas de una prueba, es importante que el coordinador de la misma haya llevado el registro de la duración en segundos de cada una de las tareas desarrolladas por los usuarios durante la prueba, de tal modo que una vez proporcionados los tiempos de las tareas de cada uno de los usuarios, sea posible hacer un análisis por zona de interés de cada una de las tareas y de la prueba en general.

4.2 Diagrama funcional del sistema software

En la figura 2 se presenta el diagrama funcional del sistema software propuesto, diferenciando las funciones que son realizadas a partir de los datos capturados al usuario de prueba y las tareas que realiza el coordinador o evaluador de la prueba de usabilidad. Lo anterior obedece a que el sistema software está pensado para ser ejecutado en el contexto de los tests de usuario desarrollados en el laboratorio de usabilidad, en los cuales mientras un usuario de prueba realiza un conjunto de tareas sobre un software determinado, el evaluador o coordinador de la prueba se encarga de observar el comportamiento del usuario y verificar el cumplimiento de las diferentes tareas.

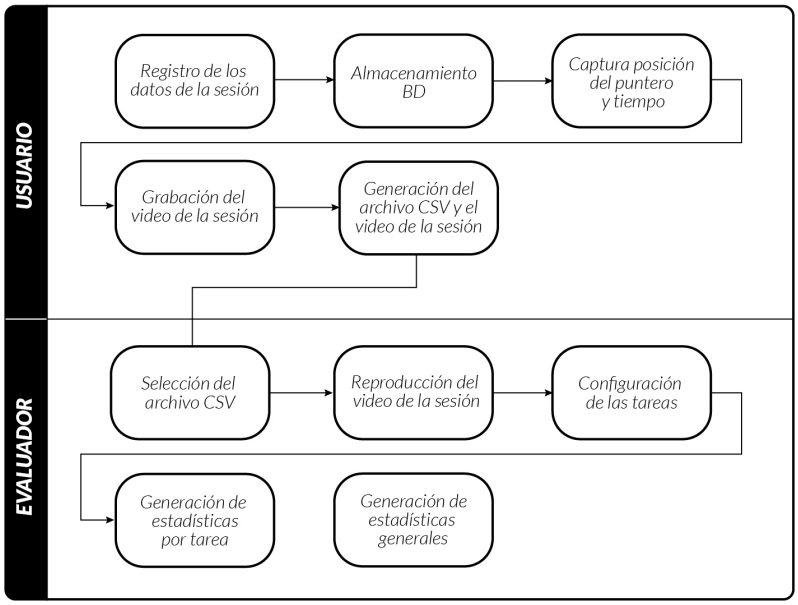


Fig. 2. *Diagrama funcional del sistema software.*

En ese orden de ideas, el sistema en primera instancia permite registrar los datos de la prueba y almacenarlos en una base de datos. A continuación, a partir de las tareas que realiza el usuario dentro del laboratorio de usabilidad, el sistema software se encarga de capturar en tiempo real las posiciones del puntero y el tiempo asociado a dichas capturas. Así mismo, mientras el usuario está realizando la interacción, el sistema

software se encarga de capturar un video de la sesión, el cual va a permitirle al coordinador de la prueba realizar un análisis detallado del comportamiento del usuario en una determinada tarea o a nivel general. Una vez terminadas las diferentes tareas de la prueba de usabilidad, el sistema genera un archivo csv que contiene información de la traza del ratón y el tiempo en el cada pareja de coordenadas han sido capturadas.

Cuando la prueba de usabilidad termina, el coordinador de la misma es el encargado de analizar los resultados de esta, cargando en primera instancia el archivo CSV generado en la prueba. Posteriormente el coordinador de la prueba puede acceder a los diferentes videos generados, con el fin de verificar el tiempo de duración de cada una de las tareas desarrolladas por un determinado usuario. A partir del diligenciamiento de los tiempos de las diferentes tareas en el sistema software, es posible visualizar el análisis por zonas en cada tarea y de manera general. Mediante estos resultados es posible tomar desiciones relacionadas con la reajuste en la ubicación de los componentes de la interfaz hacia zonas de mayor jerarquía en pantalla, buscando que el software evaluado permita una mayor eficiencia en la ejecución de las tareas.

4.3 Interfaz final del sistema software

A partir de las funcionalidades descritas en la sección 4.2 y 4.3, en la figura 3 se presenta la interfaz gráfica principal del sistema software para el seguimiento de la traza de mouse en pruebas de usabilidad. En la izquierda de la interfaz se muestran las diferentes opciones del sistema software, las cuales permiten la configuración de las sesiones, la captura de la traza de mouse, la configuración de las tareas y el análisis gráfico de la traza de mouse por tareas y de manera general. Por su parte en la parte derecha por defecto se presenta la interfaz que permite cargar un archivo CSV de una sesión previa, así como el video asociado a dicha sesión. Del mismo modo es posible digitar la duración de las diferentes tareas de la prueba asociadas a la sesión cargada, mediante las cuales es posible hacer el análisis de la traza de mouse por zonas de pantalla.

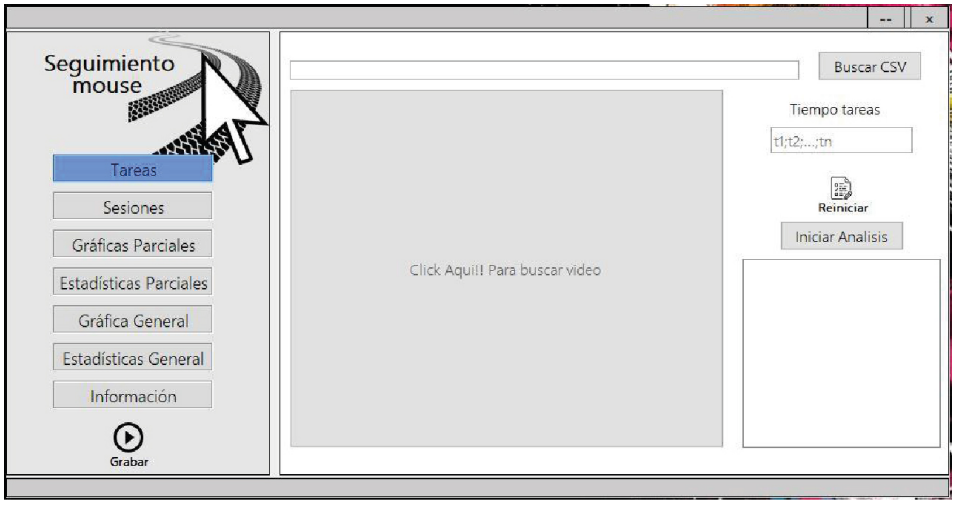


Fig. 3. *Diagrama funcional del sistema software.*

De manera posterior a la carga del archivo CSV y del registro de la duración de las diferentes tareas de la prueba, se presiona en el botón "Iniciar Análisis", lo cual permite visualizar en la opción "Gráficas Parciales", el análisis de la traza de mouse en las diferentes tareas de la prueba de usabilidad, tal como se muestra en la figura 4. Lo anterior permite al coordinador de la prueba identificar si la disposición de ciertos elementos de la interfaz involucrados en cada una de las tareas es la adecuada.

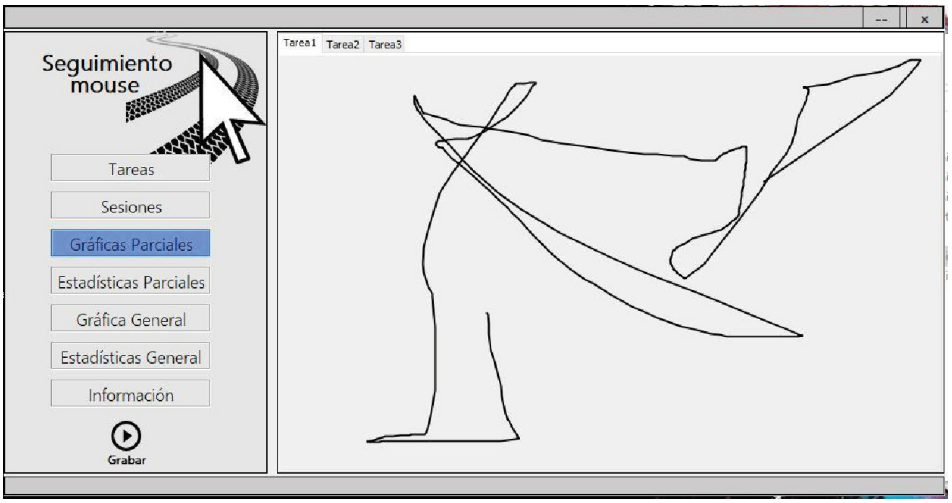


Fig. 4. *Diagrama funcional del sistema software.*

Finalmente, según se muestra en la figura 5, es posible realizar un análisis por zonas de interés en pantalla, de acuerdo a las diferentes tareas de una prueba de usabilidad. Esto permite visualizar las zonas más utilizadas en una tarea de la prueba, para posteriormente identificar si ciertos elementos de la pantalla están dispuestos de manera correcta en una tarea específica de la prueba.

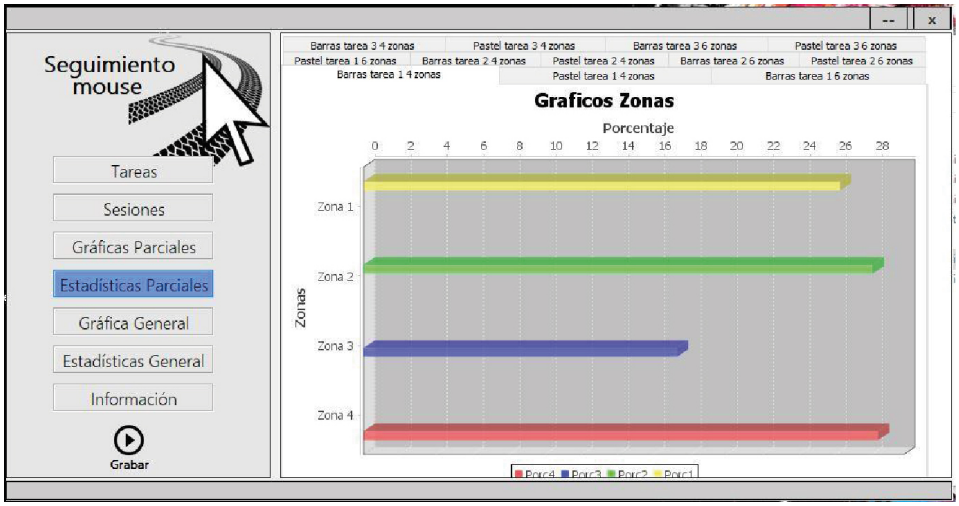


Fig. 5. *Gráfico de zonas de interés.*

V. CONCLUSIONES

El sistema software presentado en esta investigación fue construido de a acuerdo a las características de un test de usuario convencional desarrollado en un laboratorio de usabilidad, de tal modo que permita hacer el análisis de las trazas de mouse por tarea de la prueba. En este mismo sentido el sistema software tiene como aporte el uso de diferentes zonas de interés, lo cual puede ayudar a tomar decisiones sobre posibles áreas de la pantalla que no se están usando de manera adecuada. Finalmente el sistema software propuesto también presenta como aporte la generación de un video con la interacción realizada por el usuario durante la prueba, el cual le permite al coordinador de la prueba llevar un control sobre el tiempo de las tareas realizadas por el usuario.

El sistema software propuesto almacena los datos de la interacción del usuario con el software a evaluar (coordenadas del mouse y traza de tiempo) en un archivo plano tipo CSV y se encarga de procesar dichos datos de acuerdo a la duración de las tareas de la prueba de usabilidad. Lo anterior posibilita presentar y filtrar de manera gráfica la traza de mouse en las diferentes tareas, lo cual pretende ser de ayuda en la toma de decisiones del coordinador de la prueba sobre los elementos de la interfaz que pueden ser reacomodados para hacer más eficiente el software evaluado con respecto a las diferentes tareas.

El sistema software propuesto en este artículo pretende ser de apoyo en cuanto a la ejecución de pruebas de usabilidad, en el sentido que permite obtener datos complementarios que pueden generar recomendaciones acerca de la adecuada disposición de los elementos de la interfaz. Esto gracias a las ventajas que provee el sistema software con respecto al análisis por zonas y por tareas de una prueba de usabilidad.

A nivel de las tecnologías usadas para la construcción del sistema software propuesto, es importante resaltar que el lenguaje Java tiene como ventaja el hecho de contar con un conjunto de APIs que posibilitan la captura de la traza de mouse, la grabación de videos de la interacción y el almacenamiento en archivos CSV. En este sentido a diferencia de otros lenguajes, Java permite facilidad en la integración de dichas librerías al proyecto, sin requerir instalaciones adicionales. En este sentido las tecnologías empleadas pretenden servir de referencia para la construcción de proyectos para el estudio de la traza de mouse en diferentes contextos de aplicación.

La importancia de las pruebas de usabilidad bajo el enfoque de seguimiento de mouse o mouse tracking, es que mediante la mejor distribución de los elementos en pantalla, se pretende mejorar los indicadores del atributo eficiencia, en el sentido que un usuario pueda tardar menos tiempo en ejecutar una tarea determinada dentro del software evaluado.

Como trabajos futuro derivado de la presente investigación, se espera complementar el sistema software en el sentido que se permita discriminar de manera gráfica el tiempo que el puntero del ratón permaneces suspendido en la interfaz. En este mismo sentido, se pretende conducir estudios de computación afectiva a partir de la traza del mouse capturada durante una prueba de usabilidad, con el fin de contribuir con la estimación del atributo satisfacción.

REFERENCES

[1] *Sánchez, W.: La usabilidad en Ingeniería de Software: definición y características. Revista de Ingeniería e Innovación de la Facultad de Ingeniería - Universidad Don Bosco 1(2), 7-21 (2011).*

[2] *ISO 9241-11, https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en*

[3] *Mascheroni, M., Greiner, C., Petris, R., Dapozo, G., Estayno M.: Calidad de software e Ingeniería de Usabilidad. En: XIV Workshop de Investigadores en Ciencias de la Computación. Universidad Nacional de Misiones (UNM), La Plata-Argentina (2012).*

[4] *Enriquez J. G., Casas S.I.: Usabilidad en aplicaciones Móviles. Revista Informe Científico Técnico UNPA 5(2), 25-47 (2013).*

[5] *Delgado D., Girón D., Chanchí G., Márceles K.: Propuesta de una herramienta para la estimación de la satisfacción en pruebas de usuario, a partir del análisis de expresión facial 19(2), 6-15 (2018).*

[6] *Hehman, E., Stolier, R., Freeman, J.: Advanced mouse-tracking analytic techniques for enhancing psychological science 18(3), 384-401 (2014).*

[7] *Freeman, J., Ambady, N.: MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method 42(1), 226-241 (2010).*

[8] *Perurena, L., Moráguez, M.: Usabilidad de los sitios Web, los métodos y las técnicas para la evaluación. Revista Cubana de Información en Ciencias de la Salud 24(2). 176-194 (2013).*

[9] *Calvo-Fernandez A., Ortega, S., Valls, A.: Métodos de evaluación con usuarios. Editorial Universidad de Cataluña, Cataluña-España (2004).*

[10] *Mouseflow, https://mouseflow.com*

[11] *UserTrack, https://www.usertrack.net/*

[12] *IoGraph, http://iographica.com/*

[13] *ClickTale, https://www.clicktale.com/*

Un docente, enseña o educa: la Ingeniería con pensamiento sistémico

Gloria Amparo Thomé Ponce
Manuel Ernesto Bolaños González
Nelson Antonio Jaramillo Enríquez

Resumen

En el presente artículo, sírvase referirse a los conceptos; esta es la base para el conocimiento nueva producción (Anderson, 1984; Anderson y Pearson, 1984), los conceptos son importantes frente al contenido de programación del plan de estudios de ingeniería sobre el aprendizaje y proceso de enseñanza. Los conceptos relevantes marcarían el camino hacia una ingeniería generación pensando en la educación como investigación constante; los conceptos son soporte para las actividades para el desarrollo de habilidades. Hay dos instancias para la construcción de los conceptos, el primero se utiliza "mentefacto" y el segundo la construcción piensa en aptitud en ingeniería puede ver presente para la teoría general del sistema correcta aplicar todos los procesos que tienen resultados cada uno muy complejo para resolver y comprender.

Palabras clave: Conceptos, educación, aprendizaje, pensamiento sistémico.

A teacher, teaches or educates: Engineering with thought systemic

Abstract

In the present article, do refer to the concepts; this is the base for the knowledge new production (Anderson, 1984; Anderson y Pearson, 1984), the concepts are important front to the programing contain of the engineering studios plan on learning and teaching process. The relevant concepts would check the way into an engineering generation thinking the education as constant research; the concepts are support to the activities for abilities development. There are two instances for construction of the concepts, the first is used "mentefacto" and second the construction thinks aptitude in engineering can see present for System General Theory correct apply all process that have results each very complex for solver and understand.

Keywords: Concepts, education, learning, system thinking.

I. INTRODUCCIÓN

"El sistema educativo se basa en un principio que es falso: que el profesor tiene la verdad"
Javier Sala i Martin

¿Qué es eso tan difícil que un estudiante de ingeniería no pueda asumir por el hecho de ser estudiante en entornos complejos de conceptos de orden sistémico? ¿Cuán importante es conocer los conceptos, y aún más, cual es la fórmula para aplicarlos en una situación problémica? Los entornos tabulan conceptos de la generalidad y se van depurando a medida que hay experiencia hacia lo particular, es el caso de la TGS. El ingeniero en formación requiere un nuevo esquema de orientación de las materias, su escenario es la universidad, pero debe demostrar dentro y fuera de la institución su desempeño como profesional. La pregunta surge entonces, ¿Qué recibe, qué gestiona, qué desarrolla el estudiante en clase? Pocas veces recibe de los docentes estrategias para desarrollar habilidades de pensamiento [2], además, las evaluaciones se califican con escalas de 0 a 10 o de 0 a 5 o con letras indicando que alguno de estos símbolos demuestra lo que el ingeniero aprende de una disciplina pero no se preocupan por evaluar el incremento del desarrollo o la evolución de habilidades para resolver situaciones problémicas. Esto es más general, merece otra forma de valoración y su percepción no es tan fácil, pero es necesaria.

1.1 La importancia de los conceptos

"Los analfabetos del siglo XXI no serán los que no sepan leer o escribir, sino los que no pueden aprender, desaprender y reaprender"

declara el filósofo francés Edgar Morín

La primera habilidad del pensamiento es crear los conceptos en nuestra mente, un ejemplo palpable es, cuando vemos a lo largo del camino a alguien; a la distancia para distinguir pensamos que es de nuestra especie, un ser humano, __ conforme se acerca vemos que es un hombre de género masculino__, y finalmente estando suficientemente cerca descubrimos que es Santiago, pero que no lo distinguimos desde el principio no solo por la distancia sino porque viene vestido con frac. Es necesario trabajar en clase con los conceptos no tanto con los contenidos. Además esos conceptos clasificarlos. En este caso Santiago tiene tres conceptos: Ser humano, hombre y la persona con un nombre, es real, es válido, por tanto, con este ejemplo notamos que nuestro cerebro clasifica los conceptos jerárquicamente.

En ingeniería los conceptos son las bases para la construcción de pensamiento, se espera en realidad que sea pensamiento nuevo, que haya nuevo conocimiento. Esto significa que el docente debe prepararse para saber evaluar no confrontando las respuestas de los estudiantes con lo que se da por sentado, sino el proceso de reflexión y posterior aplicación que hace el estudiante en la complejidad de los temas conociendo claramente los conceptos. Así entonces, el proceso de aprendizaje es propiamente del estudiante y el proceso de enseñanza es de base, de pauta, de orientación por parte del docente. Un ejemplo:

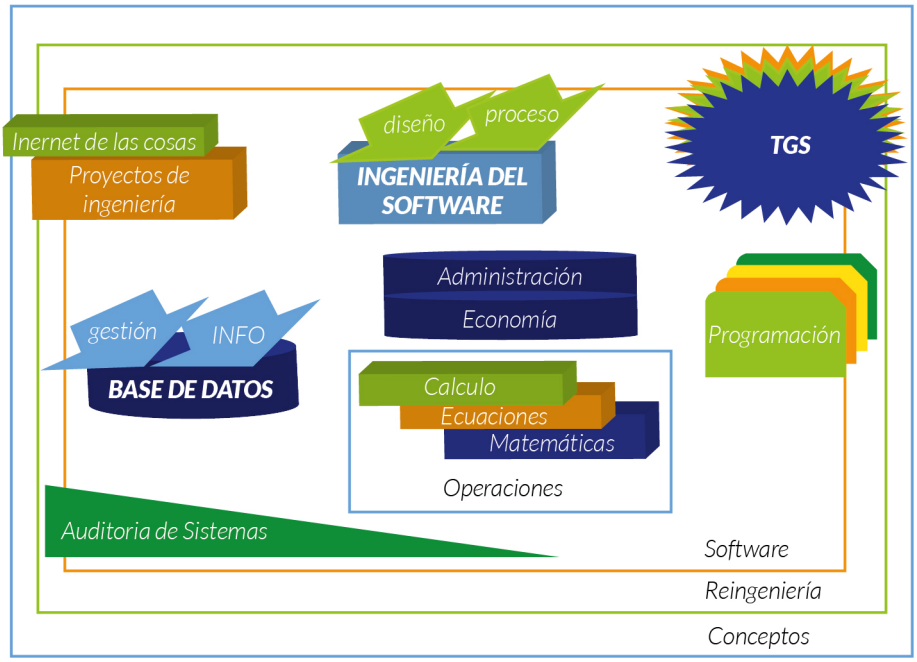


Fig. 1. *Representación gráfica de sugerencias en algunas asignaturas con enfoque de conceptos*

Fuente: Elaboración propia

En la figura 1. La transversalidad es evidente dentro del rectángulo naranja, luego algunas de las asignaturas pueden verse comprometidas, total o parcialmente, en procesos de Reingeniería en el rectángulo verde, más la totalidad de estas materias correspondientes a un pensum de Ingeniería de Sistemas están dentro de un marco que es ineludible: el marco purpura representa los conceptos.

II. APTITUDES DE PENSAMIENTO

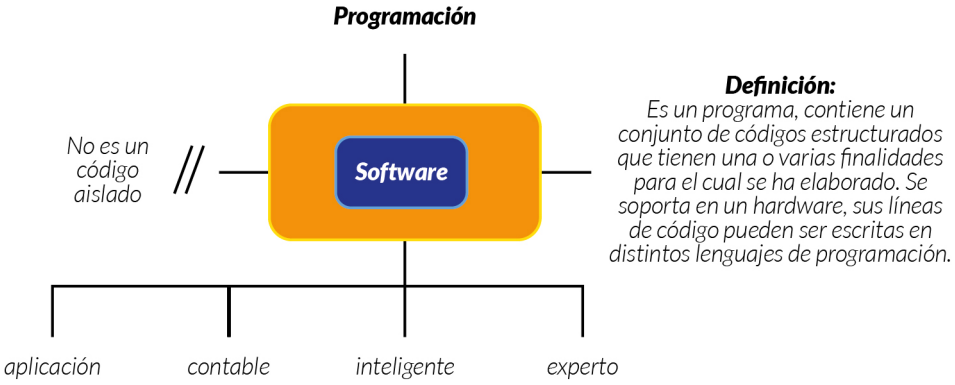
El docente de ingeniería ciertamente sigue un programa, esta labor no está mal, pero tampoco está bien que el estudiante reproduzca ese contenido. Eso no es educar. La respuesta es simple. Es un ingeniero, creador de ideas. Generador de conocimiento. "Existe consenso entre las autoridades y especialistas ligados al quehacer científico, respecto de la necesidad de generar cambios en la enseñanza de las ciencias" (Bromley, 1989; de Hart, 1991) [3]. Para corroborar con un ejemplo; imagine que hay en un Table.ro las siguientes palabras:

Datos -información- bases de datos -base de conocimientos- sistemas inteligentes- sistemas expertos- el tema para confrontar no es contenidos, son los conceptos. Consecuentemente que concepto podría crear el ingeniero en el ejercicio de la triangulación de tres de los términos del Table.ro:

En la indagación del concepto, el docente se soporta en distintas fuentes, la malla curricular, un libro de texto, un software, con una consulta amplia este conocimiento, generalmente. Es entonces tiempo de considerar: ¿El currículo en las facultades de

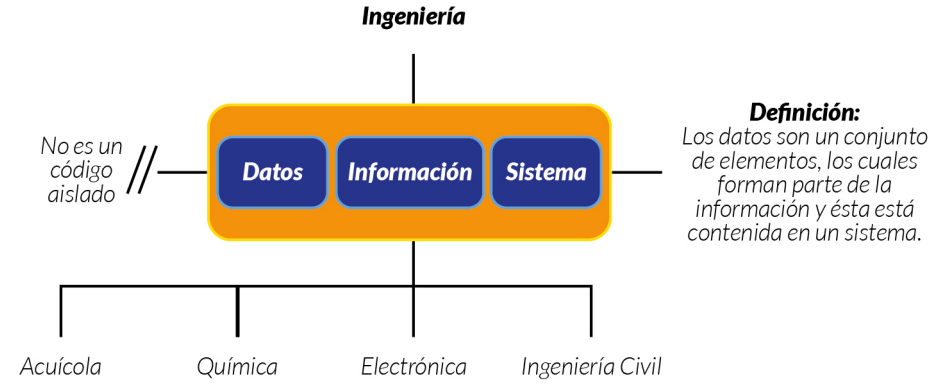
ingeniería mueve los cimientos para intencionar un nuevo concepto? ¿Se propende por el nuevo conocimiento como criterio para concertar evaluación?

En la figura 2 se ilustra un concepto general. Una de las estrategias de representación de pensamiento es el mentefacto. Esta herramienta es útil para descifrar conceptos; generalmente admite un término y permite organizar la información acerca del concepto.



Realmente no es un código suelto, aislado, sin sentido, el software tiene una definición en palabras sencillas pero con base científica) podríamos agregar el autor, la enciclopedia o enlace en internet. La subcategoría nos indica algunas clases de programas y finalmente la supra categoría agrupa todos los aspectos que se recogen con esta herramienta en un concepto más complejo: programación.

En el siguiente caso, se propone que el estudiante desarrolle con ayuda de la herramienta un concepto de tres términos.



La supra categoría es en este caso Ingeniería agrupa las subcategorías (Acuícola, Química, Electrónica, Ingeniería Civil). Que son ramas de la ingeniería. En el costado izquierdo, se dice lo que no es, un elemento suelto no puede ser, es un complejo concepto y se da un paso hacia adelante cuando el estudiante relaciona estos tres términos y los define a partir del concepto de jerarquización.

¿Realmente se desarrolla en el ingeniero en formación un lenguaje construido en el ámbito que han entretejido por años y años de experiencia los docentes en torno a

las teorías puestas en el escenario del aprendizaje? ¿De la totalidad de la población de un curso, cuántos se gradúan con un bagaje de "herramientas" desarrolladas en el ejercicio de practicar habilidades de pensamiento para resolver problemas? Si se insiste en tener en el currículo TGS, ¿Es correcto que la Teoría General de Sistemas, base de la construcción del conocimiento en ingeniería se siga compartiendo en específica teoría? y en la práctica pedagógica, ¿Cuál es la representación mental que se hace el estudiante de Ingeniería con un docente que en clase guía la lectura de un documento que data de 2004, y que en 2018 es claramente superficial y está su esencia bastante bien enfocada en la retórica? Realmente fue válida en un tiempo, perdió vigencia pese a que los conceptos se mantienen. En las universidades latinoamericanas debe darse este paso sin miedo, entre los fundamentos para promover el cambio, los actuales diseños curriculares, en muchas de ellas no reflejan la esencia de la ingeniería y tampoco sus visiones están realmente vinculadas con las necesidades de la sociedad y del hombre contemporáneo (Gardner, 1983).

De acuerdo con Pressman *"Los principios de diseño esTable.cen una filosofía general que guía el trabajo de diseño que debe ejecutarse. Deben entenderse los conceptos de diseño antes de aplicar la mecánica de éste, y la práctica del diseño en sí lleva a la creación de distintas representaciones del software que sirve como guía para la actividad de construcción que siga"* [4].

Primero visión y luego base científica para poner piso en el camino a lo largo del cual se ha de descubrir. Los docentes tendemos a asumir que el estudiante espera a que se le enseñe, tradicionalmente los estudiantes llegan a recibir, pocas veces, muy pocas veces llegan a descubrir.

2.1 Es menester desarrollar habilidades metacognitivas

¿Qué es eso tan grande que el docente enseña en su clase a un puñado de mentes en la autopista de la voluntad de aprendizaje?

Estimado docente, si usted cree que su clase es maravillosa y que sus estudiantes producen, además sus calificaciones guardan rangos de promedios noTable.s en la región o siempre han estado sobre los altos niveles, ¡felicitaciones hace un buen trabajo! __ Ahora piense en las aptitudes de pensamiento que los estudiantes desarrollaron al enfrentar los problemas de conocimiento__ se nota tendencia para poner en práctica el resolver: con o mediante la base teórica, los conceptos, con el docente, con los apuntes de clase, entre otros. ¿Cuántos de ellos se inclinan en la construcción de conceptos, toman la teoría como base del razonamiento de las posibilidades, identifican los conocimientos específicos que marcan los puntos de construcción de un mapa de soluciones heurístico?, estas líneas no tiene sino un fin: incomodar la percepción de la categorización de "buena práctica pedagógica" la que es repetitiva y tradicional.

Volver a la visión es una manera de enfrentar la cátedra, es verdad, puede ser lo suficientemente buena para activar al estudiante de ingeniería, para activarlo, no para desarrollar aptitudes con respecto del pensamiento, del pensamiento sistémico, de los mil caminos con comportamiento heurístico, de pensamiento divergente, de pensamiento algorítmico, para resolver situaciones reales. Replantear la evaluación es consecuente. Son los procesos, los que más peso tienen en la resolución de los problemas en un nivel más alto que los resultados.

2.2 Caso de estudio: Ingenieros en potencia y resultados del proceso

Se presenta a continuación un ilustrativo ejemplo de la visión en ingeniería. Las siguientes evidencias forman parte de los resultados de un proceso investigativo que buscó demostrar que jóvenes y señoritas de secundaria de grados séptimo, octavo y noveno tuvieron inmersa la ingeniería en su ser, que fueron creadores de nuevos caminos para solucionar problemas cotidianos, en este caso, problemáticas de su entorno, apoyados en herramientas de ofimática y una guía metodológica clara, el proceso en términos de tiempo inicialmente duró tres meses y para consolidar estas demostraciones se extendió a un año. Fueron dos escenarios: Robótica y Biónica. La población total de estudiantes fue de 110, con edades entre 12 a 16 años. El 100% de la población objeto participó del proceso con entusiasmo en la asignatura de Tecnología e Informática. Sus trabajos fueron expuestos y con esto se fijó una nueva manera de cursar la asignatura con estrategias pedagógicas centradas en el desarrollo de habilidades del pensamiento, además los diversos talleres contribuyeron a que se identifique la oportunidad de producción de nuevo conocimiento.

Base científica para Robótica

La robótica como herramienta para la educación de Orazio Miglino, Henrik Hautop Lund, Maurizio Cardaci Impacto de las TIC en los aprendizajes de los estudiantes. Estado del arte de Magdalena Claro. Ser competente en tecnología: ¡una necesidad para el desarrollo! MEN

Software y hardware

Ofimática: Graficador, procesador de texto, equipos de cómputo de Computadores para Educar.

Fig. 2. *Guía mediante textos, fotografías y videos sobre avances tecnológicos en materia de Robótica.*

Fig. 3. *Socializaciones de los primeros diseños sobre papel, el boceto debió entregarse en lápiz y papel bond tamaño carta.*

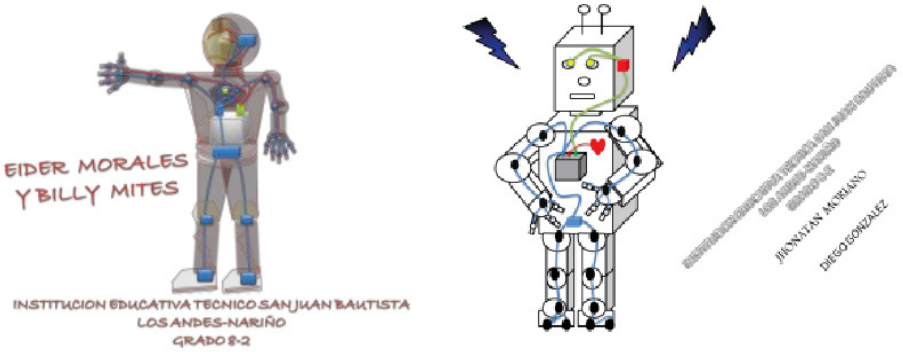


Fig. 4. *Diseño de soluciones pasan del papel a ser diseño asistido por computador, los propósitos de las creaciones fueron seleccionados libremente por los estudiantes.*

Fig. 5. *Socialización de Los resultados, los diseños fueron trasladados a prototipos elaborados con materiales reutilizables.*

Base científica para Biónica

Impacto de las TIC en los aprendizajes de los estudiantes. Estado del arte de Magdalena Claro Ser competente en tecnología: ¡una necesidad para el desarrollo! MEN Biónica, lo que nos enseña la naturaleza MEN

Guía mediante textos, fotografías, talleres y videos sobre avances tecnológicos en materia de Biónica.

Fig. 6. *Maquinas voladoras: Da Vinci, Oechiem.*

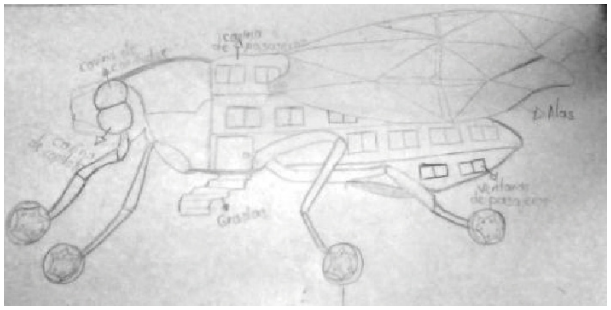


Fig. 7. *Socializaciones de los primeros diseños sobre papel, el boceto debió entregarse en lápiz y papel bond tamaño carta.*

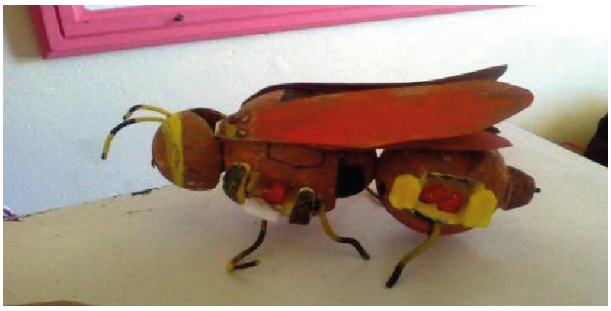


Fig. 8. *Exposición de los diseños elaborados en material de reutilizable.*

Fig. 9. *Exposición de trabajos ante padres de familia y autoridades de educación y de la municipalidad.*

Fig. 10. *Culminación del proceso de Biónica y Robótica para ingenieros en potencia.*

Los talleres iniciaron con la consulta de los temas: maquinas al servicio del ser humano en Robótica y en el caso de Biónica, el estudio de las características de los insectos y su comportamiento en el vuelo. Los diseños sobre papel tuvieron lugar aproximadamente un mes después de haber seleccionado y estudiado las necesidades del ser humano y el insecto de preferencia. Posteriormente se procedió a socializar los avances, al tiempo que se registró la valoración de manera progresiva sin fijar metas por semana. Cada estudiante de acuerdo a sus responsabilidades con otras asignaturas y la disponibilidad de tiempo avanzó y presentó los resultados satisfactoriamente en el tiempo global estipulado.

El estudiante que logró desarrollar habilidades de pensamiento de orden superior y usó el conocimiento en Robótica y Biónica de manera básica como operador de nuevo conocimiento, no necesitó, en estas circunstancias, acreditarse como "ingeniero", estuvo ejerciendo la ingeniería en los constantes procesos de creación. No al contrario, en las universidades el estudiante debe acreditarse como profesional de ingeniería para ser reconocido como un profesional creativo. Desde este punto de vista, los contenidos no son realmente la hoja de ruta de los estudios en la ingeniería, hay un retroceso en la educación de las ingenierías. Son los conceptos las luces que brillan. En otras palabras, la persona somete su intelecto a la adquisición de conocimiento en el "hacer" por tanto, los contenidos pasan a ser aspectos que se entretejen en el proceso del conocimiento.

El uso de las TIC como apoyo en estudios sobre temas que involucran creatividad, contribuye a desarrollar habilidades de pensamiento de orden superior, facilitan su desarrollo en el ejercicio del diseño, el modelamiento, en la consulta y la presentación de resultados. [5][6]

En este trabajo, la autorregulación y la metacognición enfocada en los conceptos [7], permitieron fomentar las experiencias de aprendizaje en la práctica, propendieron por las actitudes de aprendizaje.

Las evaluaciones de los resultados de aprendizajes, cuando son orientadas al avance de las aptitudes hacia el nuevo pensamiento, mejoran la disposición de los estudiantes para adquirir conocimiento. Se sintieron cómodos en el proceso. El docente en el ejercicio de su función puede validar las evaluaciones en procesos de realimentación y estas evaluaciones se vuelven adecuadas cuando permiten al estudiante avanzar en medio de sus experiencias de aprendizaje, con respecto de su experiencia, no con respecto de una escala valorativa; insistentemente cuantitativa.

2.3 Aprovechar la voluntad de aprendizaje de los estudiantes en las universidades, fortalece la inteligencia y podría disminuir noTable.mente los índices de deserción

...y si el ingeniero que se forma en las facultades tiene la iniciativa de enfrentar una problemática, ¿cuantos más podrían tomar la decisión de desarrollar aptitudes de pensamiento en tanto el plan de estudios, o la malla curricular, es ampliamente escruTable. y justificadora de temas susceptibles de investigación?

Los siguientes programas se han destacado porque mediante la práctica especial mejoran la inteligencia general: el Programa de Stenberg (1986) para el desarrollo de la inteligencia práctica.; Programa de Enriquecimiento Instrumental de Feuerstein (Feuerstein y otros, 1985); Whimbey y Lochhead (1982,1984)para estudiantes de escuelas

secundarias y universidades; y Proyecto de Inteligencia de Venezuela (Bolt Beranek y Newman, 1983) [1]. Se caracterizan por la práctica y la realimentación en tareas como: construcción de vocabulario, actividades con sinónimos y antónimos, analogías, razonamiento espacial y algunas prácticas de lógica con distintos niveles de complejidad.

En las universidades, en una reacción por mejorar el currículo, éste podría estar matizado, por ejemplo, de los mencionados aportes con un enfoque sistémico en razón de las actividades que podrían tomarse como base, sólo aquellas que se adapten a los programas de ingeniería y de esta manera empezar a construir a partir de temas propuestos conjuntamente con estudiantes y docentes, el desarrollo de habilidades y aptitudes que contribuyan a dar solución a problemáticas en contexto.

2.4 Teoría General de Sistemas extensible en el tiempo

Se trata de poner como base a la TGS para toda actividad que se desarrolle en ingeniería, en todas las ingenierías, las condiciones están dadas, en el futuro cercano habrá trabajos que aún no se han definido, pero que están haciendo su proyección y sus avances se empiezan a dilucidar demandas que no sabemos afrontar. Algunas profesiones deben ir hacia los sistemas sin reparo, no funcionarán si no hay ese movimiento sinérgico. El estudiante de ingeniería debe apropiarse de las maneras posibles que estén al alcance para globalizar el conocimiento en torno a las exigencias del día a día. El conocimiento es el arma que desarma la educación, por consiguiente, el motor de empuje es permanecer inconforme a la linealidad, los problemas son heurísticos. Examine las manifestaciones estudiantiles en las universidades públicas, son de carácter interrelacional, y por tanto merecen ser abordados de manera ramificada, y por demás, compleja. Los problemas de conocimiento requieren aprehender y orientar a los estudiantes a saber hacer preguntas, a construir preguntas, a ser constructores de preguntas que contribuyan a resolver problemas a través de la investigación.

Los estudiantes son aspirantes del ámbito profesional, necesitan saber cómo desempeñarse en el campo laboral, en consecuencia, la ética ineludiblemente debe orientar toda actuación académico - laboral, no obstante, ésta se ve determinada por la calidad de educación en las instituciones de educación superior. Los enfoques sistémicos son bastante bien enredados en muchos casos, pero al fin resuelven, las tecnología de la información y de las comunicaciones permiten la gestión de manera ágil y una aptitud de pensamiento dibuja senderos donde no los hay, siempre que se tenga en cuenta la TGS se eliminan paulatinamente los laberintos del conocimiento.

III. CONCLUSIONES

Los esquemas de representación de pensamiento, en este artículo, aludido el mentefacto, soporta en sí mismo la carga de entendimiento de los conceptos. Es una herramienta fácil de comprender en su organización pero sí requiere de la indagación y de la base científica para crear descifrar conceptos o crear nuevos.

El docente que se dispone a desarrollar aptitudes de pensamiento, entendiendo que las habilidades de pensamiento son enseñables, los resultados marcan la diferencia no

solo en las instituciones, más importante aún marcan en el ser humano una experiencia gracias a tener la oportunidad de aplicar los conceptos de una manera libre, innovadora, con base científica y enfocada a la producción de conocimiento, se convierte en un juicio de valor que era requerido para afrontar nuevas situaciones que demandan una mente que sepa resolver en pocos pasos con efectividad y eficiencia.

El desarrollo de la creatividad, mediante el ejercicio de generación de aptitudes de pensamiento en el estudiante de secundaria y media vocacional, fortalece los desempeños en ingeniería dado que las habilidades de orden superior ya están prestas a ser utilizadas y por tanto sometidas a valoración mediante la interacción en situaciones de resolución de problemas, esto es un gran avance para alcanzar mejores resultados en el desempeño ingenieril disminuyendo asimismo la deserción

Como una forma de mejorar los desempeños y el goce del aprendizaje en los estudiantes de ingeniería, Marroquín (2012), plantea que docentes estratégicos forman estudiantes estratégicos y mediante dos aspectos a tener en cuenta: autorregulación y metacognición.

REFERENCIAS

[1] *Resnick, L.: La educación y el aprendizaje del pensamiento. 1ra edn. National Academy Press, Argentina (1999).*

[2] *Amestoy de Sánchez, M.: La investigación sobre el desarrollo y la enseñanza de las habilidades de pensamiento. Revista Electrónica de Investigación Educativa versión On-line ISSN 1607-4041.*

[3] *Santelices, C._: Vía para mejorar la calidad de la enseñanza científica. Pensamiento Educativo. Vol. 24 (julio 1999) pp 235-247*

[4] *Pressman, R._: Ingenieria del Software. Un Enfoque Práctico. 7ma edn. Mac Graw Hill, México (2010)*

[5] *Mayta, R., León, W.: El uso de las TIC en la enseñanza profesional: Redalyc, Revista de la Facultad de Ingeniería Industrial Vol. 12(2): pp 61-67 (2009) UNMSMISSN: 1560-9146 (Impreso) / ISSN: 1810-9993 (Electrónico)*

[6] *Marina, J.A.: El aprendizaje de la creatividad. Brújula para Educadores. Pediatr Integral 2013; XVII(2): 138-142*

[7] *Marroquín, M. (2013). Docentes estratégicos forman estudiantes estratégicos. Editorial Publicaciones UNIMAR, San Juan de Pasto, pp. 266.*

Biografía Autor(es)

Correspondencia.

Manuel Ernesto Bolaños González. Doctorado en Ciencias de la Electrónica. Docente Asociado Departamento de Sistemas, Universidad de Nariño. E-mail: mbolanos@udenar.edu.co, mbolanosg@gmail.com. Pasto, Colombia

Una aproximación a métricas de productividad para equipo en desarrollo ágil de Software

Giovanni Albeiro Hernández Pantoja (a)
Álvaro Alexander Martínez Navarro (b)
Robinson Andrés Jiménez Toledo (c)
Franklin Eduardo Jiménez Giraldo (d)

Resumen

Los métodos ágiles han sido adoptados de manera más frecuente en el desarrollo de software. Existe literatura sobre el uso de métricas como herramienta para la mejora continua en el desarrollo ágil de software (ASD), sin embargo, es muy limitada la literatura sobre métricas que se especialicen en medir la productividad de un equipo. Este artículo presenta una revisión sistemática de literatura sobre las métricas que evalúen la productividad de los equipos que construyen software en ASD. Se identificaron 822 artículos, los cuales se redujeron a 10 artículos principales, según el protocolo que se describe en este artículo. Los resultados indican que se encontraron 21 métricas de productividad en equipo, en una mayor proporción para medir el valor que agregan las tareas al producto software y para el cálculo del esfuerzo invertido en el desarrollo de una tarea. Al clasificarlas, se identificó que principalmente se orientan al proyecto y al desempeño organizacional; y al especificar las mediciones, la mayoría se ubican en el uso de una escala comparativa. Al inspeccionar, la relación con las nuevas tendencias sobre agilidad, la mayoría propician la reflexión-experimentación.

Palabras clave: Desarrollo ágil de software, equipo, métrica, productividad.

An approach to productivity metrics for teams in agile software development

Abstract

Agile methods have been adopted more frequently in software development. There is literature on the use of metrics as a tool for continuous improvement in agile software development (ASD), however, the literature on metrics that specialize in measuring the productivity of a team is very limited. This article presents a systematic review of the literature on the metrics that evaluate the productivity of teams that develop software in ASD. 822 articles were identified, which were reduced to 10 main articles, according to the protocol described in this article. The results indicate that, 21 team productivity metrics were found, in a greater proportion to measure the value added by the tasks to the software product and to calculate the effort invested in the development of a task. When classifying them, it was identified that they are mainly oriented to the project and organizational performance; and in specifying the measurements, most are in the use of a comparative scale. When inspecting, the relationship with new trends in agility, most encourage reflection-experimentation.

Keywords: Agile development of Software, team, metrics, productivity.

I. INTRODUCCIÓN

Un factor fundamental para un equipo que desarrolla software está relacionado con la productividad. Para mejorar la productividad de un equipo, se hace necesario conocer cómo se comporta de forma regular en unos intervalos de tiempo. En este sentido, se requiere recopilar y analizar datos durante estos periodos de tiempo como una vía o camino objetivo para aprender sobre él y generar información que posibilite tomar decisiones. Este actuar debe ser permanente y continuo; con la información generada, se puede aplicar cambios o ajustes a la forma de trabajo e iniciar con una nueva recopilación y análisis de datos.

Las organizaciones para llegar a ser más productivas, han venido incorporando métodos para el desarrollo ágil de software (Por sus siglas en Inglés ASD) [1]. Estos métodos, se basan en los principios enunciados en el manifiesto por el desarrollo ágil de software [2], que buscan desarrolladores de software motivados y capacitados, que confían en la excelencia técnica y el diseño simple, agregan valor al negocio entregando software funcional a los usuarios en intervalos cortos y regulares de tiempo [3]. Estos principios han generado una serie de prácticas que se cree, ofrecen mayor valor a los clientes. En el centro de estas prácticas se encuentra la idea de equipos auto-organizados cuyos miembros no solo están colocados, sino que también trabajan a un ritmo que mantiene su creatividad y productividad [3]. Además, el principal objetivo del ASD es proporcionar productos de software rápidamente [4] que solucionen problemas al cliente, pero a la vez permita realizar cambios frecuentes a las especificaciones.

Un elemento que permite recopilar datos sobre la productividad de un equipo son las métricas. Una métrica se comprende como los datos que se pueden obtener de la ejecución de un intervalo regular de tiempo (Ciclo, incremento, iteración) construyendo software en equipo [5]. Una métrica puede ser una simple fuente de datos o un conjunto de datos de múltiples fuentes. Generalmente, las métricas poseen información sobre una variable que se desea conocer. Un ejemplo de métrica en la construcción de software en equipo es la velocidad, es decir, la capacidad que un equipo tiene sobre el tiempo, para hacer funcionalidades de un producto software y que solucionan un problema que posee un cliente.

El uso de métodos ágiles se hace más frecuente en la industria de software [6] [7] y las métricas pueden ser usadas para recolectar información relevante. Cuando se construye software desde un enfoque ágil, las personas involucradas desempeñan diferentes roles y poseen una variada experiencia [5]. En este orden de ideas, el significado de relevante pude cambiar según el rol. Por ejemplo, al gestionar el proceso de hacer software utilizando *Scrum*, *el product owner* puede considerar que lo importante es el número de funcionalidades desplegadas; y que solucionan parte del problema al cliente. No obstante, un integrante del *development team* puede considerar que lo relevante es el nivel de complejidad que poseen las funcionalidades que se desarrollan. Los dos roles persiguen el mismo propósito, pero comprenden de manera diferente la manera de medir lo que están haciendo [5].

En el enfoque ágil, los valores y principios se pueden tornar abstractos al convertirlos en mediciones que proporcionen información. Por ejemplo, dentro de los principios del manifiesto ágil propuesto por Beck y otros [2] , se plantea que el software funcionando, es la principal medida de progreso. El concepto software trabajando, resulta ambiguo, porque puede estar dando respuesta a las funcionalidades priorizadas para solucionar un problema al cliente; sin embargo, no consigue abordar atributos de calidad

para desplegar el software. Por ejemplo, la funcionalidad es hacer una transferencia de dinero de una cuenta a otra; pero la seguridad como atributo de calidad, impide que se pueda desplegar el software sin haber tenido en cuenta todos los vacíos de seguridad que se puedan presentar al realizar esta acción.

Otro punto de reflexión sobre los principios del manifiesto ágil propuesto por Beck y otros [2] radica en la entrega temprana y continua de software con valor, es decir, el desarrollo de software de forma iterativa e incremental en intervalos regulares de tiempo. Esta acción hace que un equipo de desarrollo genere datos de manera permanente, pero sin una vista unificada [5], en diferentes sitios, a través del uso de diferentes técnicas y herramientas. Por ejemplo, para hacer seguimiento al trabajo del equipo, se puede estar utilizando un Table.ro *Kanban* donde se puede visualizar de la manera más simple, las tareas: pendientes, en progreso y finalizadas. Esta manera de hacer seguimiento al trabajo en equipo, se puede llevar manualmente, a través de una plantilla o utilizando una herramienta software. Cada una de ellas, representa una forma diferente de almacenar los datos para conocer cómo el equipo: comprende el proyecto, la rapidez con que actúa y la consistencia para completar el trabajo [5].

Por otra parte, se debe partir del supuesto que, cualquier medida utilizada para decidir, tiene que ser de bajo costo [5]. En este sentido, se puede comprender que no deben tener mayor peso el recurso y esfuerzo invertido para generar información que permita resolver problemas y mejorar, que recolectarla y analizarla. De igual forma, esta premisa es relevante para definir lo importante a medir. En este sentido han surgido iniciativas como el agilísmo moderno [8] y el corazón de lo ágil [9], donde se reflexiona acerca de los métodos ágiles y se concluye que desde la aparición del manifiesto por el desarrollo ágil de software [2], se ha venido involucrando aspectos que han hecho perder el horizonte de sentido de lo ágil o han logrado superar los principios y valores, en él expuestos.

La motivación por esta investigación surgió al buscar esTable.cer métricas que permitan medir la productividad de un equipo en ASD, que reduzca la ambigüedad y abstracción de los principios del manifiesto por el desarrollo ágil de software, que se aproximen a las nuevas reflexiones sobre el agilísimo.

II. TRABAJOS RELACIONADOS

Una métrica de equipo se entiende como la forma de recopilar y analizar datos, produciendo información objetiva que permita aprender sobre un equipo, aspecto que incluye medir ajustes que se decida hacer al comportamiento del mismo [5]. El comportamiento como se expresa en [10], se asocia principalmente con la productividad o el desempeño; y en la Ingeniería de Software, la productividad se define con frecuencia, desde un punto de vista económico, entendiéndose como la efectividad del esfuerzo productivo, es decir, la tasa de producción por unidad de entrada [10]. Generalmente, la unidad de entrada es el esfuerzo invertido en el desarrollo del software y la salida es el producto software. De acuerdo con el manifiesto que guía el desarrollo ágil de software [2], los principios que se relacionan con la productividad son:

- Satisfacer al cliente mediante la entrega temprana y continua de software con valor, es decir, que el software debe solucionar un problema y que la entrega se debe hacer de manera iterativa e incremental.

- Para que la entrega del software sea temprana y continua, se debe desplegar en periodos cortos de tiempo.

- Los proyectos se desarrollan en torno a individuos motivados, es decir, que la productividad para el equipo, se basa en el interés que los integrantes del equipo tengan por invertir el esfuerzo en el cumplimiento de un objetivo que significa agregar valor rápidamente a través del software.

- El software funcionando es la medida principal de progreso, pero que además agregue valor al cliente.

- La simplicidad, o el arte de maximizar la cantidad de trabajo no realizado, es fundamental para la mejora de la productividad.

Interrelacionando la definición de productividad [10] con los principios seleccionados del manifiesto por el desarrollo ágil de software [2] y el trabajo en equipo (Ver Tabla 1), se puede esTable.cer la productividad de un equipo en ASD como el esfuerzo invertido para motivar a un equipo a entregar tempana y continuamente software, utilizando periodos cortos de tiempo, realizando únicamente las tareas necesarias. El resultado debe ser software que agregue valor y satisfaga al cliente.

Table. 1. *Principios, productividad y equipo*

Unidad de entrada (esfuerzo invertido)	***Salida (Software)***
• *Entregar temprana y continuamente.* • *Trabajar periodos cortos de tiempo.* • *Motivar a los integrantes del equipo en el logro un objetivo.* • *Maximizar la cantidad de trabajo no realizado.*	• *Software que solucione un problema.* • *Satisfacer al cliente.* • *Software funcionando que agregue valor al cliente.*

Por otra parte, como resultado de más de 4 años de cooperación entre la academia y las empresas de desarrollo de software, en [11] se hace una propuesta donde se agrupan las métricas en cinco áreas principales: proyecto, diseño, desempeño organizacional, producto y negocio. En este orden de ideas, las métricas de productividad en equipo, serán las medidas que aporten a la definición de productividad en equipo dentro de ASD según las categorías propuestas por [11].

Existen varios estudios enfocados en identificar métricas en el desarrollo ágil de software. En [12], se analizó las razones y efectos de usar métricas en la industria de software, enfocándose en las métricas para equipos ágiles. Para lograr este fin, se desarrolla una revisión sistemática de literatura, y se esTable.ce un comparativo entre el uso de métricas tradicionales y ágiles. El trabajo desarrollado permite concluir que el uso de métricas en ASD es similar al desarrollo de software tradicional. Además, se debe incluir métricas relacionadas con los proyectos y *Sprints* para ser planeados y monitoreados; la calidad debe ser medida; los problemas en el proceso de software, deben ser identificados y solucionados.

En [1], se analizó literatura sobre investigaciones relacionadas con factores que inciden en la productividad en ASD, identificando: conceptos, métodos más utilizados, niveles y métricas de productividad. El análisis se realizó mediante un estudio de

mapeo sistemático (SMS) donde se seleccionó 25 artículos. Esta investigación permitió identificar los principales factores que afectan la productividad en ASD y el contexto en el cual los estudios seleccionados fueron realizados. El método de investigación que predomina es el estudio de caso, utilizando ampliamente la encuesta y la entrevista como métodos de recolección de datos. Además, se encontró que la mayoría de los factores informados que afectan la productividad en ASD están relacionados con el equipo, y aunque la calidad debe estar relacionada con la productividad, no fue considerada en todos los estudios.

En relación con el marco del proyecto Q-Rapids (Horizonte 2020) [13], se realizó un estudio de caso múltiple en cuatro empresas Agiles, utilizando el enfoque GoalQuestion-Metric (GQM), para investigar los motivos que explican la elección de las métricas de proceso en ASD y los desafíos que se enfrenta cuando se las operacionaliza. Los resultados reflejan que las empresas están interesadas en evaluar aspectos del proceso como: velocidad, desempeño en las pruebas y la precisión de la estimación; y prefieren elaborar métricas personalizadas para estas evaluaciones. Además, las empresas utilizan las métricas como un medio para acceder e incluso capitalizar los datos, hasta ahora inaccesibles debido a limitaciones técnicas o de proceso. Sin embargo, el contexto de desarrollo de una empresa puede obstaculizar la operacionalización de las métricas, manifestándose principalmente como falta de disponibilidad de los datos.

En [14], se presenta los resultados preliminares de una revisión sistemática de literatura sobre el uso de métricas en ASD. Como resultado, se identifica que las métricas y el uso se enfocan en las siguientes áreas: planeación y seguimiento de las iteraciones, motivación y mejora, identificación de problemas del proceso, calidad de un pre-lanzamiento y cambios en el proceso o las herramientas. Los hallazgos se contrastan con los principios para el desarrollo ágil de software y muestran que el uso de métricas los apoya, con algunas excepciones. Además, se encontró poca evidencia del uso de métricas para el código; y hay mucha evidencia sobre el uso de métricas de planificación y seguimiento.

En [15], se explorar que métricas están acordes con el proceso ASD, el uso de esas métricas en la práctica, los beneficios percibidos y las herramientas relacionadas. El análisis se basó en encuestas y entrevistas aplicadas a ingenieros de software de 24 empresas de desarrollo. Además, se identificó 10 métricas que pueden ser beneficiosas para el proceso ASD, donde los beneficios superan los gastos generales involucrados.

Existen trabajos que, si bien analizan las razones de la selección, uso y efectos de las métricas en ASD, aplicando técnicas como la revisión sistemática de literatura (LSR) o a través de estudios de caso aplicando encuestas y entrevistas; se ha encontrado un vacío en relación con qué métricas permiten medir la productividad de un equipo en ASD.

III. METODOLOGÍA

Para lograr el objetivo de identificar las métricas que permiten medir la productividad de un equipo en ASD, se utilizó como técnica la revisión sistemática de literatura (LRS) descrito en [16].

3.1 Preguntas de investigación

Las preguntas de investigación formuladas, se basan en la pregunta principal ¿Cuáles son las métricas que permiten medir la productividad de un equipo en ASD?

A partir de la pregunta principal, se derivaron las preguntas secundarias que permitieron analizar y categorizar los estudios primarios:

RQ1. ¿Qué métricas fueron estudiadas?

RQ2. ¿Cómo se categorizan las métricas?

RQ3. ¿Qué mediciones plantean las métricas?

RQ4. ¿Cómo se articulan las métricas con las nuevas reflexiones sobre el agilísimo?

3.2 Proceso de búsqueda

Como parte del protocolo para buscar estudios primarios, se identificó las fuentes de información y se definió la cadena de búsqueda según la pregunta de investigación principal. Las siguientes bases de datos, se utilizaron como fuentes de información: *ScienceDirect y Springer*.

La cadena general creada para esTable.cer los criterios de búsqueda, se configura a partir de los datos mostrados en la Tabla 2.

Table. 2. *Cadena de búsqueda*

Concepto	***Palabras relacionadas***
Metric	*Metric OR Metrical OR Measures OR Measured OR Assessment*
Team	*Team OR Team work OR Team building*
Agile	*Agile OR Rapid OR Scrum*
Software Development	*(Software OR System) AND (Development OR Creation OR Construction)*

La cadena de búsqueda fue elaborada construyendo expresiones que utilizan los operadores booleanos OR y AND. El operador OR, se lo utilizó para incorporar sinónimos del concepto de búsqueda, mientras que el operador AND permite agregar las palabras relacionadas en la cadena de búsqueda.

Para la ejecución de las búsquedas, se examinó: título, resumen y palabras clave, dentro de los resultados obtenidos a través del uso de los motores de búsqueda de cada fuente de datos seleccionada. La revisión de artículos se limitó a aquellos, que se encuentren escritos en inglés y dentro de una ventana de observación entre el 2013 y 2018.

3.3 Proceso de selección

Después de realizar el proceso de búsqueda, se procedió a seleccionar los estudios relevantes, es decir, aquellos artículos que permiten dar respuesta a las preguntas de investigación planteadas. Para determinar la relevancia de los estudios, se esTable.cieron unos criterios de inclusión y exclusión, como se pueden observar en la Tabla 3.

Table. 3. *Criterios de selección de estudios*

Criterios de inclusión	Criterios de exclusión
1. *Artículos se relacionan con métricas de desempeño en equipo para ASD.* **2.** *Artículos publicados en una ventana de observación entre 2013 y 2018.* **3.** *Artículos escritos en inglés.* **4.** *Artículos resultado de estudios primarios.*	**1.** *Artículos que no se relacionen con métricas en ASD.* **2.** *Reportes técnicos.* **3.** *Estudios duplicados.*

Para la aplicación de los criterios de inclusión y exclusión, como parte de la selección, se definieron los filtros que se muestran en la Tabla 4.

Table. 4. *Estrategia de selección*

Filtro	Descripción	Criterio aplicado	
		Inclusión	Exclusión
1F	*Buscar los artículos aplicando la cadena de búsqueda en los motores de las fuentes seleccionadas.*	*2 y 3*	
2F	*Leer el título, palabras clave y resumen del artículo aplicando los criterios de inclusión y exclusión.*	*1 y 4*	*1 y 2*
3F	*Leer el título, palabras clave y resumen del artículo aplicando los criterios de inclusión y exclusión.*	*1 y 4*	*1 y 2*
4F	*Eliminar los estudios duplicados.*		*3*
5F	*Leer el artículo completo y aplicar los criterios de inclusión y exclusión.*	*1 y 4*	*1 y 2*

En la Fig. 1, se puede observar los resultados obtenidos de la aplicación de los filtros definidos en la Tabla 4.

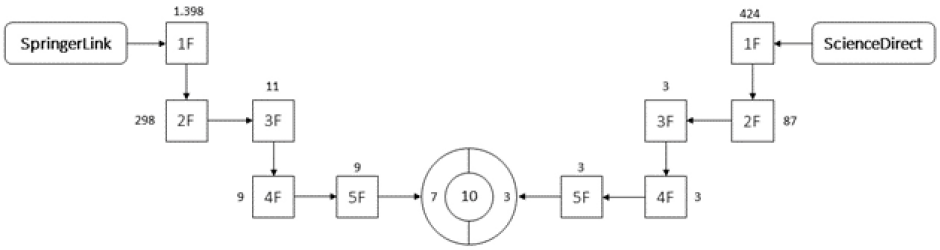


Fig. 1. *Resultados de la revisión sistemática*

3.4 Proceso de evaluación de la calidad

Posterior al proceso de selección, se realiza una nueva tarea para asegurar la calidad de los artículos encontrados, que corresponde a inspeccionar en los artículos seleccionados, un conjunto de criterios que se muestran en la Tabla 5.

Table. 5. *Criterios de evaluación de la calidad*

Criterio	*Categoría*
Los objetivos y preguntas de investigación, se describen de forma explícita, son claros y relevantes.	*Calidad del reporte*
La investigación presenta un diseño metodológico que le permite alcanzar los objetivos.	*Rigor*
El procedimiento de recopilación de datos es coherente con el diseño metodológico.	*Rigor*
Los resultados presentados son claros y coherentes con el diseño metodológico propuesto.	*Credibilidad*
El estudio es valorado por otros investigadores.	*Relevancia*

Para evaluar la calidad de los artículos, se esTable.ció una escala para inspeccionar el nivel de cumplimiento de los criterios, de la siguiente manera: Alto (2 puntos), Medio (1 punto) y Bajo (0 puntos). Los artículos que cumplieron con un 80% o más, del total de los puntos posibles, son lo que finalmente se eligieron. Al finalizar el proceso de evaluación de la calidad de los artículos, se incluyeron 10 artículos (Ver Apéndice A).

3.5 Extracción de datos

Posterior a la evaluación de la calidad de los artículos, se procedió a realizar la extracción de la información. En esta actividad, se obtuvo de cada estudio: autores, título,

año de publicación, tipo de documento, nombre de la conferencia o revista donde se presentó y publicó y la fuente donde se encontró. En la Tabla 6, se presenta las variables analizadas por cada estudio, en relación con las preguntas de investigación formuladas.

Table. 6. *Datos extraidos*

Variable	Descripción	Pregunta de investigación
Tipo	*Tipo de métricas utilizadas en el estudio.*	*RQ1*
Clasificación	*Categorías en las que se cataloga a las métricas.*	*RQ2*
Medición	*Tipo de medidas utilizadas.*	*RQ3*
Tendencias	*Métricas que surgen en relación con nuevas tendencias del agilismo (Corazón de lo ágil, agilismo moderno o ambos).*	*RQ4*

3.6 Síntesis de los datos

Finalizada la actividad de extracción de los datos, se procedió a categorizar la información de acuerdo con cada variable, reuniendo la información en relación con cada uno de los artículos que daban respuesta a las preguntas de investigación. Los resultados obtenidos, se presentan en la siguiente sección.

IV. RESULTADOS

Se analizaron 10 estudios primarios. En la Fig. 2 se muestra una distribución de los documentos incluidos por año y tipo de documento. Los artículos analizados incluyen 6 artículos de conferencia (60%) y 4 artículos de revistas (40%); donde el 60% son trabajos realizados entre los años 2017 y 2018. A continuación, se presentan los resultados encontrados en relación con las preguntas de investigación.

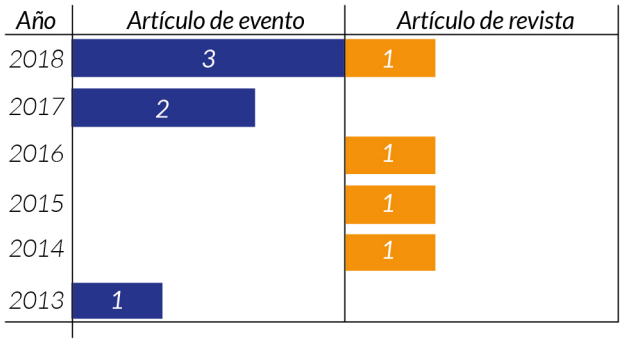


Fig. 2. *Estudios por año y tipo*

4.1 Métricas estudiadas (RQ1)

Al identificar las métricas utilizadas en los 10 estudios primarios, se encontraron 64 métricas en ASD. Del conjunto métricas encontradas, se procedió a extraer aquellas que se relacionan con productividad en equipo para ASD, cumpliendo los criterios que se plantean en la Tabla 7, obteniendo un total de 21.

Table. 7. *Criterios para seleccionar las métricas de productividad en equipo*

Criterios de inclusión	Criterios de exclusión
1. *Mide la motivación de un equipo.*	**1.** *No se relaciona con productividad en equipo para ASD.*
2. *Calcula esfuerzo por entregar software temprana y continuamente.*	**2.** *Métrica está duplicada.*
3. *Evalúa el esfuerzo en periodos cortos.*	**3.** *El cálculo es ambiguo*
4. *Valora el esfuerzo en el desarrollo de tareas.*	
5. *Mide el valor que agregan las tareas.*	
6. *Evalúa el valor agregado al cliente.*	
7. *Valora la satisfacción del cliente.*	

Como se puede observar en la Fig. 3, la mayoría de las métricas encontradas miden el valor que agregan las tareas al producto software (38.1%), seguido por aquellas que calculan el esfuerzo invertido en el desarrollo de una tarea (23.8%). Cabe destacar que no se encontraron métricas para medir motivación en el equipo.

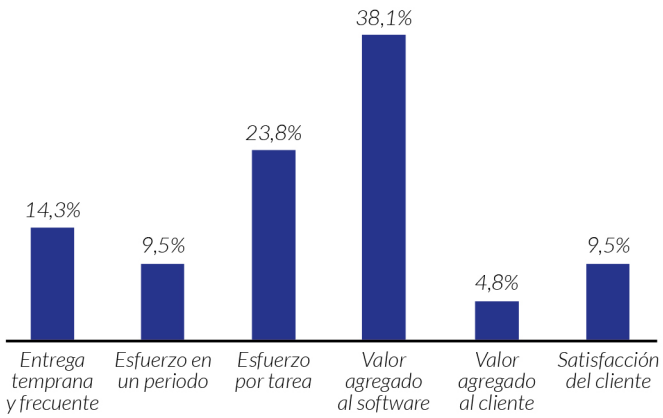


Fig. 3. *Tipo de métricas*

4.2 Clasificación de las métricas (RQ2)

Una vez, se estableció la relación de las métricas con productividad en equipo para ASD, se procedió a clasificarlas en relación con la propuesta planteada en [11], donde se definen las áreas: desempeño organizacional, diseño, negocio, producto y proyecto.

Para realizar la clasificación, se efectuó un contraste entre la descripción de cada una de las 21 métricas seleccionadas con la información proporcionada por cada sub-área de la propuesta de [11]. Como se observa en la Fig. 4, la mayoría de las métricas de productividad en equipo, se encuentran en primer lugar, orientadas al proyecto (38.1%), donde se valora la calidad, integración y pruebas hechas al producto (Valor agregado al software). En segunda instancia, se mide el desempeño organizacional (33.3%), orientado expresamente a calcular la efectividad y eficiencia del equipo (Esfuerzo por tarea o periodo para le entrega temprana y frecuente de software). En una menor proporción (19%), las métricas se orientan al negocio, donde se la evalúa los defectos en el uso de software y el retorno a la inversión (Satisfacción y valor agregado al cliente).

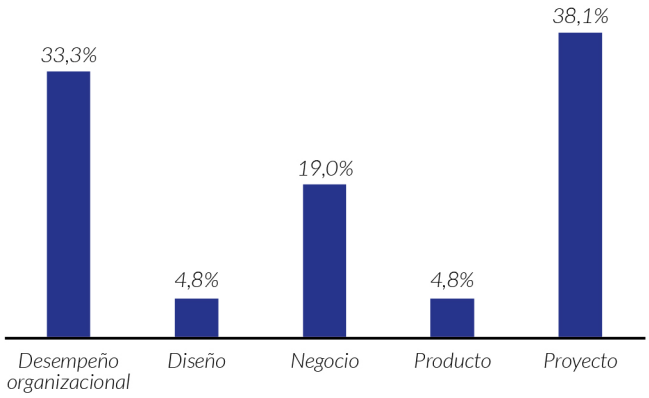


Fig. 4. *Métricas clasificadas por área*

4.3 Medición y medidas (RQ3)

La medición es un proceso donde se realiza una abstracción del mundo empírico al mundo formal y relacional [17]. En este sentido, una medida corresponde a un valor o un símbolo asignado a una entidad, como resultado de la abstracción, que permite caracterizar un atributo de la entidad [17]. Una forma de asignar un valor o un símbolo, son los esquemas de clasificación subjetiva [17], entre los que se encuentran: escala likert, clasificación forzada, escala de frecuencia verbal, escala ordinal, escala comparativa y escala numérica. En la Tabla 8, se muestra un ejemplo para cada medida.

Table. 8. *Ejemplo de un esquema de clasificación subjetiva*

Medida	Ejemplo
Escala Likert	*Muy alto, Alto, Medio, Bajo, Muy bajo*
Clasificación forzada	*1=Poco complejo, 5=Muy complejo*
Escala de frecuencia verbal	*Muy frecuente, Frecuente, Algunas veces, Casi nunca, nunca*
Escala ordinal	*Totalmente satisfecho, Satisfecho, Ni satisfecho, Ni insatisfecho, Insatisfecho, Totalmente insatisfecho*
Escala comparativa	*Muy superior=1 … Igual=5 … Muy inferior=10*
Escala numérica	*1=Sin importancia … 5=Muy Importante*

En [17], se plantea que para especificar mediciones en Ingeniería de Software, se debe identificar la entidad, atributo y medición. En la Fig. 5, se puede observar que la mayoría de las métricas de productividad en equipo se orientan a realizar mediciones utilizando una escala comparativa.

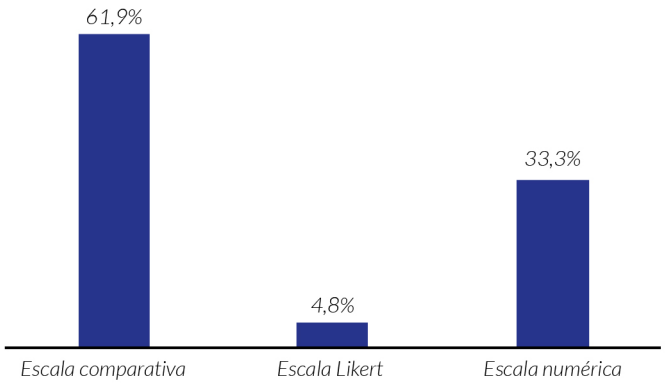


Fig. 5. *Tipo de mediciones*

En la Fig. 6, se puede observar que al clasificar las métricas de productividad para equipo en ASD encontradas, por atributo, se encontró que lo más utilizado corresponde a porcentajes, seguidos por frecuencias que se representan a través de cantidades y duración.

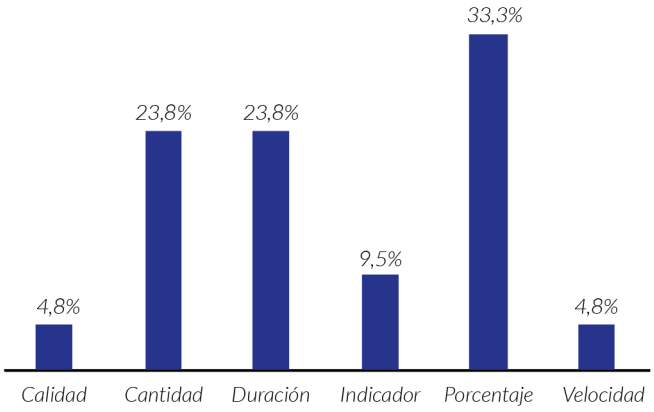


Fig. 6. *Tipo de mediciones por atributo*

4.4 Articulación con las nuevas tendencias (RQ4)

Heart of agile [8] plantea simplificar el ASD, porque considera que ha sido excesivamente decorada. En este orden de ideas, plantea realizar 4 acciones [8] que permiten retornar al centro de la agilidad, la cuales son: a.) Colabora, es decir, trabajar en equipo generar y desarrollar mejores resultados en las primeras versiones de las ideas, b.) Entrega, se entiende como, realiza pequeñas pruebas iniciales para aprender sobre el domino del problema, corrigiendo y direccionando objetivos, c.) Reflexiona,

entendiéndose como, pensar en la forma que se aprende trabajando en equipo y cuál ha sido la incidencia en las entregas, d.) Mejora, a partir de la reflexión, se transforma ideas, técnica y proceso. Las cuatro acciones invitan a pensar y actuar, permanentemente. En sí, las métricas favorecen recopilar datos sobre la productividad en equipo (Colaboración y entrega) y transformarlos en información (Reflexión) que permite ajustar (Mejora) el ASD. Estas acciones están perfectamente alineadas con el fin último de las métricas de productividad en equipo.

En *modern agile* [9], se plantean un conjunto de principios, que buscan, en primer lugar, hacer que las personas que intervienen en ASD puedan desplegar toda su potencialidad. En segunda instancia, se proyecta hacer que la seguridad psicológica sea un elemento fundamental, antes de realizar cualquier esfuerzo. En tercer lugar, se debe entregar valor de manera segura, continuamente. Finalmente, se plantea que se debe aprender rápidamente y para lograrlo, se debe experimentar frecuentemente, siendo este principio base para alcanzar los dos primeros principios.

Con *modern agile* aparecen nuevos retos, porque surge un principio basado en la seguridad psicológica [9], donde se argumenta que la productividad mejora cuando no se tiene miedo de ser uno mismo, de tomar riesgos, cometer errores, plantear problemas, hacer preguntas y estar en desacuerdo con los demás. En este sentido, surgen desafíos para el diseño de métricas, identificando atributos y medidas que se necesitan para hacerlo. La entrega de valor, el aprender rápidamente y experimentar, son acciones que se alinean con la recopilación sistemática de datos a través de métricas, para generar información que posibilite tomar decisiones para la mejora continua.

Haciendo una abstracción y síntesis de las nuevas tendencias en ASD, los principios que están promoviendo son: colaboración, entrega, reflexión-experimentación y seguridad psicológica. Al inspeccionar, la relación que existe entre las métricas de productividad en equipo identificadas en este estudio con las nuevas tendencias, se encontró que el 76.2% (16) de las métricas propician la reflexión-experimentación y un 23,8% (5) se orientan a la entrega de software con valor. No se encontraron métricas para colaboración y seguridad psicológica.

V. CONCLUSIONES

El desarrollo del proceso de revisión sistemática de literatura en dos fuentes de información (*Springer y Science Direct*), permitió identificar 10 estudios primarios relacionados con métricas de productividad en equipo para ASD, de los cuales, se extraen 21 métricas.

Al establecer la relación entre los principios por ASD y las métricas identificadas, se determina en una mayor proporción, métricas para el valor que agregan las tareas al producto software, seguido por aquellas que calculan el esfuerzo invertido en el desarrollo de una tarea. Además, no se encontraron métricas para medir motivación en el equipo.

Al clasificar las métricas de productividad en equipo para ASD, se encontraron principalmente, aquellas que se orientan al proyecto, donde se valora la calidad, integración y pruebas hechas al producto; y las que se mide el desempeño organizacional, orientado expresamente a calcular la efectividad y eficiencia del equipo.

Al especificar las mediciones de las métricas de productividad en equipo para ASD, la mayoría se orientan al uso de una escala comparativa y en relación con el atributo a medir, se encontró que lo más utilizado corresponde a porcentajes, seguidos por frecuencias que se representan a través de cantidades y duración.

Al inspeccionar, la relación que existe entre las métricas de productividad en equipo para ASD, con las nuevas tendencias sobre agilidad, la mayoría propician la reflexión experimentación. No se encontraron métricas para colaboración y seguridad psicológica.

La principal limitación de este estudio correspondió al uso de dos repositorios de estudios primarios, sería importante ampliar esta investigación, utilizando este mismo protocolo con otras fuentes de datos como ACM Library e IEEE Xplore Digital Library.

REFERENCIAS

[1] *S. Ramírez y H. Oktaba, «Productivity in Agile Software Development: A Systematic Mapping Study,» de 2017 5th International Conference in Software Engineering Research and Innovation (CONISOFT), Mérida, Yucatán, México, 2017.*

[2] *K. Beck y Otros, «Manifesto for Agile Software Development,» 2001. [En línea]. Available: http://agilemanifesto.org/. [Último acceso: 12 5 2017].*

[3] *T. Dingsøyr, S. Nerur, V. Balijepally y N. Moe, «A decade of agile methodologies: Towards explaining agile software development.,» The Journal of Systems and Software, nº 85, pp. 1213-1221, 2012.*

[4] *S. Yamada y R. (. 1.-5. Kii, «Software quality analysis for agile development,» de 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), Noida, India, 2015.*

[5] *C. Davis, Agile Metrics in Action, New York - USA: Manning Publications Co., 2015.*

[6] *P. Rodríguez, J. Markkula, M. Oivo y K. Turula, «Survey on Agile and Lean Usage in Finnish Software,» de Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Lund, Sweden, 2012.*

Biografía Autor(es)

a. Magíster en Docencia Universitaria. Docente, Universidad Mariana. ORCID: https://orcid.org/0000-0001-7903-6513. E- mail: gihernandez@umariana.edu.co. Pasto, Colombia.

b. Magíster en Docencia Universitaria. Docente, Universidad Mariana. ORCID: https://orcid.org/0000-0002-6341-4863. E- mail: amartinez@umariana.edu.co. Pasto, Colombia.

c. Magíster en Docencia Universitaria. Docente, Universidad Mariana. ORCID: https://orcid.org/0000-0003-2218-0850 . E-mail: rjimenez@umariana.edu.co. Pasto, Colombia.

d. Magíster en Software Libre. Docente, Universidad Mariana. ORCID: https://orcid.org/0000-0002-3071-7198. E-mail: fjimenez@umariana.edu.co. Pasto, Colombia

UNIVERSIDAD
DIOS
CIENCIA Y SERVICIO
CESMAG

EDITORIAL
UNIVERSIDAD CESMAG