RedUNCI

RED DE UNIVERSIDADES CON CARRERAS EN INFORMÁTICA

Computer Science & Technology Series

XXII Argentine Congress of Computer Science Selected Papers

Patricia Mabel Pesado | Marcelo Gustavo Estayno | María Fabiana Piccoli
(Eds.)

edulp
Editorial de la Universidad de La Plata

Computer Science & Technology Series

XXII Argentine Congress of Computer Science
Selected Papers

Computer Science & Technology Series

XXII Argentine Congress of Computer Science
Selected Papers

Patricia Mabel Pesado / Marcelo Gustavo Estayno / María Fabiana Piccoli
(Eds.)

Editorial de la Universidad de La Plata

Computer science & technology series: XXII Argentine Congress of Computer Science Selected Papers / María Fabiana Piccoli... [et al.]; compilado por Patricia Mabel Pesado; Marcelo Estayno; María Fabiana Piccoli. - 1a ed . - La Plata: EDULP, 2017.
364 p.; 24 x 15 cm.

ISBN 978-987-4127-28-0

1. Actas de Congresos. I. Piccoli, María Fabiana II. Pesado, Patricia Mabel, comp. III. Estayno, Marcelo, comp. IV. Piccoli, María Fabiana , comp.
CDD 005.1

Computer Science & Technology Series
XXII Argentine Congress of Computer Science
Selected Papers

Diagramación: Andrea López Osornio

Edulp
Editorial de la Universidad de La Plata

Editorial de la Universidad de La Plata (Edulp)
47 N.° 380 / La Plata B1900AJP / Buenos Aires, Argentina
+54 221 427 3992 / 427 4898
editorial@editorial.unlp.edu.ar
www.editorial.unlp.edu.ar

Edulp integra la Red de Editoriales Universitarias Nacionales (REUN)

Primera edición, 2017
ISBN 978-987-4127-28-0
Queda hecho el depósito que marca la Ley 11.723

Impreso en Argentina

Topics

XVII Intelligent Agents and Systems Workshop
Chairs Sergio A. Gómez (UNS), Marcelo Arroyo (UNRC), Guillermo Leguizamón (UNSL)

XVII Distributed and Parallel Processing Workshop
Chairs Fabiana Piccoli (UNSL), Laura De Giusti (UNLP), Carlos García Garino (UNCuyo)

XV Information Technology Applied to Education Workshop
Chairs Cristina Madoz (UNLP), Sonia Rueda (UNS), Alejandra Malberti (UNSJ), Claudia Russo (UNNOBA)

XIV Graphic Computation, Images and Visualization Workshop
Chairs Silvia Castro (UNS), Roberto Guerrero (UNSL), Javier Giacomantone (UNLP)

XIII Software Engineering Workshop
Chairs Patricia Pesado (UNLP), Elsa Estevez (UNU), Alejandra Cechich (UNCOMA), Horario Kuna (UNaM)

XIII Database and Data Mining Workshop
Chairs Hugo Alfonso (UNLPam), Rodolfo Bertone (UNLP), Nora Reyes (UNSL)

XI Architecture, Nets and Operating Systems Workshop
Chairs Jorge Ardenghi (UNS), Carlos Buckle (UNPSJB), Orlando Micolini (UNC)

VIII Innovation in Software Systems Workshop
Chairs Pablo Fillottrani (UNS), Dante Zanarini (UNR), Jorge Finocchietto (UCAECE), Marcelo Estayno (UNLZ)

VII Signal Processing and Real-Time Systems Workshop
Chairs Oscar Bría (INVAP), Osvaldo Spositto (UNLM), Horacio Villagarcía Wanza (UNLP), Emmanuel Frati (UNChilecito)

V Computer Security Workshop
Chairs Javier Díaz (UNLP), Hugo Ramón (UNNOBA), Antonio Castro Lechtaler (UBA)

V Innovation in Computer Science Education Workshop
Chairs Cecilia Sanz (UNLP), Beatriz Depetris (UNTDF), Marcelo De Vincenzi (UAI), Uriel Cukierman (UTN)

V ETHICOMP LatinAmerica
Chairs Guillermo Feierherd (UNTF), William Fleischman (U. Vilanova - USA)

Scientific Committee

Abásolo, María José (Argentina)
Acosta, Nelson (Argentina)
Alfonso, Hugo (Argentina)
Ardenghi, Jorge (Argentina)
Arroyo, Marcelo (Argentina)
Astudillo, Hernán (Chile)
Baldasarri, Sandra (España)
Balladini, Javier (Argentina)
Barbosa, Luiz (Portugal)
Bertone, Rodolfo (Argentina)
Bria, Oscar (Argentina)
Brisaboa, Nieves (España)
Buckle, Carlos (Argentina)
Castro Lechtaler, Antonio (Argentina)
Castro, Silvia (Argentina)
Cechich, Alejandra (Argentina)
Chavez, Edgar (México)
Coello Coello, Carlos (México)
Constantini, Roberto (Argentina)
Cuevas, Alfredo Simón (Cuba)
Cukierman, Uriel (Argentina)
De Giusti, Armando (Argentina)
De Giusti, Laura (Argentina)
De Vincenzi, Marcelo (Argentina)
Deco, Claudia (Argentina)
Depetris, Beatriz (Argentina)
Diaz, Javier (Argentina)
Dix, Juerguen (Alemania)
Doallo, Ramón (España)
Docampo, Domingo (España)
Dujmovic, Jozo (USA)
Echaiz, Javier (Argentina)
Esquivel, Susana (Argentina)
Estayno, Marcelo (Argentina)
Estevez, Elsa (Argentina)
Falappa, Marcelo (Argentina)
Feierherd, Guillermo (Argentina)
Fillottrani, Pablo (Argentina)
Finocchietto, Jorge (Argentina)
Fleischman, William (USA)
Frati, Emanuel (Argentina)
García Garino, Carlos (Argentina)
García Villalba, Javier (España)
Género, Marcela (España)
Giacomantone, Javier (Argentina)
Gómez, Sergio (Argentina)
Gröller, Eduard (Austria)
Guerrero, Roberto (Argentina)
Janowski, Tomasz (Naciones Unidas)
Kantor, Raul (Argentina)
Kuna, Horacio (Argentina)
Lanzarini, Laura (Argentina)
Leguizamón, Guillermo (Argentina)
Lopez Gil, Fernando (España)
Loui, Ronald Prescott (EEUU)
Luque, Emilio (España)
Madoz, Cristina (Argentina)
Malberti, Alejandra (Argentina)
Malbran, Maria (Argentina)
Manresa Yee, Cristina (España)
Marco, Javier (España)
Marín, Mauricio (Chile)
Mas Sansó, Ramón (España)
Matrángolo, Carlos (Argentina)
Micolini, Orlando (Argentina)
Motz, Regina (Uruguay)
Naiouf, Marcelo (Argentina)
Navarro Martín, Antonio (España)
Olivas Varela, José Angel (España)
Padovani, Hugo (Argentina)
Pandolfi, Daniel (Argentina)
Pardo, Álvaro (Uruguay)
Pesado, Patricia (Argentina)
Piattini, Mario (España)
Piccoli, María Fabiana (Argentina)
Printista, Marcela (Argentina)
Puppo, Enrico (Italia)
Ramón, Hugo (Argentina)
Rexachs, Dolores (España)
Reyes, Nora (Argentina)
Riesco, Daniel (Argentina)
Roig Vila, Rosabel (España)
Rossi, Gustavo (Argentina)
Rosso, Paolo (España)
Rueda, Sonia (Argentina)
Ruiz, Francisco (España)
Russo, Claudia (Argentina)
Sanz, Cecilia (Argentina)
Saroka, Raúl (Argentina)
Simari, Guillermo (Argentina)
Spositto, Osvaldo (Argentina)
Steinmetz, Ralf (Alemania)
Suppi, Remo (España)
Tarouco, Liane (Brasil)
Tirado, Francisco (España)
Vaquila, Isidoro (Argentina)
Velho, Luiz (Brasil)
Vendrell, Eduardo (España)
Vénere, Marcelo (Argentina)
Villagarcia Wanza, Horacio (Argentina)
Zanarini, Dante (Argentina)

Preface

CACIC Congress

CACIC is an annual Congress dedicated to the promotion and advancement of all aspects of Computer Science. The major topics can be divided into the broad categories included as Workshops (Intelligent Agents and Systems, Distributed and Parallel Processing, Software Engineering, Architecture, Nets and Operating Systems, Graphic Computation, Visualization and Image Processing, Information Technology applied to Education, Databases and Data Mining, Innovation in Software Systems, Security, Innovation in Computer Education, Computer Science Theory, Signal Processing, Real time Systems and Ethics in Computer Science).

The objective of CACIC is to provide a forum within which to promote the development of Computer Science as an academic discipline with industrial applications, trying to extend the frontier of both the state of the art and the state of the practice.

The main audience for, and participants in, CACIC are seen as researchers in academic departments, laboratories and industrial software organizations.

CACIC started in 1995 as a Congress organized by the Network of National Universities with courses of study in Computer Science (RedUNCI), and each year it is hosted by one of these Universities. RedUNCI has a permanent Web site where its history and organization are described: *http://redunci.info.unlp.edu.ar.*

CACIC 2016 in San Luis

CACIC'16 was the 22th Congress in the CACIC series. It was organized by the Computer Science Department at the School of Mathematics, Physics and Natural Sciences of the San Luis National University. (http://unsl.edu.ar/)
The Congress included 13 Workshops with 136 accepted papers, 2 Conferences, 2 invited Tutorials, different meetings related with Computer Science Education (Professors, PhD students, Curricula) and an International School with 6 courses. (http://www.cacic2016.unsl.edu.ar/).
CACIC 2016 was organized following the traditional Congress format, with 13 Workshops covering a diversity of dimensions of Computer Science Research. Each topic was supervised by a committee of 3-5 chairs of different Universities.
The call for papers attracted a total of 185 submissions. An average of 2.5 review reports were collected for each paper, for a grand total of 462 review reports that involved about 176 different reviewers.
A total of 136 full papers, involving 457 authors and 79 Universities, were accepted and 30 of them were selected for this book.

Acknowledgments

CACIC 2016 was made possible due to the support of many individuals and organizations. The Computer Science Department at the School of Mathematics, Physics and Natural Sciences of the San Luis National University, RedUNCI, the Secretary of University Policies, the National Ministry of Science and Technology, and CONICET were the main institutional sponsors.

This book is a very careful selection of best qualified papers. Special thanks are due to the authors, the members of the workshop committees, and all reviewers, for their contributions to the success of this book.

Ing. Armando De Giusti
RedUNCI

Table of Contents

XVII

Intelligent Agents and Systems Workshop

A Basic Framework for Stream Reasoning with Argumentation Systems

JOSÉ N. PAREDES, FABIO R. GALLO, GERARDO I. SIMARI, MARCELO A. FALAPPA

Departamento de Ciencias e Ingeniería de la Computación,
Universidad Nacional del Sur (UNS) and
Instituto de Ciencias e Ingeniería de la Computación
(CONICET–UNS)
San Andrés 800, (8000) Bahía Blanca, Argentina
{jose.paredes, fabio.gallo, gis, mfalappa}@cs.uns.edu.ar

Abstract. Advances in information technology make it easier to accurately generate and process data about what happens in a complex domain. The applications that operate in these environments are characterized by the reception of large amounts of data in a short time, high rates of change, uncertainty, and incompleteness; those source of data—which cannot be stored and processed later—are called "*streams*". In this work, we present a proposal for carrying out stream reasoning via argumentative systems, by means of a novel combination of (i) complex events and stream processing techniques, (ii) structured argumentation tools, (iii) probabilistic reasoning, and (iv) belief revision operators.

Keywords: Stream Processing, Structured Argumentation, Reasoning under Uncertainty, Belief Revision.

1. Introduction and Motivation

Nowadays, a large amount of very frequent, dense, heterogeneous, and digital information is produced when people carry out their daily activities, leaving traces composed of call records, text messages, GPS usage, social network posts, and many other sources. Clearly, those traces can be used to describe less obvious details of the surrounding environment, always taking care not to violate each person's privacy constraints.

At the same time, the fact that more and more devices that produce abundant information are becoming available to more people cause companies and other organizations to implement technologies that generate data about what happens in everyday environments. In this context, an application domain arises that tries to take advantage of this situation but requires that the *stream processing* (SP) often take place within a short amount of time. However, due to the characteristics mentioned above, it is not possible (or necessary) to process *all* the incoming information; Therefore, systems that operate in

those environments use a concept known as *data window*, which allows to define the data that is available for processing over a period of time. If one wants to use this information to take advantage of the semantic value in the best possible way, conventional SP is not enough. Hence, we propose the use of knowledge representation and reasoning techniques for this purpose. Argumentation fits into this category, and its origins are linked to the philosophers of antiquity who studied the nature of human dialogue. Specifically, argumentation is based on dialectical processes that allow to elaborate arguments in favor and against certain a given position. Then, it is possible to weigh each argument in relation to others that can weaken this position and favor an opposing one. As a result, one can determine the winning conclusions after completing this process. This approach provides a way to deal with incomplete and possibly contradictory information; as an additional advantage, the process yields *explanations* that were used to arrive at a certain conclusion. Different ways of carrying out this process have been proposed in the literature, which are divided mainly into abstract [8] and concrete [11]; the former is characterized by the analysis of available arguments and the relation of attacks among them, while the latter assumes the availability of the *internal structure* of each argument.

In this work we propose to combine the areas of stream processing and argumentation, giving rise to a novel development within the area of stream reasoning (SR) [5]. We base our work on the DeLP3E formalism proposed in [23,24], which divides the knowledge base (KB) into two different but interrelated parts: the *environmental model* (EM), which defines relationships among events that, by their nature, must be handled with uncertainty models; on the other hand, the *analytical model* (AM) handles potentially contradictory information through an argumentative analysis based on the *DeLP with presumptions* (PreDeLP) [17] formalism. The latter also has the possibility of offering the explanations that justify answers to a given query. From the point of view of SP, the contents of both EM and AM can be affected by the high rates of incoming information. Therefore, the main issues to deal with are a combination of the hurdles that arise in each of the following areas:

(i) *Incompleteness, inconsistency and uncertainty*: often we must reasoning with information that is either insufficient to reach accurate conclusions offering contradictory data, or perhaps describing inherently uncertain phenomena (such as financial markets).

(ii) *Volume and speed*: the amount of information coming from different sources at high rates makes it impossible to store all of it for its subsequent careful processing; one must then decide what to store and what to discard in order to make the system work in the best possible way.

(iii) *Belief dynamics*: as a consequence of the first two points, we can derive the need to adequately manage the dynamics of the beliefs held by the system; revision and consolidation operators in general are designed for environments with much lower rates of change.

(iv) *Computational tractability*: the use of rich tools for reasoning and knowledge representation usually involves paying a cost in terms of computational complexity; this is even more central in probabilistic reasoning models. Mechanisms must be developed whereby approximations can be obtained and in which quality degradation is as graceful as possible; moreover, the user should ideally know the effects of making compromises between execution time and quality of the answers obtained.

Example 1. Consider the self-driving car developed by Google [13]; to be able to transit the streets of a city it requires a large amount of information about the surrounding environment. For this purpose it has various sensors:

- LIDAR ("*Laser Imaging Detection and Ranging*"), which allows to identify objects and measure distances using light beams. Specifically, Google uses the high-definition Velodyne HDL-64E S2, which has 64 laser beams and rotates 360 degrees continuously up to 900 turns per minute in order to monitor the situation around the car; it observes 1.3 million points per second, in order to construct a three-dimensional image of the entities involved (pedestrians, other vehicles, street lighting, trees, etc.). It has a range of 50 meters for pavement and 120 meters for cars and trees.
- GPS and an inertial unit, which measures the acceleration and the angular velocity using accelerometers, gyroscopes and magnetometers. That is, it accurately identifies car's movements.
- Four radars that use radio waves; three of them are located on the front bumper (one in the center and the other two on the side corners). The fourth is located in the rear bumper—its function is to identify objects and measure distances around the car.

Together, sensors offer the ability to know what happens in real time around the vehicle, which means that this information varies at *every instant* in time, and much of it is only relevant for a fairly short time, such as the position of pedestrians, other cars, and closed roads (for social protests, repairing, etc.). On the other hand, the system should also take into account information that does not generally change over time, such as map data. Google invested many resources towards elaborating maps, both for the development of *Google Maps* and *Street View*; both contain semantic value that goes beyond a traditional map.

Sensors generate a large amount of data; the LIDAR produces 1.3 million points per second for a 360 degree horizontal vision range and a vertical vision range of 26.8 degrees—the total output generated is 100 Mbps of UDP packets [6]. This is a large volume that cannot be stored of processed in its entirety. In this context, the system that drives the car must make certain decisions considering that the information available is affected by uncertainty. In addition, it must have mechanisms to manage newly incorporated information (such as current climate conditions), being aware that contradictions can occur with the data that is already present.

2. Towards a Combination of Tools

In this work we propose to develop a stream processing formalism combining argumentation systems with conventional stream processing techniques. As discussed in the previous section, belief revision and probabilistic reasoning are central to achieving this goal. Each one of these topics has been treated in the literature, generating combinations between them; however, we do not know of any proposal that combines all at same time. In this section we first review the most significant works in this sense (Sections 2.1–2.4), and then we go on to describe our proposal in Section 2.5.

2.1 Probabilistic Structured Argumentation

In the literature there are several formalisms that have been proposed for structured argumentation; in this work we choose *defeasible logic programming* (DeLP); we will make a brief introduction and refer the reader to [11] for a deeper treatment. A DeLP program is composed of sets of facts, strict rules, and defeasible rules. This is usually denoted by $\Pi = (\Theta, \Omega, \Delta)$, where Θ is the set of facts, Ω is the set of strict rules, and Δ is the set defeasible rules. Facts are basic literals (i.e., without variables) that represent atomic information or its negation; they are always true and cannot be contradictory. Strict rules represent cause and effect information that is always true; they are specified by the combination of ground literals and have the form $L_0 \# L_1,\ldots, L_n$, where L_0 is a ground literal and $\{L_i\}_{i>0}$ is a set of ground literals. On the other hand, defeasible rules include knowledge that is true if contradictory information is not available—they are similar to strict rules, but represent weaker knowledge, and have the form $L_0 ! L_1,\ldots,L_n$, where L_0 is a ground literal and $\{L_i\}_{i>0}$ is a set of ground literals. Strong negation is allowed for both strict and defeasible rules in order to represent contradictory information. Using these elementary concepts we can now introduce the notion of argument; an *argument* for a literal L is a pair $\langle A, L\rangle$, where $A \subseteq \Delta$ consists of a proof for L such that: (1) L is defeasibly derived from A, (2) $A \cup \Omega \cup \Theta$ is not contradictory, and (3) A is a minimal subset of Δ that satisfies (1) and (2). Literal L is called the *conclusion* supported by the argument, and A is the *support*. An argument $\langle B, \mathrm{L}\rangle$ is a *sub-argument* of $\langle A, L'\rangle$ iff $B \subseteq A$.

In order to handle probabilistic uncertainty we can take the model defined in [23], which combines DeLP with random variables. This gives rise to so-called *DeLP3E* programs, denoted with $P = (\Pi_{EM}, \Pi_{AM}, af)$, where Π_{EM} allows the representation of uncertain knowledge subject to probabilistic events, Π_{AM} allows the representation of strict and defeasible knowledge (DeLP program), and *af* is a function known as an *annotation function*, such that its domain is the set Π_{AM} and its image is a subset of the ground logical formulas than can be formed from the elements defined in Π_{EM}. At the same time, the set of all logical formulas formed by basic atoms in Π_{EM} defines a set of *possible worlds*. For each of these worlds

there is a subset of elements of Π_{AM} that are valid, which determines a (non-probabilistic) DeLP program that exists in that world. The output is a literal L which can be *warranted* in some worlds and not in others. Thus, the probability that a literal L is warranted results from the sum of the probabilities of the worlds in which L is warranted.
Specifically Π_{EM} must be supported by a model of representation of random events, such as Bayesian Networks [21], Markov Logic Networks [7], Nilsson Probabilistic Logic [20], etc.

2.2 Belief Revision

The area of belief revision deals with how an agent's epistemic states must change when new epistemic input arrives; in other words, how beliefs should be revised in the presence of new information that possibly contradicts the beliefs established so far. Traditionally, epistemic states have the form of either belief sets [1,12] (which are closed under the consequence operator) or belief bases (which are not closed) [9,14,15,16]. It is clear that belief revision appears constantly in the real world, and any intelligent system that works with data must be prepared to carry out some kind of revision. The typical methodology consists of the development of operators that take the current knowledge base and the epistemic input and produce a new knowledge base that corresponds to the result of revising the beliefs. These operators are generally characterized by the properties they must satisfy (expressed in the form of postulates); then, algorithmic constructions are proposed and it is formally proved that the two characterizations are equivalent—this type of result is called a *representation theorem*. As mentioned in the previous section, it is necessary to investigate operators that are especially suitable to operate in an SP environment and, in particular, with quantitative aspects to make the most of the underlying probabilistic model; as far as we know, the only work in this research area is a recent proposal in [24].

2.3 Stream Reasoning

Although more and more information is available through the Web or the use of mobile devices, answering questions that at first appear simple becomes extremely complicated. For example, one could obtain information regarding the clicks that a user makes when browsing the Web, and based on this one could learn their interests, what news caught their attention, and perhaps generally determine the behavior readers regarding related news. However, this type of query requires systems that can handle the quickly-changing nature of the real world, which provides obstacles at a semantic level. Although rapidly-changing data can be analyzed by specialized real-time *stream processing systems* [2,4], these systems cannot perform *intelligent* (complex) processing tasks. For this purpose, systems with such characteristics have been investigated and developed in the last decade [5,3]. On the other hand, in [10] different techniques for *complex event* processing under uncertainty are analyzed. The authors considered

techniques based on automata and logic, considering the uncertainty related to the imperfection in the rules that define events, and the incompleteness and errors derived from the stream. As can be seen, the main difference with our proposal is that these techniques only perform processing with very low semantic content. These developments have proposed initial solutions and outlined the most important problems to be solved; however, the area of stream reasoning area is still in its early phases of evolution.

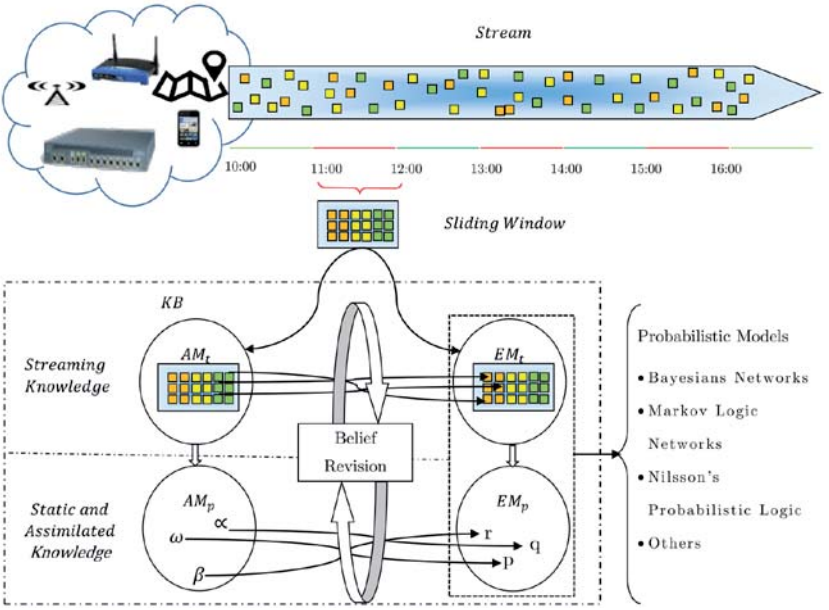


Fig. 1. Outline of a formalism to combine techniques and tools from structured argumentation, belief revision, probabilistic reasoning, and stream processing.

2.4 Structured Argumentation Systems in Dynamic Environments

The problem of performing stream reasoning with significant semantic value is extremely complex due to the high data volume and frequency. In the research line proposed in [22,19,18] the authors proposed a formalism to represent and reason about changing knowledge by *activating* and *deactivating* arguments according to available evidence. However, although the model takes into account a changing context, the knowledge dynamics considered assumes neither high speed nor large amounts of input data, so that computational complexity is not a central aspect. Furthermore, they do not consider data affected by uncertainty.

2.5 Proposed Model

Our proposal is summarized in Figure 1, which shows the general operation of the proposed model. It has a stream processing system that generates events and notifies either AM_t or EM_t. AM_p contains facts, presumptions, and strict and defeasible rules relating to knowledge that generally does not change over time and is already assimilated.

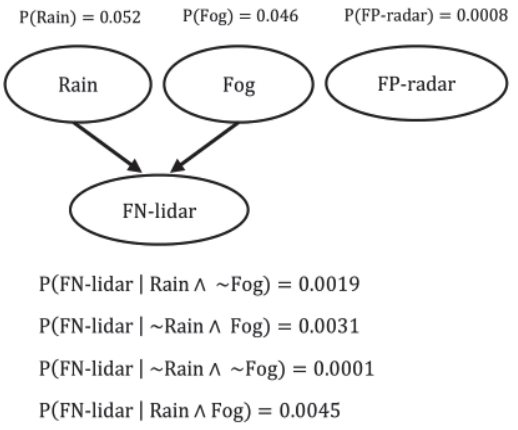


Fig. 2. Bayesian network used in the EM_p and EM_t models from Example 2.

	AM_t	*Anotation Function*
Θ:	Θ_1 = my_location(t1, position, direction)	True
	Θ_2 = pedestrian_ahead	~FP-radar
	Θ_3 = ~pedestrian_ahead	~FN-lidar
	AM_p	
Θ:	*"Set of facts representing the map of the city"*	True
Ω:	Ω_1 = stop ← pedestrian_ahead	True
Δ:	Δ_1 = give_way –≺ car_crossing, ~traffic_lights	True

Fig. 3. AM_t and AM_p from Example 2, along with the corresponding annotation function. The rule sets of AM_t are empty.

On the other hand, EM_p contains a probabilistic model in relation to random variables whose information also generally does not change over time and is already assimilated. AM_t contains facts, presumptions, and strict and defeasible rules relating to knowledge that generally changes over time and comes from the stream. It is important to emphasize that the elements of AM_t

can become part of AM_p if they acquire relevance. EM_t contains a probabilistic model in relation to random variables whose information generally changes over time and that comes from the stream. As in *AM*, some elements of EM_t can become part of EM_p.

Example 2. If we want to apply the model presented in Figure 1 to the scenario in Example 1, we could think of the next simplified situation: there is a stream produced from the information obtained from the self-driving car's sensors, such as the location of pedestrians and other vehicles at a distance considered relevant, there direction and speed of movement, closed streets, traffic signals, etc. This setting is represented and modeled in AM_t. The system also has assimilated knowledge, such as the location of each street, the direction of circulation, the location of traffic lights, among others; this knowledge is contained in AM_p.
Meanwhile, the information provided by the LIDAR is affected by uncertainty—e.g., its reliability is influenced by climatic conditions such as the presence of heavy fog or rain. This could be represented by three random variables that indicate the likelihood that the LIDAR delivers information categorized as a false positive, the likelihood that there is fog at a given time, and the likelihood that there is heavy rain. One way to model this is with a Bayesian network as shown in Figure 2; this knowledge is part of EM_p. Finally, information is available from random events but has limited validity over a period of time; for example, information that there is actually fog at the current position of the vehicle—this is part of EM_t.
In Figure 3 we present a simple example of the content of AM_t and AM_p, together with an annotation function. This model makes the assumption that EM_t and EM_p contain the following atoms: *FN-lidar* (probability of LIDAR producing a false negative), *FP-radar* (probability of radio wave-based radar produces a false positive), *Rain* (probability of heavy rain), and *Fog* (probability of fog). Suppose the car perceives the presence of fog; in this case, the variable *Fog* is instantiated in EM_t with value *true*. As can be seen in Figure 3, there are two contradictory atoms in AM_t: *pedestrian_ahead* and *~pedestrian_ahead.* However, their annotations are different; the first one exists in the worlds in which the radar does not deliver a false positive, while the second one holds in the worlds where the LIDAR does not deliver a false negative. Thus, the system should decide which is most likely according to the probability that a false negative from the LIDAR has occurred (taking into account that there is fog) and the probability of a false positive from the radar.

3. Discussion and Future Work

In this work we have presented and discussed the difficulties and challenges involved in the generation and processing of knowledge that is available in the form of a *stream*; i.e., information with semantic value in contexts where large volumes of data are produced with high frequency. The main

contribution is the proposal of a model as a starting point for stream reasoning, using argumentation techniques to take advantage of the semantic content by leveraging its capacities to handle contradictory information and to provide explanations for the answers provided, which can be very useful or even necessary in certain domains. The novelty of the proposal is that until now there are no developments combining stream processing, structured argumentation, belief revision, and reasoning under uncertainty.
As part of this line of research, current and future work involves the development of algorithms that integrate existing tools for stream and complex events processing with argumentation systems in an efficient way, since computational tractability is a crucial aspect for its effectiveness. Another main objective is the development of belief revision operators that are suitable for environments with highly changing and uncertain knowledge, which will also depend on the proper use of efficient probabilistic models.

Acknowledgments. This work was supported by funds provided by CONICET, Agencia Nacional de Promoción Científica y Tecnológica, and Universidad Nacional del Sur, Argentina. Some of the authors of this work were also supported by the U.S. Department of the Navy, Office of Naval Research, grant N00014-15-1-2742. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Office of Naval Research.

References

1. Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. Journal of symbolic logic 50(2):510–530 (1985)
2. Arasu, A., Babu, S., Widom, J.: The CQL continuous query language: Semantic foundations and query execution. VLDB Journal 15(2), 121–142 (2006)
3. Barbieri, D.F., Braga, D., Ceri, S., Della Valle, E., Grossniklaus, M.: C-SPARQL: A continuous query language for RDF data streams. International Journal of Semantic Computing 4(1):3–25 (2010)
4. Cugola, G., Margara, A.: Processing flows of information: From data stream to complex event processing. ACM Computing Surveys 44(3):15 (2012)
5. Della Valle, E., Ceri, S., Harmelen, F.v., Fensel, D.: It's a streaming world! Reasoning upon rapidly changing information. IEEE Intell. Sys. 24(6):83–89 (2009)
6. Deyle, T.: Velodyne HDL-64E laser rangefinder (LIDAR) pseudodisassembled. Online blog: http://www.hizook.com/blog/2009/01/04/velodyne-hdl-64e-laser-rangefinder-lidar-pseudo-disassembled (2009)
7. Domingos, P., Kok, S., Lowd, D., Poon, H., Richardson, M., Singla, P.: Markov Logic, pp. 92–117. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
8. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. Artificial Intelligence 77(2):321–357 (1995)

9. Falappa, M.A., Kern-Isberner, G., Reis, M.D., Simari, G.R.: Prioritized and non-prioritized multiple change on belief bases. Journal of Philosophical Logic 41(1), 77–113 (2012)

10. Gal, A., Wasserkrug, S., Etzion, O.: Event processing over uncertain data. In: Reasoning in Event-Based Distributed Systems, pp. 279–304. Springer (2011)

11. García, A.J., Simari, G.R.: Defeasible logic programming: An argumentative approach. Theory and practice of logic programming 4(1–2):95–138 (2004)

12. Gärdenfors, P.: Belief revision, vol. 29. Cambridge University Press (2003)

13. Guizzo, E.: How Google's self-driving car works. IEEE Spectrum Online, October 18 (2011)

14. Hansson, S.: Belief Base Dynamics. Uppsala University. Ph.D. thesis, (1991)

15. Hansson, S.: Semi-revision. J. Appl. Non-Class. Log. 7(1-2), 151–175 (1997)

16. Hansson, S.O.: Taking belief bases seriously. In: Logic and philosophy of science in Uppsala, pp. 13–28. Springer (1994)

17. Martinez, M.V., García, A.J., Simari, G.R.: On the use of presumptions in structured defeasible reasoning. In COMMA, pp. 185–196 (2012)

18. Moguillansky, M.O., Rotstein, N.D., Falappa, M.A., García, A.J., Simari, G.R.: Dynamics of knowledge in DeLP through argument theory change. Theory and Practice of Logic Programming 13(6):893–957 (2013)

19. Moguillansky, M.O., Rotstein, N.D., Falappa, M.A., García, A.J., Simari, G.R.: Argument theory change through defeater activation. In: COMMA, pp. 359–366 (2010)

20. Nilsson, N.J.: Probabilistic logic. Artificial Intelligence 28(1):71–87 (1986)

21. Pearl, J.: Probabilistic reasoning in intelligent systems: Networks of plausible inference. (1988)

22. Rotstein, N.D., Moguillansky, M.O., García, A.J., Simari, G.R.: A dynamic argumentation framework. In COMMA, pp. 427–438 (2010)

23. Shakarian, P., Simari, G.I., Moores, G., Paulo, D., Parsons, S., Falappa, M.A., Aleali, A.: Belief revision in structured probabilistic argumentation: Model and application to cyber security. Annals of Mathematics and Artificial Intelligence 78(3–4):259–301 (2016).

24. Simari, G.I., Shakarian, P., Falappa, M.A.: A quantitative approach to belief revision in structured probabilistic argumentation. AMAI 76(3–4):375–408 (2016)

Immune Algorithm for Solving the Dynamic Economic Dispatch Problem

VICTORIA S. ARAGÓN[1, 2], SUSANA C. ESQUIVEL[1]

[1]Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)
Universidad Nacional de San Luis
Ejército de los Andes 950 - (5700) San Luis, ARGENTINA
[2]CONICET

Abstract. In this paper, an algorithm inspired on the immune system is presented, IA DED stands for Immune Algorithm Dynamic Economic Dispatch, it is used to solve the Dynamic Economic Dispatch problem. IA DED uses as differentiation process a redistribution power operator and the outputs power are integer values. The proposed approach is validated using three problems taken from the specialized literature. Our results are compared with respect to those obtained by several other approaches.

Keywords: Artificial immune systems, dynamic economic dispatch problem, metaheuristics.

1. Introduction

One of the early problems in power system optimization is the Dynamic Economic Dispatch (DED) problem. Its main objective is to determine the optimal schedule of output powers of on line generating units, over a certain period of time (T intervals), to meet power demands at minimum operating cost [10]. Besides, several constraints associated to the system have to be satisfied, such as load demands, ramp rate limits, maximum and minimum limits, and prohibited operating zones.

As DED problem is nonlinear different heuristics have been used to solve it, such as, artificial immune system [8], genetic algorithm [7], particle swarm optimizer [7], algorithm bee colony [7], harmony search [14], [12], [2], imperialist competitive algorithm [11], an hybridized differential evolution [3], among others. Surveys about these techniques can be found in [16] and [9].

In this paper, we propose an algorithm to solve DED problem which is inspired on the immune system. Considering T intervals, the problem is regarded as a sequence of T problems. But, each problem (at time i) depends on its predecessor (at time $i - 1$) and it conditions to its successor (at time $i + 1$). The algorithm applies a redistribution power operator in order to improve a solution at time t with the aim of keeping such a solution feasible at a low computational cost.

The remainder of this paper is organized as follows. Section 2 defines the DED problem. In Section 3, we describe our proposed algorithm. In Section 4, we present the test problems used to validate our proposed approach and parameters settings. In Section 5, we present our results and we discuss and compare them with respect to other approaches. Finally, in Section 6, we present our conclusions and some possible paths for future research.

2. Problem Formulation

The DED problem minimizes the total production cost (TC) associated with N dispatch units for a time period:

$$TC = \sum_{t=1}^{T} \sum_{i=1}^{N} F_i\ (P_i^t) \qquad (1)$$

where *TC* is the fuel cost over the whole dispatch period, *T* is the number of intervals in the period, *N* is the number of generating units in the system, P_i^t is the power of i^{th} unit at time t (in MW) and F_i is the fuel cost for the i^{th} unit (in \$/h).
A smooth fuel cost function can be expressed as a single quadratic function:

$$F_i(P_i^t) = a_i\ (P_i^t)2 + b_i\ P_i^t + c_i \qquad (2)$$

where a_i, b_i and c_i are the fuel consumption cost coefficients of the i^{th} unit. But, if the valve-point effects are taking into account, the fuel cost function of the i^{th} unit is expressed as the sum of a quadratic and a sinusoidal function in the form:

$$F_i(P_i^t) = a_i\ (P_i^t)2 + b_i\ P_i^t + c_i + |\ e_i \sin(\ f_i(P_{mini} - P^t_i\ ))\ | \qquad (3)$$

where e_i and f_i are the fuel cost coefficients of the ith unit with valve-point effects.
Regardless the function considered (Eq. 2 or Eq. 3), its minimization is subjected to:

1. Power Balance Constraint: the power generated has to be equal to the power demand required. It is defined as:

$$\sum_{i=1}^{N} P_i^t - P_D^t - P_L^t = 0 \qquad (4)$$

where t = 1, 2, . . . , *T*. P^t_D is the total power demand at time *t*, and P^t_L is the transmission power loss at time *t* (in MW). P^t_L is calculated using the B-matrix loss coefficients, and the general form of the loss formula using B-coefficients is:

$$P_{L,t} = \sum_{i=1}^{N} \sum_{j=1}^{N} P_i^t\ B_{ij}\ P_j^t + \sum_{i=1}^{N} B0_i\ P_i^t + B00 \qquad (5)$$

If transmission power loss is not considered, $P_{L,t}=0$.

2. Operating Limit Constraints: units have physical limits about the mínimum and maximum power they can generate:

$$P_{mini} \leq P_i^t \leq P_{maxi} \quad (6)$$

where P_{mini} and P_{maxi} are the minimum and maximum power output of the i^{th} unit in MW, respectively.

3. Ramp Rate Limits: they restrict the operating range of all on-line units. Such limits indicate how quickly the unit's output can be changed:

$$\begin{cases} P_j^t - P_j^{(t-1)} \leq UR_j \ if \ P_j^t > P_j^{(t-1)} \\ P_j^{(t-1)} - P_j^t \leq DR_j \ if \ P_j^t < P_j^{(t-1)} \end{cases} \quad (7)$$

where $P_j^{(t-1)}$ is the output power of the j^{th} unit at previous hour and, UR_j and DR_j are the ramp-up and ramp-down limits of the j^{th} unit in MW, respectively. Due to ramp-rate constraints, Eq. 6 is modified as:

$$\max(P^t_{minj}, P_j^{(t-1)} - DR_j) \ d \ P^t_j \ d \ \min(P^t_{maxj}, P_j^{(t-1)} + UR_j) \quad (8)$$

such that

$$\begin{cases} P^t_{minj} = P_{minj}, \ P_j^{(t-1)} - DR_j \\ P^t_{maxj} = P_{maxj}, P_j^{(t-1)} + UR_j \end{cases} \quad (9)$$

4. Prohibited Operating Zones: they restrict the operation of the units due to steam valve operation conditions or to vibrations in the shaft bearing. Thus, an unit with prohibited operating zones has a discontinuous input-output power generation characteristic which gives rise to additional constraints on the unit operating range.

$$\begin{cases} P_{mini} \leq P_i^t \leq PZ^L_{i,1} \\ PZ^U_{i,k-1} \leq P_i^t \leq PZ^L_{i,k} \ k = 2, 3, .., n_i \\ PZ^U_{i,n1} \leq P_i^t \leq P_{maxi} \end{cases} \quad (10)$$

where n_{ji} is the number of prohibited zones of the i^{th} unit, k is the index of the prohibited operating zones of the i^{th} unit. $PZ^L_{i,k}$ and $PZ^U_{i,k}$ are the lower and upper bounds of the k^{th} prohibited operating zones of unit i.

3. Our Proposed Algorithm

In this paper, an adaptive immune system model based on the immune responses mediated by the T cells from the immune system is presented. These cells present special receptors on their surface called T cell receptors (TCR), they are responsible for recognizing antigens bound to major histocompatibility complex (MHC) molecules.) [13].
The model considers some processes that T cells suffer. These are proliferation (to clone a cell) and differentiation (to change the clones so that they acquire specialized functional properties); this is the so-called activation process. IA DED (Immune Algorithm for Dynamic Economic Dispatch problem) is an adaptation of an algorithm inspired on the activation process [1], which was proposed to solve the economic dispatch problem. IA DED operates on one population which is composed of a set of T cells.

For each cell, the following information is kept:

1. TCR: it identifies the decision variables of the problem (TCR $\in \Re^N$). Each thermal unit is represented by one decision variable, each variable is encoded
   by an integer value.
2. objective: objective function value for TCR, (TC(TCR)).
3. prolif: it is the number of clones that will be assigned to the cell, it is N for all problems.
4. differ: it is the number of decision variables that will be changed when the differentiation process takes place (if applicable). This level is calculated as U(1,N).
5. TP: it is the power generated by TCR ($\sum_{i=1}^{N} TCR_i$).
6. P'_L: it is the transmission loss for TCR TCR, according to Eq. 5.
7. ECV: it is the equality constraint violation for TCR. At *t* time (|TP - P_D -P_L|). If ECV > 0, then the power generated is bigger than the demanded power, and if ECV < 0 then the power generated is lower than the required power.
8. ICS: it is the inequality constraints sum, $\sum_{i=1}^{N} \sum_{i=1}^{ni} poz(TCR_i, i, j)$

$$poz(p,i,j) = \begin{cases} min\left(p - P_{i,j}^{L}, P_{i,j}^{U} - p\right) if\ p \in \left[P_{i,j}^{L}, P_{i,j}^{U}\right] \\ 0\ otherwise \end{cases}$$

   where n_i is the number of prohibited operating zones and [$PZ^{L}_{i,j}$, $PZ^{U}_{i,j}$] is the j^{th} prohibited range for the thermal i^{th} unit.
5. feasible: it indicates if the cell is feasible or not. A cell is considered as feasible if: 1) ECV = 0 for problems without transmission network loss and 0 d ECV < ε for problems with transmission loss and 2) ICS = 0 for problems which consider prohibited operating zones.

Differentiation for feasible cells - Redistribution Process

The idea is to take a value (called d) from one unit (say i) and assign it to another unit (say j). i^{th} and j^{th} units are modified according to: $cell.TCR_i = cell.TCR_i - d$ and $cell.TCR_j = cell.TCR_j + d$, where $d = U(1, (int)](P_1 * min(cell.TCR_i - P^t_{mini}, P^t_{maxj} - cell.TCR_j)))$, $U(w_1,w_2)$ refers to a random number with a uniform distribution in the range (w_1,w_2) and P_1 is a change factor ($P_1 \in [0, 1]$). Besides, a probability, P_2, determines if ith and jth units will be modified.

Differentiation for infeasible cells

For infeasible cells, the number of decision variables to be changed is determined by their differentiation level (differ). Each variable to be changed is chosen in a random way and it is modified according to: cell.TCR2$_i$ = cell.TCR$_i$ ± m, where cell.TCR $_i$ and cell.TCR2$_i$ are the original and the mutated decision variables, respectively. m = U(1, (int)](cell.ECV + cell.ICS)). In a random way, it decides if m will be added or subtracted to cell.TCR$_i$. If the procedure cannot find a TCR2$_i$ in the allowable range, then a random number with a uniform distribution is assigned to it (cell.TCR2$_i$ = U(cell.TCR$_i$, P^t_{maxi}) if m should be added or cell.TCR2$_i$ = U(P^t_{mini}, cell.TCR$_i$), otherwise). If the resulting clon is feasible then differentiation process stops, otherwise, the process is applied to the resulting clon instead the original infeasible cell. This methodology follows until prolif differentiations have been applied or a feasible clon has been reached.

Algorithm 1 IA_DED Algorithm

```
1: P ← Initialize_Population();
2: Evaluate_Constraints(P);
3: Evaluate_Objective Function(P);
4: for t ← 1 to T do
5:     top ← 0;
6:     while A predetermined number of evaluations has not been reached and
top <
       5 * 10^7 do
7:         Proliferation_Population(P);
8:         Differentiation_Population(P);
9:         top + +;
10:    end while
11:    bestt ← Search best at Population(P, t);
12:    t + +;
13:    Update_limits(bestt);
14:    Repair_output power(P);
15:    Update_output power(P);
16:    Evaluate_Constraints(P);
17:    Evaluate_Objective_Function(P);
18: end for
19: BEST ← (best_1, best_2, . . . , best_T );
```

The algorithm works in the following way (see Algorithm 1). First, the TCRs are randomly initialized within the limits of the units (Step 1). Then, ECV and ICS are calculated for each cell (Step 2). Only if a cell is feasible, its objective function value is calculated (Step 3). Next, the following steps are repeated T times (Step 5 to 17): while a predetermined number of objective function evaluations had not been reached and $5*10^7$ iterations are not performed the cells are proliferated and differentiated according to their feasibility. After activation process, best solution at t time is recorded. The time is increased and new operational limits are updated according Eq. 3. Those units which outputs power falling out the new operational limits are changed by random values from the valid limits. Finally, (Step 19) the whole final solution is sequence of solutions found in time 1, time 2, to time T.

4. Validation

IA_DED performance was validated with three test problems, 5-unit system [11], 15-unit system [5], 54-unit system [4]. Table 1 provides their most relevant characteristics and the maximum number of function evaluations. IA DED was implemented in Java (version 1.6.0 24) and the experiments were performed in an Intel Q9550 Quad Core processor running at 2.83GHz and with 4GB DDR3 1333Mz in RAM.

The required parameters by IA DED are: size of population, maximum number of objective function evaluations, change factor (P1) and probability for redistribution operator (P_2). To analyze the effect of the first and third parameters on IA DED's behavior, we tested it with different parameters settings. Some preliminary experiments were performed to discard some values for the population size parameter. Hence, the selected parameter levels were: a) Population size (C) has four levels: 5, 10 and 50 cells, b) Probability P_1 has three levels: 0.1, 0.5 and 0.9 and c) Probability P_2 has two levels: 0.01 and 0.1.

Thus, we have 18 parameters settings for three problems. They are identified as C<size>-P1<Prob>-P2<Prob>, where C, P1 and P2 indicate the population size and the probabilities, respectively. For each problem, 100 independent runs were performed.The box plot method was selected to visualize the distribution of the objective function values for each power system. This allowed us to determine the robustness of our proposed algorithm with respect to its parameters. Figure 1 show in the x-axis the parameter combinations and the y-axis indicates the objective function values for each problem. We can see for 5-unit system and 15-unit system the results are robust. For 54-unit system, to increase the change factor improve the results and to increase the probability of application of the redistribution process and size population deteriorate the results. So, the settings parameters were used to compare the results got by IA DED with those produced by other approaches are: for 5-unit system C=10 and P_1=0.1, P_2=0.1, for 15-unit system C=50 and P_1=0.9, P_2=0.1, for 54-unit system C=5 and P_1=0.9,

P_2=0.01. Also, we set ε=2.0 for those problems which consider loss transmission.

Table 1. Test Problems Characteristics

Problem	Thermal Units	Objective	P_L	Prohibited Zones	P_D (MW)	Evaluations
5-unit system	5	non-smooth	Yes	No	14577	19000
15-unit system	15	smooth	Yes	No	60981	19000
54-unit system	54	non-smooth	No	yes	111600	30000

5. Comparison of Results and Discussion

Several methods are compared with IA DED. They are listed next with the maximum number of function evaluations performed (if the value was available): AIS [8] (300000), GA [7] (not found), PSO [7] (not found), ABC [7] (not found), HS [14] (50000), ICA [11], for 5-unit system, (20000), DE-SQP [3] (50000), NPAHS. Additionally, the problem dimensionality does not seem to affect the performance of our proposed approach either. For problems which consider transmission loss, rate ramp limits and prohibited zones, SYS_6U_a and SYS_15U, the standard deviations increase with the problem dimensionality. For the only problem which considers transmission loss but not rate ramp limits or prohibited zones, SYS_20U, the standard deviation is lower than SYS_15U's standard deviation. [12] (50000) CSADHS [2] (250000), SAMF [6] (not indicated), OCD [15] (not indicated), ICA [11] (80000), for 54-unit system.

Table 2 shows: the best, worst, mean, standard deviation and running times obtained by the approaches. For IA DED only four decimal digits are shown due to space restrictions. For all the test problems, our proposed IA DED found feasible solutions in all the runs performed.

Table 2. Comparison of results. The best values are shown in **boldface**. - denotes that the value was not available in the literature.

Problem/ Algorithm	Best	Worst	Mean	Std.	Time(s)
5-unit system					
IA_DED	43716.6386	46611.1333	44879.9749	649.60	**2.228s**
AIS[8]	44385.43	45553.7707	44758.8363	-	4min
GA[7]	44862.42	-	-	-	-
PSO[7]	44253.24	-	-	-	-
ABC[7]	44045.83	-	-	-	-
HS[14]	44376.23	-	-	-	2.8min
ICA[11]	43117.055	**43209.533**	**43144.472**	19.821	-
DE-SQP[3]	**43161**	-	-	-	-
15-unit system					
IA_DED	**759385.9148**	**759665.1300**	**759478.5852**	55.08	**2.08s**
NPAHS[12]	759603.089	759988.390	759779.467	-	250.0
CSADHS[2]	759689.220	759845.739	759766.233	-	3.36min
SAMF[6]	759406.42	-	-	-	2.951s
54-unit system					
IA_DED	**1718177.0643**	**1718695.4568**	**1718424.5910**	130.85	8.565s
OCD[15]	1772724.032	**-**	-	-	**0.132s**
ICA[11]	1807081.174	1811388.285	1809664.219	-	-

The running time of each algorithm is affected by both the hardware environment and the software environment. That is the reason why the main comparison criterion that we adopted for assessing efficiency was the number of objective function evaluations performed by each approach. For having a fair comparison of the running times of all the algorithms considered in our study, they should all be run in the same software and hardware environment (something that was not possible in our case, since we do not have the source code of several of them). Clearly, in our case, the emphasis is to identify which approach requires the lowest number of objective function evaluations to find solutions of a certain acceptable quality.

However, the running times are also compared in an indirect manner, to give at least a rough idea of the complexities of the different algorithms considered in our comparative study. Analyzing Table 2, IA DED, for 5-unit system, is outperformed by ICA [11] and DE-SQP [3], but it can found quickly an acceptable solution performing less evaluations of objective function. For 15-unit system, IA DED outperformed all approaches which we compare it, it found the best solution both running time and total fuel cost. For 54-unit system, IA DED outperformed all approaches which we compare it, taking into account the total fuel cost. It needs 8.565s to find this solution, however it costs $54547 less than OCD's solution.

6. Conclusions and Future Work

This paper presented an adaptation of an algorithm inspired on the T-Cell model of the immune system, called IA DED, which was used to solve dynamic economic dispatch problems. IA DED is able to handle the five types of constraints that are involved in this kind of problems: power balance constraint with and without transmission loss, operating limit constraints, ramp rate limit constraint and prohibited operating zones, and different types of objective function: smooth and non-smooth.

At the beginning, the search performed by IA DED is based on a simple differentiation operator which takes an infeasible solution and modifies some of its decision variables by taking into account their constraint violation. Once the algorithm finds a feasible solution, a redistribution power operator is applied. This operator modifies two decision variables at a time, it decreases the power in one unit, and it selects other unit to generate the power that has been taken, always integer values.

The approach was validated with three test problems having different characteristics and comparisons were provided with respect to some approaches that have been reported in the specialized literature. Our proposed approach produced competitive results in all cases, being able to outperform some approaches while performing a lower number of objective function evaluations.

As future work, we are interested in redesigning the redistribution operator in order to maintain the solutions' feasibility when a problem involves prohibited operating zones.

Acknowledgements. The authors acknowledge support from Universidad Nacional de San Luis, Project no. P330214/22F435.

References

1. V.S. Aragon, S.C. Esquivel, and C.A. Coello Coello. An immune algorithm with power redistribution for solving economic dispatch problems. Information Sciences, 295(0):609 – 632, 2015.
2. R. Arul, G. Ravi, and S. Velusami. Chaotic self-adaptive differential harmony search algorithm based dynamic economic dispatch. International Journal of Electrical Power & Energy System, 50:85–96, 2013.
3. A.M. Elaiw, X. Xia, and A.M. Shehata. Hybrid de-sqp and hybrid pso-sqp methods for solving dynamic economic emission dispatch problem with valve-point effects. Electric Power Systems Research, 103:192–200, October 2013. ISSN 0378-7796.
4. M.Shahidehpour [Online]. Available from: motor.ece.iit.edu/data/SCUC 118test.xls, March 2009. [accessed July 2016].

5. Z.L. Gaing and Ting-Chia Ou. Dynamic economic dispatch solution using fast evolutionary programming with swarm direction. In 4th IEEE Conference on Industrial Electronics and Applications. ICIEA 2009, pages 1538–1544, 2009.
6. S. GAnesan and S. Subramani. Dynamic economic dispatch based on simple algorithm. International Journal of Computer and Electrical Engineering, 3(2):1793–8163, 2011.
7. S. Hemamalini and S. Simon. Dynamic economic dispatch using artificial bee colony algorithm for units with valve-point effect. European Transactions on Electrical Power, 21:70–81, 2011.
8. S. Hemamalini and S. Simon. Dynamic economic dispatch using artificial immune system for units with valve-point effect. International Journal of Electrical Power & Energy System, 33:868–874, 2011.
9. C. Kumar and T. Alwarsamy. Dynamic economic dispatch - a review of solution methodologies. European Journal of Scientific Research, 64(4):517–537, 2011.
10. R. Kumar, D. Sharma, and A. Sadu. A hybrid multi-agent based particle swarm optimization algorithm for economic power dispatch. International Journal of Electrical Power and Energy Systems, 33(1):115–123, 2011.
11. B. Mohammadi-Ivatloo, A. Rabiee A., Soroudi M., and Ehsan. Iteration PSO with time varying acceleration coefficients for solving non-convex economic dispatch problems. International Journal of Electrical Power & Energy Systems, 42(1):508 – 516, 2012.
12. Qun Niu, Hongyun Zhang, Kang Li, and George W. Irwin. An efficient harmony search with new pitch adjustment for dynamic economic dispatch. Energy, 65:25–43, 2014.
13. Leandro Nunes de Castro and Jonathan Timmis. Artificial Immune Systems: A New Computational Intelligence Approach. Springer-Verlag, New York, 2002.
14. V.R. Pandi and B.K. Panigrahi. Dynamic economic load dispatch using hybrid swarm intelligence based harmony search algorithm. Expert Systems with Applications, 38:8509–8514, 2011.
15. Abbas Rabiee, Behnam Mohammadi-Ivatloo, and Mohammad Moradi-Dalvand. Dynamic economic dispatch: A review. IEEE Transactions on Power Systems, 29(2):982–983, 2014.
16. X. Xia and A. M. Elaiw. Dynamic economic dispatch: A review. The Online Journal on Electronics and Electrical Engineering (OJEEE), 2(2):234–245, 2009.

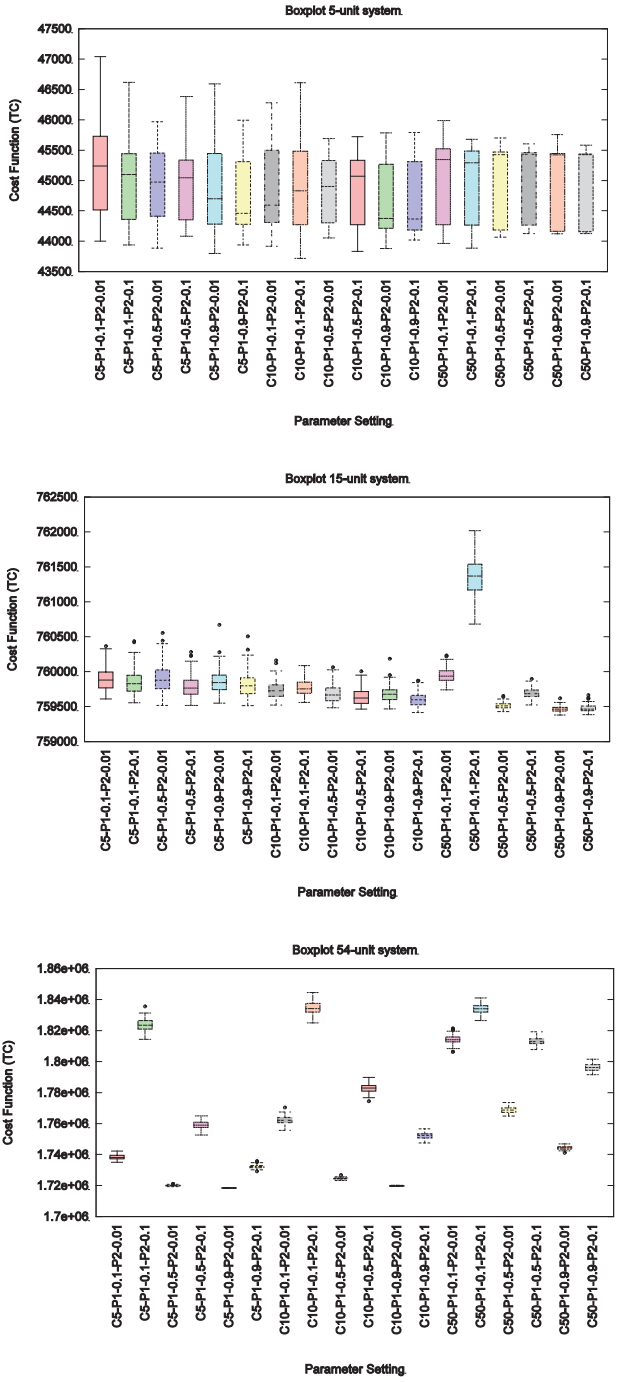


Fig 1. Box plots for the test problems with the best parameters combination

Optimization of addition chains

FERNANDO AQUINO[1], GUILLERMO LEGUIZAMÓN[2]

[1] Universidad Tecnológica Nacional – Facultad Regional Concepción del Uruguay
Fernando.aquino@hotmail.com.ar
[2] Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC)
Universidad Nacional de San Luis
legui@unsl.edu.ar

Abstract. The work addresses the problem of optimal computation of addition chains, widely studied with different methods and approaches (both deterministic and stochastic) being also a problem of interest in the field of cryptography. In this paper, we propose the use of the Grey Wolf Algorithm (GWO) to deal with this problem in order to compare the results obtained with other approaches of the state of the art. Although the problem in question has been dealt with by different strategies and for different types of exponents, this proposal particularly focuses on the optimization of addition chains associated with exponents of moderate size.

1. Introduction

Asymmetric cryptographic algorithms use modular exponentiation to encrypt and decrypt data. The computational cost resulting from the large number of multiplication operations, which involves resolving exponents of a considerable order number, can be solved by using the concept of "addition chains" [2]. However, an exponent does not have a single valid addition chain associated and therefore finding a valid chain and having a minimum length is a NP-hard problem.
Modular exponentiation [8] is called the operation whose purpose is to obtain a value b (integer, positive) to satisfy the following equation:

$$b = A^e \bmod P \quad (1)$$

Let A be an integer value, arbitrary in the range [1, P-1] and be *e* an integer number, positive and arbitrary, modulating exponentiation is defined as the problem of finding a single positive integer value b in the range [1, P-1], capable of satisfying equation (1), being for this problem (1) the objective function to be minimized in order to obtain a minimum b value (i.e., a minimum length addition string).

The inherent difficulty is that as the exponent (*e*) is increased, the number of operations to be performed to solve equation (1) increases, which requires an alternative to optimize the process.
A possible alternative is to use addition chains in order to reduce the number of multiplications to solve a exponentiation operation. Example: Given the exponent e = 91, instead of multiplying the base 91 times it is feasible to have a chain associated with the number 91, such as: C = {1, 2, 4, 6, 8, 14, 22, 36, 58, 80, 88, 90, 91} and then solve the exponent would consist of multiplying the base by each of the exponents in C, i.e., for this example only 13 operations.
However, considering that there is no a single chain for an exponent, finding a minimum length chain is a combinatorial optimization problem and can be addressed using a metaheuristic to avoid testing by brute force to find an optimal value (minimum length chain).

1.1 Background

In this section we provide a summary of the state of the art of the problem under study in the field of metaheuristics. Table 1 summarizes the results of previous studies in which the optimization of the objective function defined by equation (1) was approached through the use of various well-known metaheuristics, chronologically ordered by year and indicating ranges of the exponent (*e*) used for each approach:

Table 1. Summary of the state of the art of the problematic object of study

Applied metaheuristics	**ACO [7]**	**AG [8]**	**AIS [13]**	**PSO [14]**	**EP [17]**	**Hybrid GA [22]**	**AG [23]**
Year	2004	2005	2008	2009	2011	2011	2016
e μ	Exponent Ranges						
< 128	X	X	X	-	X	X	X
[128,512]	X			-	X	-	X
[512,1024]	X			-	X	-	

Considering that the present proposal focuses on the use of GWO metaheuristic as an alternative strategy for the problem of generation of

addition chains, it is interesting to compare with proposals of the state of the art also based on metaheuristics.

Nedjah and De Macedo-Mourelle in 2004 experimented with ACO [7] to solve the problem of finding chains of optimal length, working with exponents up to a length of 128-bits without addressing tests with exponents of greater length. However, in a later work they applied ACO to exponents of up to 1024-bits. In 2005, Cruz-Cortés and others used GA [8] to find optimal chains for exponents of all ranges less than 4096 (and < 4096). In 2008, Rodríguez-Henríquez and Coello Coello proposed AIS [13], obtaining minimum chains for (e) in [1,4096], being possible to obtain results for all ranks but better for those of limited length in [128,512], leaving proposed as future work achieve better results for larger exponents. León-Javier in 2009 implemented PSO [14] but only reported experiments with small exponents, leaving the other cases as pending work. Isidro Domínguez in 2011 in his thesis adapted the algorithm EP [17] achieving to minimize chains for exponents of the ranges 128, 256, 512, and 1024 bits. The same year a hybridization alternative of GA with Simulated Annealing [22] was proposed improving with this the convergence of the GA, but nevertheless the results do not surpass previous proposals as the case of PSO with which its performance was compared. Finally in 2016 an GA was proposed with a series of improvements in the operators of variation, a novel representation of solutions and a solution repair strategy [23]. Optimized exponents in the range $[2^{37} - 3, 2^{127} - 3]$.

2. Algorithm based on the behavior of grey wolves (Grey Wolf Optimizer)

In [21] is proposed a metaheuristic dominated Grey Wolf Optimizer (GWO[1]), inspired by the behavior of grey wolf packs (Canis Lupus) and its social organization for hunting prey.

Grey wolves by nature prefer to live in a group of between 5 and 12 members on average. The herd is made up of two leaders called alpha (one male and one female) and the rest of the herd. In this social structure, alpha members are responsible for making decisions about hunting, place to sleep, time to wake up, and the rest of the members must follow their orders.

The alpha member of a herd should not necessarily be the strongest but the best in terms of group leadership and decision-making ability.

[1] For reasons of international nomenclature in the field of metaheuristics will be maintained through the article the acronym GWO.

The second level in the hierarchy of the group are the beta members (subordinates of the alpha), who must collaborate with the decision making and the fulfillment of the orders of the alpha. These members are the candidates to become alpha when one of these dies or ages. There is another classification that is found below in the hierarchy of the flocks of grey wolves and is called omega, these must be submitted to the alpha and beta members obeying them, in many cases usually it is seen in this type to members who officiate of babes of the hatchlings, within the herd [24].
The main stages of hunting in grey wolf commands are as follows:

- To follow, persecute, and approach the prey
- To chase, to surround, and to harass the prey until it "paralyzes".
- To attack the prey.

This hunting technique was modeled mathematically in [21] with the purpose of proposing an optimization method and to use it in the resolution of complex problems.

2.1 The mathematical model and the algorithm:

The mathematical model of the social hierarchy of grey wolf herds is described below based on the 3 main actions related to the activity of the hunt: to follow, surround, and attack.

Social hierarchy of the herds:

In order to mathematically model the social hierarchy of wolves in the GWO design, it is considered that the solution most suitable for the problem is the alpha (±). Consequently, the second and third best solutions are called beta (²) and delta (´), respectively. The rest of the candidate solutions are assumed to be omega (É). In the GWO algorithm the hunting activity is translated as an optimization process, and it is guided by ±, ², ´, whereas the wolves É follow the other 3 types.

Barracks of the prey:

The grey wolves surround the prey during the hunt. Mathematically this behavior is modeled as follows:

$$D = |C. Xp(t) - X(t)| \quad (2)$$

$$X(t+1) = Xp(t) - A . D$$

Where t denotes the current iteration, A and C are vectors of coefficients, Xp is the position vector of the prey, and X indicates the vector position of a given grey wolf. On the other hand, vectors A and C are calculated as follows:

$$A = 2\,a.\,r1 - a \tag{3}$$

$$C = 2\,.\,r2$$

According to the original formulation of GWO [21] the components of vector A decrease linearly from 2 to 0 over the course of the iterations and r1, r2 are random vectors in the range [0,1].
The mathematical model of GWO describes the approach operation as a grey wolf in a certain position (in the n-dimensional space) can update its position with respect to the prey, moving the location of the wolf (or agent) through the adjustment of the value of vectors A and C.

Hunting (prey attack):

Grey wolves have the ability to recognize the prey location and surround them. In nature hunting is usually guided by the alpha member, although the beta and delta members who follow him in hierarchy, may eventually participate in hunting.
In order to mathematically simulate hunting behavior of grey wolves, we assume that the alpha member (best candidate solution), beta and delta have a better knowledge about the possible location of the prey. Therefore, we keep the first three best solutions obtained so far and force the other search agents (including the omega members) to update their positions according to the position of the best search agents. In this sense the following formulas are proposed:

$$D\pm = |C1.\,X\pm - X|,\; D^{2} = |C2.\,X^{2} - X|,\; D' = |C3.\,X' - X| \tag{4}$$

$$X1 = X\pm - A1.\,(D\pm),\; X2 = X^{2} - A2.\,(D^{2}),\; X3 = X' - A3.\,(D')$$

$$X(t+1) = (X1 + X2 + X3)/3$$

As each agent updates its position according to: alpha, beta, and delta in a search space, it can be said that the role of alpha, beta, and delta is to estimate the position of the prey and the other wolves update their positions at random around the same prey.

Exploration (search for the prey) vs. exploitation (prey attack) in GWO

The wolves approach the prey and attack it when it stops moving. The mathematical model of such behavior approaching the prey, is translated as the decrease in the value of vector A (coefficient). It should be noted that the fluctuation range of vector A is also reduced by the value of "a" (see equation 3).
When | A | <1, GWO forces the wolves to attack the prey. It should be noted that this feature is related to the *exploitation* of the algorithm.
GWO is prone to stagnation in local solutions with these operators. Although it is true that the trap mechanism generates exploration to some extent, it needs more operators to emphasize such exploration.
Wolves have different positions with respect to the prey but they converge to the same position when attacking the prey. The mathematical model of this divergence is reflected in the use of random values for the vector A (greater than 1 and less than -1) to ensure divergence and thus the search better encompasses the search space (this feature is related to *exploration*, that is when | A | > 1 is forcing the wolves to diverge from the prey).
In summary, the search process in GWO begins with the creation of population of random wolves (candidate solutions). In the course of iterations, the wolves: alpha, beta, and delta, estimate the probable position of the prey. Each candidate solution updates its distance from the prey. The parameter "a" is decreased from 2 to 0 in order to emphasize the exploration and exploitation, respectively. The candidate solutions diverge with respect to the prey when | A | > 1 and converge towards the prey when | A | < 1. Finally, the GWO algorithm stops by satisfying an end criterion.
As described in [21] GWO has been applied to a number of optimization problems and within them was applied to a real problem in the field of optical engineering [21].

The general scheme of the algorithm:

Initialize population of wolves Xi (with i = 1, 2, ..., n)
Initialize ±, A and C
Calculate fitness for each search agent
X± = the best agent.
X² = the second best agent.
X´ = the third best agent.
While (t < maximum number of iterations) For i = [1..n]
Update the position of each agent using equation (7)
End For
Update ±, A and C
Calculate the fitness of all search agents
Update X±, X² and X´
t=t+1
End While
Return X±

Algorithm 1: Pseudocode GWO algorithm.

2.2 Our proposal

The objective of the present work is to address the optimization problem of generation of addition chains of minimal length by appliying a metaheurístic not explored or scarcely explored such as GWO. In addition, is important to note that GWO is one of the metaheuristics of more recent emergence used for generation of addition chains.
For our implementation we used a programmed function on MATLAB© language, able to generate chains of addition associated with exponents that receives as input parameters [17] (see Algorithm 2) and it was optimized by GWO.

1- Assign to $u_0 = 1$ and $u_1=2$
2- Randomly select 3 or 4 and assign it to u_2.
3- Complete the sequence by invoking the function FILL(U, Rand(3,4),e) // parameters used by the function to "fill" the chain.
4- Return (U, L)

Algorithm 2. Produce a valid addition chain

Where:

e: exponent for which the addition chain is generated.
U: (u_0,u_1, u_2 …. e): complete addition chain of length L.

Rand(3,4): method that randomly takes the value 3 or 4 to generate the third component of the chain.
FILL: function fill, which is used to complete the chain, described below the structure of the same:

```
FILL(U, Rand(3,4),e)
i = k – 1
while u_i ` e do:        // While the position of the chain is distinct from "e".
 if FLIP(f) then:
    u_{i+1} = 2u_i          // apply "folded" (rule 1)
 else
    if FLIP(g) then:
       u_{i+1} = u_i + u_{i-1}   // apply (rule 2)
    else
       u_{i+1} = u_i + u_rand // apply (rule 3)
    end if
 end if
 while (u_{i+1}) > e do:    // For the next position to be greater than "e".
    j = i – 1
    u_{i+1} = u_i + u_{i – j}
    j = j – 1              // Is decremented so as not to exceed the value of "e".
 end while
end while
return (U)
```

Algorithm 3. Function: complete chain.

As in [17] individuals of the population will be generated applying the 3 rules mentioned below, always with the premise that the first two components of the chain are 1 and 2, in this same order:

1- It is allowed that an element of the chain results from adding to the previous element by itself (folded).
Thus, in a chain U, a component u_{i+1} can be equal to:
2* u_i a ($u_{i+1} = u_i + u_i$). For example, for the chain {1, 3, 6, 9, 12, 15, 21}, the third element consists of adding the second element to itself.

2- Sum of two previous numbers, so that: $u_{i+1} = u_i + u_{i-1}$. Taking as an example the same chain as in point 1, the fourth element (9) is equal to the second added to the third.

3- Sum of a previous number plus a random number of the chain: $u_{i+1} = u_i + u_{rand}$. Following the same previous example: the element (15) is formed from the sum of the fifth (12) and the second element (3).

In order to adapt the behavior of GWO to solve the problematic object of study, we will establish that our population of wolves (solutions) is a population of valid addition chains for a given exponent, i.e., numerical sequences of integers and that correspond to the rules mentioned above. On the other hand, taking into account that the purpose of optimizing chains for a given exponent is to obtain chains of minimum length, the fitness function to be used, will be the length of the chains.
In this way, the FILL function applies the above rules to certain probability values. For the experiments we have taken as reference the values reported by the approach [17] that has shown competitive results for several ranges of exponents, in order to be able to compare our proposal under similar conditions in terms of solution generation and performance of the proposed heuristic . These values are:

- Rate of determination of an element of the chain from the sum of the previous number by itself (rule 1) = 0.7
- Sum rate of previous items (rule 2) = 0.2

In the next section, we present the results of a series of statistical tests performed on GWO metaheuristics applied to the problem under study.
All results obtained using the GWO approach have been generated by applying the following parameter settings:

- SearchAgents = 10 (number of search agents)
- Max_iteration= 30

3. Results

This section describes how the performance of GWO was tested.

For all experiments was used GWO metaheuristic on MATLAB© v.7.8.0 (R2009a) under Operating System Windows 8.1 Enterprise 64 bits, in a computer: Intel Core i-3. 2.10 GHz., 4 Gb. RAM.

3.1 Experiment 1

The first experiment was to calculate the total accumulated lengths of additive chains for sets of small length exponents in order to compare the performance of GWO algorithm with other results obtained by deterministic methods of the literature and other metaheuristics of the state of the art:

Table 2. Average of results obtained by GWO on the following ranges of exponents with 30 independent executions.

Exponent	Accumulated length	Average	Median	Deviation
[1, 128]	960	7,5	4	2,159
[1, 256]	2459	9,6	9	2,986
[1, 512]	4994	9,8	10	1,9501

Table 2 summarizes the best results obtained by enumeration for the exponents of the expressed ranges, taking the best of 30 independent runs, in order to be able to compare such results with other proposals for which experiments have been carried out by using the same range of exponents. They are described below:

Table 3. Cumulative addition chains for all lengths of exponents "e" õ [1,512]:

e	[1, 512]
Optimal [13]	4924
Binary [13]	5388
Çuaternary [13]	5226

GWO Results

Accumulated lengths	4994
Average	9,77
Median	10
Deviation	1,95

Table 4. Comparison of the best results accumulated for GWO with other approaches, for the same range of exponents.

e õ [1, 512]	AIS	GA	PSO	EP	GWO
	4924	4924	----	4924	4994

Table 3 shows a comparison of the results for GWO with respect to the deterministic state of the art methods taken from [13] whose results were used to compare with the AIS approach [13], on the other hand Table 4 provides a comparison of the best accumulative addition chain results for the same range of exponents [1, 512] obtained by stochastic methods and compared in [17], except PSO that did not report results for this range of exponents, with respect to the proposed GWO.
In Table 3 it can be seen that GWO approximates the result of the "Optimal" method [13], which is the best performance for that range of exponents according to [13]. On the other hand, in Table 4 it is observed that GWO arrives at its best run to match the best results that yielded AIS, GA, PSO, EP for range [1, 512]. In comparison with the results of the deterministic methods, it can be observed that in its best execution, it is close to the best result of "Optimal", whereas in the comparative results with the other stochastic methods it approximates the optimum of the other proposals, not exceeding the by a percentage of H0,9%.

3.2 Experiment 2

The experiment raises the comparison of GWO's performance with a recent proposal [22] of the state of the art.

Table 2. Comparison of GA Annealing [22] and GWO for the same set of exponents.

Exponent	Chain	GA Annealing	GWO
e		Length	
23	1, 2, 4, 5, 10, 20, 21, 23	7	7
55	1, 2, 3, 6, 12, 24, 27, 54, 55	9	8
130	1, 2, 4, 8, 16, 32, 64, 128, 129, 130	11	9
250	1, 2, 3, 5, 10, 20, 30, 50, 100, 150, 250	13	10
768	1, 2, 3, 6, 12, 24, 48, 96, 192, 384, 768	23	10

Table 5 shows a comparison of the results of GWO with GA Annealing [22] for a number of different exponents and especially difficult to optimize. It is considered in this category to those exponents for which it is not possible to find chains of minimum length by means of deterministic methods.

It can be observed that in the 5 cases, GWO has obtained chains of length equal to or less than the approach proposed in [22]. Although the methods applied in [13], [14] and [17] have also been tested with this class of diverse exponents, they are not the same as those used in [22]. For this reason the comparison is made in Table 5 and in the next experiment is done with respect to the other methods of the state of the art.

3.3 Experiment 3

The experiment raises the comparison of the performance of GWO with other approaches of the state of the art.

Table 3. Best results obtained by AIS [13], PSO [14], EP [17] and GWO for a set of different exponents.

Exponent		Length			
E	Chain	AIS [13]	PSO [14]	EP [17]	GWO
5	1, 2, 3, 5	3	3	3	3
7	1, 2, 4, 5, 7	4	4	4	4
11	1, 2, 4, 8, 10, 11	5	5	5	5
19	1, 2 ,4, 8, 16, 18, 19	6	6	6	6
29	1, 2, 3, 6, 7, 14, 28, 29	7	7	7	7
47	1, 2, 4, 8, 9, 18, 36, 45, 47	8	8	8	8
71	1, 2, 3, 6, 7, 14, 21, 35, 70, 71	9	9	9	9
127	1,2, 3, 6, 12, 18, 30, 60, 63, 126, 127	10	10	10	10
191	1, 2, 3, 6, 12, 18, 19, 38, 57, 95, 190, 191	11	11	11	11

379	1, 2, 3, 6, 9, 18, 27, 45, 72, 117, 189, 378, 379	12	12	12	12
607	1, 2, 3, 6, 9, 15, 30, 60, 61, 121, 182, 303, 606, 607	13	13	13	13
1087	1, 2, 3, 6, 9, 18, 36, 54, 108, 216, 324, 540, 541, 1082, 1085, 1087	14	14	14	15

Table 6 summarizes the results of the third experiment, corresponding to the best results obtained by GWO for a set of diverse exponents and difficult to optimize as the experiment of Table 5 but comparing those results with those obtained by AIS [13], PSO [14] and EP [17] for the same cases.

GWO has been compared by taking some exponents of the metaheuristics mentioned in the previous paragraph with the exception of [8] since, as mentioned in [17], all others ([13], [14]) exceed their results.

Although GWO does not exceed the results obtained for the listed exponents by the use of the other approaches (see Table 6), it has not produced results of lower quality in terms of chain length, except for the e = 1083 exponent used in this lot of test cases.

Besides the majority of the cases taken for the experiment with the results of other metaheuristics, these results correspond to chains of minimum length as a function of the table of "various" exponents presented in [13]:

Table 4. Set of exponents associated with an optimum chain length.

Chain Length	Exponents
1	2
2	3, 4
3	5, 6, 8
4	7, 9, 10, 12, 16
5	11, 13, 14, 15, 17, 18, 20, 24, 32
6	19,21,22,23,25,26,27,28,30,33,34,36,40,48,64
7	29,31,35,37,38,39,41,42,43,44,45,46,49,50,51,52,54,56,60,65,66,68,72,80,96, 128
8	47,53,55,57,58,59,61,62,63,67,69,70,73,74,75,76,77,78,81,82,83,84,85,86,88,90,92 ,97,98,99,100,102,104,108,112,120,129,130,132,136,144,160,192,256

9	71,79,87,89,91,93,94,95,101,103,105,106,107,109,110,111,113,114,115,116,117, 118,119,121,122,123,124,125,126,131,133,134,135,137,138,140,145,146,147,148, 149,150,152,153,154,156,161,162,163,164,165,166,168,170,172,176,180,184,193, 194,195,196,198,200,204,208,216,224,240,257,258,260,264,272,288,320,384,512

Table 7 summarizes the most well-known and proven examples of the literature for which optimal length chains are known, indicating the length of these optimal values (first column of the table). As a function of this, it is verified that for example the exponent e = 71 which was used for the experiment of Table 6, for which the chain {1, 2, 3, 6, 7, 14, 21, 35, 70, 71} whose length is 9 coindicate with the length indicated in Table 7, beyond there being coincidence in the length also obtained by [13], [14] and [17] for the same case.

4. Conclusion

In this work we proposed the use of the GWO heuristic [21] to find chains of addition of minimum length, associated to exponents.
GWO [21] demonstrated competitive results to address the problem, which is evident from the comparison of its results, compared to other proposals of the state art. In Experiment 1 it is observed that for the range of exponents [1, 512], the accumulative minimum addition chain far exceedes two of the deterministic methods reported in [13] for that range and is only slightly exceeded by "Optimal "[13]. With respect to the proposals for the stochastic approach, in the second experiment it has been compared directly with the same cases of exponents reported by [22], in order to match and overcome this proposal, in most cases. Finally, the proposal has been compared for a group of diverse exponents and particularly difficult to optimize, in relation to the performance of other metaheuristics in the literature: AIS [13], PSO [14] and EP [17], matching the reported results except for the largest exponent.
As future work, we will analyze the parameters required by GWO [21] or even hybridizations or other techniques required, to extend the results of the metaheuristic to exponents of greater length.

Bibliography

[1] Rivest. R.L.; Shamir, A.; Adlman, L. 1978. *A method for obtaining digital signatures and public-key cryptosystems*. Cambridge.

[2] Bos, J.; Coster, M. 1990. *Addition Chain Heuristics*. Centrum voorWiskunde en Infomatica. Amsterdam, Netherland.

[3] Kaya Koc, C. 1994. *High-speed RSA implementation*. Technical report, RSA. Laboratories, Redwood City, CA. USA.

[4] Rotger, L.H.; Coma, J.R.; Tena-Ayuso, J.G. *Criptografía con Curvas Elípticas*. Universitat Oberta de Catalunya. España.

[5] Liu, Y.; Passino, K.M. 2002. *Biomimicry of Social Foraging Bacteria for Distributed Optimization: Models, Principles, and Emergent Behaviors*. Journal of Optimization Theory and Applications: Vol. 115, No. 3, pp. 603–628. Ohio State University, Columbus, Ohio.

[6] Nicosia, G.; Cutello, V.; Bentley, P.; Timmis, J. 2004. *Artificial Immune Sytems.* Third International Conference, ICARIS 2004. Springer- Verlag Berlin Heidelberg, Germany.

[7] Nedjah, N.; De Macedo-Mourelle, L. 2004. *Finding Minimal Addition Chains Using Ant Colony*. Department of Systems Engineering and Computation Faculty of Engineering, State University of Rio de Janeiro. Río de Janeiro, Brasil.

[8] Cruz-Cortés, N; Rodríguez-Henríquez, F.; Juárez-Morales, R.; Coello Coello, C. 2005. *Finding Optimal Addition Chains Using a Genetic Algorithm Approach*. México.

[9] Karaboga, D. 2005. *An Idea Based on Honey Bee Swarm for Numerical Optimization*. (Technical Report). Erciyes University, Engineering Faculty Computer Engineering Department. Kayseri, Turquía.

[10] Karaboga, D.; Basturk, B. 2007. *A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm*. Department of Computer Engineering, Erciyes University, Kayseri, Turkey.

[11] Washington, L. 2008. *Elliptic Curves: Number Theory and Cryptography*. Second Edition. Discrete Matemathics and Its applications, series editor: Kenneth H. Rosen. Taylor & Francis Group. Boca Ratón, USA.

[12] Mezura-Montes, E.; Hernández- Ocaña, B. 2008. *Bacterial Foraging for Engineering Design Problems: Preliminary Results*. Laboratorio Nacional de Informática Avanzada (LANIA A.C.) - Universidad Juárez Autónoma de Tabasco. México.

[13] Cruz-Cortés, N.; Rodríguez-Henríquez, F.; Coello Coello, C. 2008. *An Artificial Immune System Heuristic for Generating Short Addition Chains*. México.

[14] León. A.; Cruz-Cortés, N.; Moreno-Armendáriz, M.; Orantes-Jiménez, S. 2009. *Finding Minimal Addition Chains with a Particle Swarm Optimization Algorithm*. Center for Computing Research, National Polytechnic Institute. México.

[15] Yang, X.S. 2010. *Nature – Inspired Metaheuristic Algorithms*, Second Edition. University of Cambridge. Luniver Press, United Kingdom.

[16] Gómez Bello, M. 2011. *La aritmética modular y alguna de sus aplicaciones*. Universidad Nacional de Colombia. Bogotá, Colombia.

[17] Domínguez, I. 2011. *Optimización de Cadenas de Adición en Criptografía utilizando Programación Evolutiva*. Tesis de Maestría. Laboratorio Nacional de Informática Avanzada Centro de Enseñanza LANIA, Xalapa, Veracruz, México.

[18] Binitha, S.; S Siva Sathya. 2012. *A Survey of Bio inspired Optimization Algorithms*. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-2. India.

[19] Baro, M. 2013. *Swarming: La Comunicación en múltiples direcciones y múltiples etapas*. Razón y palabra. Primera Revista Electrónica en Iberoamérica Especializada en Comunicación. Centro Avanzado de Comunicación - 25 Aniversario Eulalio Ferrer. Número 83, Junio – Agosto.

[20] Tall, A.; Sanghare, A.Y. 2013. *Efficient computation of addition-subtraction chains using generalized continued Fractions*. African Institute for Mathematical Sciences. Senegal.

[21] Mirjalili,S.; Mirjalili, S.M.; Lewis, A. 2014. *Grey Wolf Optimizer*. School of Information and Communication Technology, Griffith University, Nathan Campus, Brisbane QLD 4111, Australia. Department of Electrical Engineering, Faculty of Electrical and Computer Engineering, ShahidBeheshti University, G.C. 1983963113, Tehran, Iran.

[22] Pogan•i•, M. 2014. *Evolving Minimal Addition Chain Exponentiation*. University of Zagreb - Faculty of Electrical Engineering and Computing. Zagreb, Croacia.

[23] Picek, S.; Coello Coello, A.; Jakobovic, D.; Metens, N. 2016 - *Evolutionary Algorithms for Finding Short Addition Chains: Going the Distance*. Volume 9595 of the series Lecture Notes in Computer Science pp 121-137.

[24] Mech., D. 1999. *Alpha Status, Dominance, and Division of Labor in Wolf Packs*. Canadian Journal of Zoology. Jamestown, Dakota del Norte.

XVII

Distributed and Parallel Processing Workshop

Optimization and Parallel Computing to Improve River Flow Forecasting

ADRIANA GAUDIANI[1], EMILIO LUQUE[2], PABLO GARCÍA[3], MARCELO NAIOUF[4], ARMANDO DE GIUSTI[4,1]

[1]Instituto de Ciencias, Univ. Nac. de General Sarmiento, Argentina
agaudi@ungs.edu.ar
[2]Depto. de Arquitectura de Computadores y Sistemas Operativos, Universitat Autónoma de Barcelona, España, emilio.luque@uab.es
[3]Instituto Nacional del Agua, Argentina, pabloegarcia@gmail.com
[4]Instituto de Investigación en Informática LIDI (III-LIDI), Universidad Nacional de La Plata, Bs. As., Argentina, {mnaiouf, degiusti}@lidi.info.unlp.edu.ar

Abstract. The uncertainty in the values of the input parameters of a model is a fundamental problem when performing model calibration. Finding optimized parameters is a problem of high computational cost when we deal with a large number of these parameters. In this case, we are dealing with a combinatorial problem with a huge search space. In this article, we present a method of optimization through simulation to treat this problem and improve the prediction of a river channel simulator, taking advantage of the valuable contributions of the parallel and distributed computation to this area of knowledge. The solution we propose uses optimization techniques in the analysis of simulations, reducing the search space and identifying regions that allow finding an optimal configuration in reduced times. This methodology provides an acceleration of the optimal solution search, boosted with the techniques of high-performance computing, and a reduction of the prediction errors of the river model.

Keywords: computational simulation; optimization via simulation; parallel computation; automatic calibration;

1. Introduction

Physical systems are dynamic systems that modify their variables value constantly. For example, simulation and prediction of floods caused by the overflowing of water that carries the channel of rivers, being a critical issue for the world's population [1]. Only the successive calibrations and validations of the model allow handling input errors, such as parameters uncertainty, in order to improve the model behavior. Such tasks are usually very expensive in terms of human resources, time and money.

[1] Principal Investigator, CONICET

The objective of this work is to optimize the forecasting of a hydrological simulator, using optimization via simulation techniques. We present a calibration and tuning methodology for a riverbed model based on computational methods, which are addressed by using high performance computing (HPC).

This work take advantage of the ongoing research developed by the research group *High Performance Computing for Efficient Applications & Simulation,* Universitat Autonoma of Barcelona[2] [2], with the participation of the engineers of the *Instituto Nacional del Agua,* at Argentina (INA), who created the hydraulic model EZEIZA.

We optimize the model prediction looking for the optimal parameter vectors for the simulation by minimizing an objective function. These parameter set is sought "automatically" in order to minimize the difference between the simulated and observed data. We chose the parameter settings by selecting the best from a set of candidate parameter settings, but choosing a good configuration remains a challenging problem. Due to the high search space dimension, an exhaustive search is computationally intractable.

To achieve the objective previously stated, we present an optimization via simulation (OvS) approach using a two-steps scheme, which combines optimization heuristics and simulation analysis. The first phase uses an iterative Monte Carlo heuristic (MMC) plus a K-Means clustering method, based on a neighborhood structure of the problem and finding a reduced search space. The second phase seeks the best solution, by means of a reduced exhaustive search on a reduced feasible region. The proposed methodology involves a large number of simulation executions and require a considerable execution time. We used HPC techniques, in order to obtain results in a reasonable time. The optimized prediction scheme provided a reduction of 13-19% in prediction errors as compared to the classical simulation.

In this paper, we analyze the improvements obtained by using the proposed OvS scheme in terms of prediction quality and the significant simulation speed-up achieved. The speed-up calculation consists of two complementary parts: the speed-up obtained by the OvS method and that using HPC technologies. The total speed-up reached varies from 10-20X.

The paper is organized as follow. The identification of the problem that motivate the research is presented in section 2. The optimized prediction method is explained in section 3. Section 4 includes the experimentation performed and the speed-up obtained. The main conclusion are reported in section 5.

[2] http://grupsderecerca.uab.cat/hpc4eas/

2. Identifying the problem

Riverbed simulation and flood inundation models play a central role in real-time flood forecasting, when the river bursts its banks. The behavior of these natural events are described by mathematical models that work with discharge and water level as upstream, downstream and/or internal boundaries. The model accuracy depend, in great part, on to have updated system data. The model parameter uncertainty is one of the most important error source in the river model simulation [3]. In this paper, we deal with this problem by proposing a tuning methodology to find an optimized parameter set of input. We propose an automatic model calibration based on the OvS scheme, as further discussed in next section.

2.1 Input data uncertainty

Hydrological models used in practical river engineering, such us flood inundation models, produce both hydrographs and inundation information as model output. These models input data are required to compute the output: the parameter values, the morphological data, the boundary and initial conditions. Flood inundation models produce both hydrographs and inundation data as model output.
The model parameters are evaluated from field data, for example, hydro-meteorological observations, topography maps, soil types, etc. However, complete field data are rarely available to validate and determine accurate parameters values. On the other hand, these parameters values change over time due to the river system evolution.

2.2 EZEIZA: A computational model of Paraná River

The main application of the hydrodynamic model EZEIZA has been for the Hydrological Warning System of the Rio de la Plata Basin, in South America, a service provided by INA. Specifically, the Paraná River has been modeled from Yacyretá dam down to Rosario city, in Argentina, including part of the Paraguay River, for a total of about 1,500 km. Interested readers can find more information in [4].
The model flow net, that define the topology of the modelled river system, is shown in Figure 1. The most sensitive parameters in flood routing models are the rugosity value, or Manning coefficient, and the levees height, and The Manning value is calculated independently in both main channels and flood plains [5].

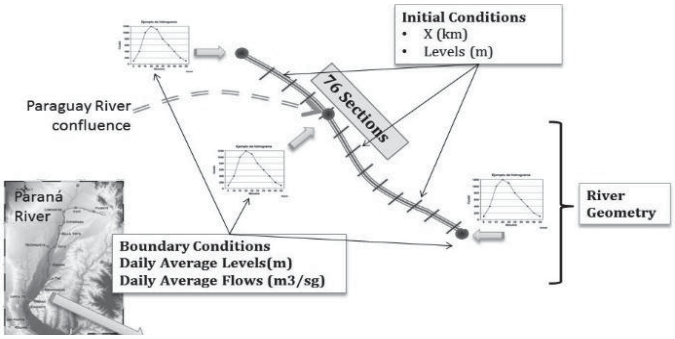


Figure 1: Flow net topology

3. Optimization via Simulation to improve river flow model

To improve EZEIZA prediction quality, this optimization method launch the simulation using input scenarios based on different sets of parameters, in an iterative way. The main goal of OvS applied to solve optimization problems is to obtain the best possible value of decision variables associated to the computational model of the system under study that results in the best performance measures. At this point, we want to highlight that the computational model, which represent the river system, can be modified without affecting the procedure used to solve the optimization problem [6] [7].

3.1 The optimization problem

The hydrological model calibration problem can be stated as a multi-objective optimization problem.
Formal optimization is associated with the specification of a mathematical objective function (called *f*) and a collection of parameters that should be adjusted to optimize the objective function. Mathematically an optimization problem can be stated as:

$$max / \min f(x)$$

$$subject\ to\ x \in S \qquad (1)$$

Where *x* is the variable; *f* is a function $(f : S \rightarrow \mathbb{R})$; S is the constraint set, and $\exists\, x_0 \in S$ such that $f(x_0) \leq f(x)\ \forall x \in S$, for minimization, and $f(x_0) \geq f(x)\ \ \forall x \in S$, for maximization.
In this context, the optimization problem is a search problem and it is expressed as follow: We find the parameters vector $\overrightarrow{x^*} = [x_1^*, x_2^*, \ldots, x_N^*]$, N-

dimensional, which optimize the objective function $f(x), \forall x \in S$ where $f: \mathbb{R}^p \rightarrow \mathbb{R}$. The search space of the problem is represented by the domain $S \subseteq \mathbb{R}^p$. The domain is quite common specified by a set of conditions or constraints that its elements have to satisfy. These elements are known as *candidate solutions*, which define a *feasible region*.

The optimization goal consists in seeking the best solution. We mean to find the parameters configuration that minimizes the difference between the simulated data and the observed data and the simulation runs with a scenario determined by the best of the candidate solutions. The OvS scheme remains appropriate when objective function cannot be calculated analytically.

EZEIZA is considered a black box for the OvS method. This means that each time we run a simulation, a full scenario is entered into the program and, when it has finished, we receive the output data. In order to do optimization, a *quality index* have to be set. This index is set for each scenario and it provides a measure of the simulation quality. We mean the scenario that allows us to reach the shortest distance between observed data and simulated results.

The calibration parameters values are set for each of the 76 sections that were used to discretize the Parana River domain. Three parameters were selected to carry on the OvS scheme, they are floodplain Manning value (*Mp*), riverbed Manning value (*Mc*) and levees height (*L*). The range value for each parameter were discretized, therefore the parameters real values are treated as discrete valued. The optimization problem has become in a discrete combinatorial optimization problem.

When the number of considered sections grows, then we have to lead with a combinatorial explosion of possible solutions, or different vectors of decision variables (scenarios). If we considered only Mc and Mp decision variables, Table 1 shows how the number of scenarios could quickly increase in function of the sections considered. We calculated the number of scenarios as $(Mc * Mp)^{\#Sections}$

Table 1: Combinatorial explosion of scenarios that were regarded for the simulation.

Cardinality		Number of river sections					
Mc	**Mp**	**2**	**4**	**6**	**8**	**10**	**76**
2	4	64	4096	262144	1.68E+07	1.07E+09	4.31E+68
3	5	225	50625	1.14E+07	2.56E+09	5.77E+11	2.42E+89
4	5	400	160000	6.40E+07	2.56E+10	1.02E+13	7.56E+98
5	5	625	390625	2.44E+08	1.53E+11	9.54E+13	1.75E+106

3.2 The metrics used to assess the OvS proposal

The index quality (*Ic*) used to evaluate the OvS methodology, is determined by the output or simulated data (Q_{sim}) and the observed data (Q_{obs}). Ic is calculated for each of the 15 monitoring stations located in the considered area of the channel. The time step used in the simulation is in days (n). We set the Ic in function of each local index measured for each station, I_{ε_j}, $j: 1..15$, as follow:

$$\mathrm{Ic} = \frac{\sum_{j=1}^{15} I_{\varepsilon_j}}{15}$$

$$I_{\varepsilon_j} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (Q_{sim}^{i} - Q_{obs}^{i})^2}$$

We focus our work in improving the performance of the current simulation results achieved by INA. In other words, we refer to the scenario used by INA when running EZEIZA to forecast the river flow dynamics. It is necessary to define a new metric, I_M, to calculate the number of the improved monitoring stations with respect to the INA scenario. The mathematical expression of I_M index is the next equation:

$$I_M = (\#\ \text{Stations} : (Ic_j^{Sim} < Ic_j^{INA})_{j=1}^{15}$$

3.3 The scheme for search space reduction

As we explained before, the OvS method for tuning the model EZEIZA, consists of two phases, as we show in Figure 2. The first phase combines a MMC heuristic and a K-Means clustering method. This phase identifies promising regions for optimization based on a neighborhood structure of the problem. The MMC algorithm, which is based on Metropolis algorithm [8], detects in an iterative way, those scenarios that have a better average mean value and standard deviation than the previous ones. K-Means identifies the promising region from the scenarios selected by MMC. The second phase performs an exhaustive search in such promising region, in order to find the optimal solution that satisfies the constraints set by the optimization method. This methodology is further explained in [9].

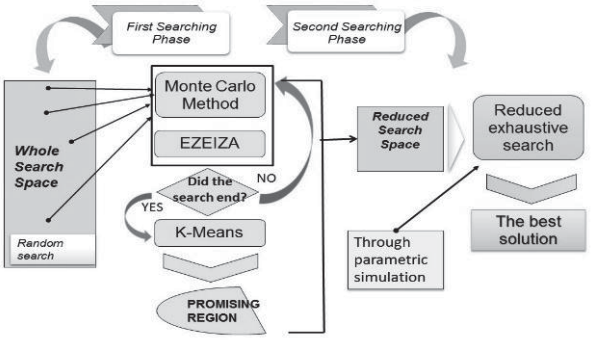


Figure 2: OvS methodology in two phases

4. Experimentation

In this section, we explain one of the experiences used to validate the proposal methodology. This simulation was run for a simulation period of 360 days, using data of year 1999. This year was characterized by periods of drought and flooding.
The parameters values ranges are:

- Mc: [0.015 .. 0.045], steps of 0.015 - #Mc = 3.
- Mp: [0.2 .. 0.4], steps of 0.1 - #Mp = 4.
- L: [0 .. 15], steps of 5 - #L = 4

We implemented de simulation combining the parameters values in three sections: 72, 74 and 76. The search space dimension is $(3 * 4 * 4)^3$, that is 110,592 configurations.

This optimization problem is expressed mathematically in equation:

$$\begin{aligned} &Minimize\ (Ic) && f(Q_{Sim}, Q_{obs}) \\ &subject\ to && 0.2 \leq M_p \leq 0.4 \\ & && 0.015 \leq M_c \leq 0.045 \\ & && 0 \leq L \leq 15 \end{aligned}$$

The I_M index expression is as follow:

$$\begin{aligned} &Minimize\ (I_M) && f(Q_{Sim}, Q_{obs}, Q_{INA}) \\ &subject\ to && \tfrac{1}{k}.\sum(Ic_s^{Sim} - Ic_s^{INA}) < tol \end{aligned}$$

The variable *tol* is a tolerance value set in 0.1, and the variable k is the number of stations, *s,* that fulfill the next constrain:

$$(Ic_s^{Sim} - Ic_s^{INA}) < \text{tol}$$

The decision variables maintain the constraints imposed by Equation (6).
In the first step, MMC process 52 groups, of 100 points each one, returning the best group found. This process required 5200 executions of EZEIZA. The group selected was characterized by the minimum index values, both for Ic and for I_M. Each point is a vector, whose elements are the decision variables of a given scenario. Afterward, K-Means identified the three clusters shown in Figure 3. The best cluster is the one which contains 8 points, that is to say 8 vectors of decision variables. These points are used to delimit the region where the minimum is located. The reduced search space found is limited by the Mc, Mp and L values, determined by those 8 points. This new reduced space contains 192 points, and one of them is the best solution. Finally, (the second step) an exhaustive search is used to find the optimal solution in the new reduced feasible space.
In Figure 4 we compare, for each one of the monitoring stations, the combined quality index Ic&I_M, reached when using the best scenario provided by the OvS method and when using the INA scenario. It is important to emphasize the improvements achieved for the three stations located at the lower reaches of Parana River.

We present the optimum solution, provided by the OvS scheme, in

Table 2. This solution was obtained when the objectives functions (5) and (7) were resolved, as a multi-objective optimization problem. This solution is the scenario determined by the minimum I_C that maximize I_M at the same time. The optimum I_C is 0.801 and I_M is 3 stations, as we can see in

Table 2.

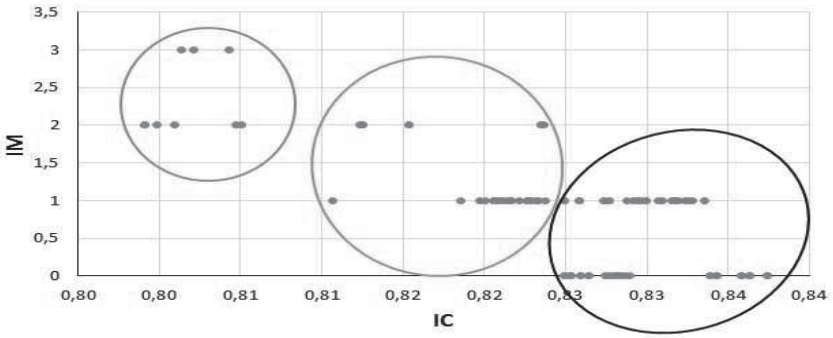


Figure 3: K-means identified three clusters for average I_C&I_M

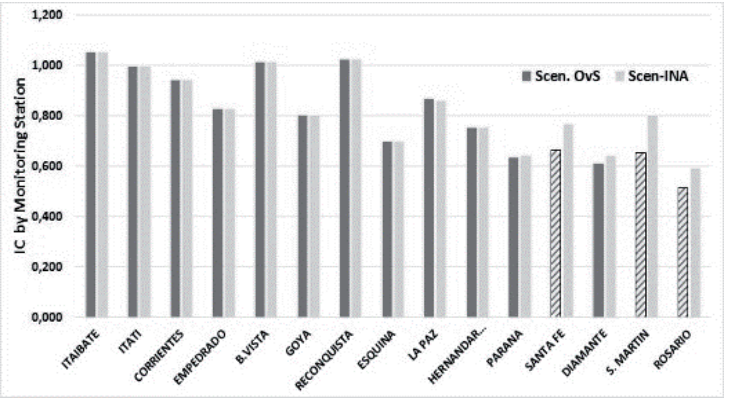


Figure 4: I_C index compared with INA scenario for each station when I_C&I_M are the objective. The three best resulting stations are highlighted.

Table 2: Solution reached when I_C&I_M improvement.

Optimal solution			
Section	**Mp**	**Mc**	**L**
1	0.02	0.2	5
2	0.02	0.2	5
3	0.03	0.1	0

Table 3: The best OvS method are the objective functions.

	Scenario		
Station	**OvS**	**INA**	**Gain**
Santa Fe	0.662	0.764	13%
S. Martín	0.651	0.799	19%
Rosario	0.512	0.59	13%

4.1 Parallel implementation of the optimization proposal

All the riverbed simulations were carried out on the high performance computer architectures in the Research Institute III-LIDI, Universidad Nacional de La Plata. The experiences were implemented on a multicore HP Blade, using 16 nodes of 2 processors quad-core Intel Xeon-2GHz and 10GB of RAM, with GNU Linux/Debian for 64bits.

5. Performance achieved with the proposal methodology

The tuning methodology, based on the OvS scheme proposed in this paper, allowed us to enhance the simulation performance and to reduce the total execution time. The improvement achieved involves two phases, during the work development, and both contribute to the total gain. In the first phase,

our proposal achieves a significant search space reduction measured in terms of number of simulations to be performed. In a second phase, our efforts were oriented to take advantage of the computing power provided by HPC systems. The speed-up reached at this step adds to the first one.

Table 4: OvS methodology improvement

Exhaustive search solution **(Simulations)**	Resulting simulations by Monte Carlo (Reduced)	Resulting simulations by K-Means (Reduced)	Search over the reduced search space **(Reduced)**	**Sum of simulated scenarios**
110.592	**5200**	8	**192**	**5200+192**
⇦ *MMC Speed-Up: 21.26* ⇨				
⇦ *Global Speed-Up:* ***20.51*** ⇨				

The total gain, in both phases, is represented in Table 4. The Table 5 compares the execution times resulting of the experience explained in section 4. These execution times are presented using an exhaustive search, over the whole search space, as baseline case.

Table 5: Execution time and speed-up reached at each OvS phase.

Number of simulations	**Sequential**		**16 Processors**		**Speed-up**	**1000 Proc**
	min.	**days**	**min.**	**days**		**min.**
110592	243302	169	15896	11	15.306	~ 240
5200	11440	8	756	0,5	15.132	~11
8	(solution selected by K-Means - *They aren't run*)					
192	422	0.3	27	0	15.63	~0.4

6. Conclusions

In this paper, we expose a methodology to tune and calibrate automatically a river flow model. Our proposal contributes to improve the simulator prediction quality through an OvS methodology, further enhanced by computing methods and HPC technologies. The improvement percentage of our implemented scheme, regarding the prediction obtained with the scenario used by INA, ranges from 13% to 19%. In other words, we are lowering about 40-60 cm the errors of water depth and it is a very important achieve when a flood event occurs.

We have presented a computational method highly scalable, as we can see in Table 4; however, our purpose is to improve the simulator prediction quality by applying computing knowledge and skills, and using the least possible amount of computational resources. As we presented before, this aim was achieved because we have got an improvement that results in an improvement from two sources. We are working now on considering longer

periods of simulation to tune the system for successive periods of flood and downspout. We are interested in setting validity ranges for the decision variables, in order to perform a more accurate prediction using the river flow model, for example, decreasing the percentage of error in more monitoring stations at once.

Acknowledgement. The MINECO Spain, under contract TIN2014-53172-P, has supported this research and it was partially supported by the research program of Informatics Research Institute III-LIDI, Universidad Nacional de La Plata. We are very grateful for the collaboration of INA Hydraulic Laboratory.

References

[1] S. Jonkman and J. Vrijling, "Loss of life due to floods," *Journal of Flood Risk Management,* vol. 1, no. 1, pp. 43-56, 2008.

[2] E. Cabrera, E. Luque, M. a. E. F. Taboada and M. Iglesias, "Optimization of emergency departments by agent-based modeling and simulation," in *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on*, 2012.

[3] J. J. Warmink, J. A. Janssen, M. Booij and M. Krol, "Identification and classification of uncertainties in the application of environmental models," *Environmental Modelling & Software,* vol. 25, no. 12, pp. 1518-1527, 2010.

[4] A. Menéndez, "Three decades of development and application of numerical simulation tools at INA Hydraulic lab," *Mecánica Comutacional,* vol. XXI, pp. 2247-2266, 2002.

[5] G. Arcement and V. Schneider, "Guide for selecting Manning's roughness coefficients for natural channels and flood plains," *United States Geological Survey Water-Supply Paper 2339,* 1989.

[6] L.-F. Wang and L.-Y. Shi, "Simulation Optimization: A Review on Theory and Applications," *Acta Automatica Sinica,* vol. 39, no. 11, pp. 1957-1968, 2013.

[7] J. F. Santucci and L. Capocchi, "Optimization via Simulation of Catchment Basin Management Using a Discrete-event Approach," *Simulation,* vol. 91, pp. 43-58, 2015.

[8] D. P. Kroese, T. Taimre and Z. I. Botev, Handbook of Monte Carlo Methods, vol. 706, John Wiley \& Sons, 2013.

[9] A. Gaudiani, E. Luque, P. García, M. Re, M. Naiouf and A. De Giusti, "How a Computational Method Can Help to Improve the Quality of River Flood Prediction by Simulation," in *Advances and New Trends in Environmental and Energy Informatics*, Switzerland, Springer International Publishing, 2016, pp. 337-351.

A Parallel Proposal for SEIR Model Using Cellular Automata

FACUNDO CASARES, CRISTIAN TISSERA, FABIANA PICCOLI [1]

[1] LIDIC, Universidad Nacional de San Luis, Ejercito de los Andes 950 – 5700
San Luis – Argentina.
{ptissera, mpiccoli}@unsl.edu.ar

Abstract. Cellular Automata have been used with success in simulations of simples and complex systems belonging to different scientific areas, such as chemistry, biochemistry, economy, physics, etc.. In this work, we propose to use it in order to specify and implement a simulation model that allows to investigate behavioral dynamics for seasonal flu. This work presents a general solution where parallel programming techniques of shared memory are applied. Finally some experimental results about performance and flu performance are showed.

Keywords: Parallel Cellular Automata, Simulation, Seasonal Flu.

1. Introduction

Throughout times, the diffusion and spread of disease were a major concern of the human being. There were cases where a disease caused the disappearance of an entire population and important demographic changes, some of them were the plague (Europe, XIV century), yellow fever (Buenos Aires, XIX century) and cholera (Asia, XIX century). Today the situation continues, there are diseases monitored in certain regions such as malaria, and other are new as Influenza A or persistent as AIDS. For all this, it is important and priority the study and control of these diseases and their mode of transmission or contagion. One way to address the problem is analyzing how the disease is distributed in a specific population. The study and analyses of complex real systems like this can be done through some simulation models.

When it comes to simulate discrete dynamical systems, cellular automata (CA) has been successfully used in simulation of diffusion process and, it is a valid alternative when we work with discrete dynamic systems which have complex behaviors from a simple set of rules. These rules allow to specify the new state of a component based on its state and its neighborhood. In this way, it is possible to model complex dynamic systems from the specification of the local dynamics of each component. Besides, the state of each of them can be calculated simultaneously, i.e. in parallel. The performance of CA can display graphically the system evolution, allowing an easy comprehension of

the studied dynamics. Moreover, CA has demonstrated to be very useful simulation tools at the time of constructing artificial scenes, mainly in domains not suitable for other approaches. In other word, CA can be used to simplify complicated relationships by means local interactions. For example, transmission of rumor, diseases or computer viruses can be reduced to local interactions among individuals/computers, as the case. In this work, we focus in spread of diseases, particularly the influenza or flu [1, 3].
Different techniques have been developed to study the spread of diseases. This technique divide population into different types considering the characteristics of the disease: susceptible, exposed (with or without symptoms), infected, infectious, recovered, vaccinated, isolated, diagnosed, etc.. According to the disease and its infectious agent, an individual can be in some of above states. From these states and the dynamics governing their compartments, different mathematical models arise, some of them are: SI (Susceptible-Infected), SIS (Susceptible-Infected-Susceptible), SIR (Susceptible-Infected-Recovered), SEIR (Susceptible- Exposed-Infected-Recovered), and other variants (SIRS, SEIRS, SEIQR, etc.). Particularly in this work we focus on the SEIR model. It is the most suitable for epidemics whose infectious agent is a virus. Generally when infectious agents are viruses, the individuals recovered or cured can achieve a state of resistance for same virus [4, 5].
Taking into account the previous aspects and the parallel nature of CA, we present a parallel solution to simulate the flu propagation using CA and SIRS models. This solution will allow to study, in short time, the spread of influenza or any virus mutation in different environments, considering type of population, its distribution and other characteristics; and to take decisions in consequence, such as vaccination campaigns, isolation, quarantine, etc.
The paper is organized as follows: the next sections describe all the previous concepts. Sections 3, 4 and 5 sketch the characteristics of seasonal flu, our parallel proposal, and its empirical performance. Finally, the conclusions and future works are exposed.

2. Previous Concepts

In this section, we explain the main concepts to develop this work.

2.1 Cellular Automata

In their research about the machines with auto-replication capabilities, John Von Neumann and Stanislaw Ulam were the first in formulate the Cellular Automata(CA). But was in 1970 when these systems received special attention, Jhon H. Conway proposed the game of life, the most known CA. Since that date, the CAs have grown in popularity within of the scientific community and actually they are considered to solve problems with different nature [6].

A CA is a mathematical system with discrete values in space, time and state. It has different characteristics; some of them are auto-replication, universal computation capabilities and auto-organization e effects. This last property plays a very important role at the time of explaining certain kind of behaviors observed in physical and biological phenomena [7, 8], in consequence, the CA have been used, for example, to simulate different phenomena as chemical reactions, diffusion processes, hydrodynamic, mechanic, filtration, chaos theory and others.

A CA is a discrete dynamic system with capacity to develop complex behaviors from a simple set of rules. It represents a grid of locally connected finite automata, each of them produces an output from several inputs, and next state is a result to apply a transition function. The upcoming state of a CA cell depends of its own current state and the states of its neighboring cells [9].

Intuitively, we consider a CA as a system composed by an array of cells A. Each cell c_i in A represents a finite automaton with a set of states Q, an input alphabet $\tilde{A}$ and a transition function δ: $Q \times \tilde{A}' \quad Q$. The input alphabet is given by all the possible combination of the cell states of the adjacent (neighboring) cells. If we denote N_{ci} to the set of cells that we consider as neighbor of cell c_i, and $|N^{ci}| = n$ is the number of adjacent cells, then the input alphabet is $\tilde{A}\ a\ Q^n$. Usually, a cell c_i and its adjacent are considered and represented as a unique set $N = \{ c_i \} \cup N^{ci}$. N is referenced as the *neighborhood*.

By all above exposed, a CA is defined as a 4-tuple $M =< A; Q; \delta; N >$ where:

- A is a D-dimensional array, and each component (cell of the array) has associated a finite automaton.
- Q is a finite set of states (of the automaton) of a cell.
- N is the specification of which cells are included in a neighborhood, $N = \{ c_i \} \cup N^{ci}$ such that N^{ci} are adjacent cells to c_i.
- Let $\tilde{A}\ a\ Q^n$ where $n = |N^{ci}|$ is the number of adjacent cells to c_i. The *transition function* of states, δ: $Q \times \tilde{A}' \quad Q$ is a mapping such that if $q_i \in Q$ is the state of the cell c_i in time t and $q_{i+1}, q_{i+2}, \ldots, q_{i+n} \in \tilde{A}$ are the states of adjacent cells to c_i and

  $$\delta (q_i , q_{i+1}, q_{i+2}, \ldots, q_{i+n}) = q'_i$$

  where $q_i , q_{i+1}, q_{i+2}, \ldots, q_{i+n}$ are the states of the central cell and its neighborhood at the time t and q'_i is the state of the central cell at the time $t + 1$.

In some cases, it is possible to specify probabilistic transition rules, where an arbitrary probability p can be associated to a transition rule. The semantic of this kind of rules establishes that, always that a cell matches the configuration of the specified neighborhood in a probabilistic rule, the cell will have at time $t + 1$ the new state specified in such rule with probability p.

2.2 SIR Model and Derivatives

In the modeling of diseases, several considerations must be taken into account, among them are important to consider: the infectious agents (they are responsible for transmitting the disease and condition the states through which passes an individual a affected) and transmission modes (they can be person-to-person, by the environment, by some vectors such as insects or agents, or among animals of the same or different species).
Because all the numerous factors involved in a disease, it is impossible to study them of same way. A start point is to classify states in that an diseased individual can be. A set of possible states is:

- *S:* Healthy individuals and Susceptible to be infected.
- *E:* Exposed individual to disease, infected but not infect others (i.e., the disease is latent).
- *I:* Infected individuals who infect others.
- *R:* Resistant individuals to diseases (normally, it happens after that a person recovers from illness or vaccinates).

In a same time, an individual can be in a single stage of the disease, therefore for a population of N persons, if we consider the above set of state, the following equation must be satisfied:

$$S + E + I + R = N$$

Kermack and McKendrick in 1927a formulated a simple model, SIR model, which consists of three stages: Susceptible, Infected and Recovered. The SIR model is easily written using ordinary differential equations (ODEs), this implies a deterministic continuous model. It assumes encounters among infected (I)and susceptible(S) individuals at a rate proportional to their respective numbers in the population.
The analytical techniques are good to address problems in a basic way. But, in the case of disease epidemic study, the system is complex, and, in consequence, more realistic solutions with high level of detail are necessary.

2.3 Multi-threaded Parallel Programing

Multi-threading and parallel are different concepts, in this work we apply them together to obtain good performance in a CA solution.
The multi-thread programming implies a single process and this process generates many threads. All of them share the same space memory, and they can be able to execute independently and at the same time: in parallel. The traditional multi-threading was used to do time-slicing or take advantage of the CPU idle time, i.e. while one of the threads waits, another thread could execute.
By its side, parallel programming allows explicitly breaks the task down into smallest units, where each unit can be executed in parallel on a single CPU core. When it is possible divide the task and its sub-tasks share the same

memory space and run in parallel, the problem can be solved applying Multi-threaded Parallel Programming.
To solve a problem by applying Multi-threaded Parallel Programming, we can use OpenMP(Open Multi-Processing)[10]. It is an API that supports multi-platform shared memory multiprocessing programming, it achieves parallelism via multi-threading and shared-memory.
In this work, we present a simulation system for u transmission using CA and SEIR as models and multi-threaded parallel programming techniques.

3. Previous Concepts

Influenza is a viral infectious disease that affects, primarily, the respiratory tract of humans. Usually it accompanied by other symptoms such as sore throat, weakness, dry cough, fever, and muscle aches, of stomach and head . In some cases, it may be complicated and derive in pneumonia becoming fatal. This can occur in certain age groups, such as young children and elderlies. There are three types of seasonal influenza: A, B and C. The influenza virus A and B are the most common, they are classified into subtypes according to the combination of two proteins in virus surface (H and N)[11, 12].
Virus transmission is done person-to-person, mainly through particles ejected when a sick person coughs, sneezes or talks. Also, it can be transmitted by means blood or contact with surfaces or objects contaminated. Besides, u virus is resistant in a dry and cold environment; this property allows its rapid spread mainly in autumn and winter, seasons when it becomes seasonal epidemic. The virus can keep its infections level by about one week at body temperature, however, there are patients that require 15 days of recovery. Most people recover without medical treatment. Antibiotics are only useful if there is a bacterial infection.
An infected person with the u virus goes through an incubation period (approximately from two to four days). The contagious period begins one day before that person has symptoms (this is a serious problem, a person could be spreading the influenza without knowing who is sick). After a week, the transmission power is reduced, even it disappears. The figure 1 summarizes how the disease evolves in a person from he/she is susceptible until his/her recovering or, in the worst case, death.

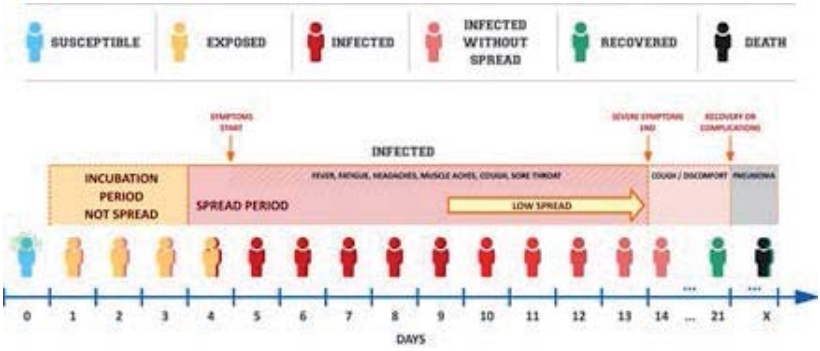


Fig. 1. Influenza progression in people.

The most effective way to prevent the u and its serious consequences is vaccination. In healthy adults, it can provide reasonable protection, while in elderlies can reduce its severity, the incidence of complications and deaths. As C influenza cases are much less frequent, generally the vaccines try influenza A and B. They are usually trivalent, they contain purified and inactivated proteins of the three strains most common in the following epidemic: two subtypes of influenza A and one B. The vaccination effectiveness depends on the match between the vaccine virus and surrounding virus. Moreover, a vaccine made one year may not be effective to the next by two reasons: the virus change and mutate rapidly, and the strains have variable dominance [12].

4. Parallel Simulation of Seasonal Flu Epidemic

The epidemiology studies the factors of potentially harmful infectious agents that a affect a particular population, and tries to explain and predict how the disease evolves in the time. As explained earlier, the simulation models that use CA concepts are ideal to represent this type of real systems. Through setting the main features of the problem, CA recreates a virtual world that comes alive and gives an approximation of what would happen in the real world.

In this section, we describe a model based on epidemiological model with SEIR approach and CA. By means this model, it is possible to analyze the effects of seasonal influenza in a population with a certain territorial distribution.

4.1 SEIR-CA Model

The SEIR CA model has a cellular space defined by a finite two-dimensional lattice. Each cell of the automata is a place busy by only one person. In this work, the cellular space (2-D array) represents a social space in which the

individuals can interact, this means, two adjacent cells occupied represent two people in touch. To perform the simulation, it is necessary to establish the following considerations:

- *Neighborhood:* For these systems, we consider the known Moore neighbor-hood: eight cells surrounding the central cell de ne the neighborhood. With this, every individual interacts with at most 8 people by once.
- *Cell State:* a cell is in one state of $Q = \{B; F; S; E; I; W; R; D\}$ where:
  - *B:* Automata limits (when borders are absorbent).
  - *F:* Free Cell (There is not any person and it can be selected to occupy). S: The person is susceptible to contract u.
  - *E:* When the person is in incubation period but not spread. I: The person is infected and, he/she spreads the u
  - *W:* The person is infected but does not spread (Infected without spread). *R:* When the person is recovered.
  - *D:* Dead person. Generally, the deaths can be by any complication. The transition through each of the states is shown in Figure 1.
- *Initial Configuration:* Before simulation begins, it is necessary to set relevant information such as size of population surface (CA Size), how the population is distributed in the surface, which is the infection percentage of population and their ages.
- *Virtual Clock:* Time is discrete, at the beginning of simulation, an interval of time is set. During this interval, a person can move to any of the neigh-boring cells and relate with other people. Generally, this movement follows some probabilistic pattern.
- *Model Evolution Rules: There are three kinds of rules which are: the rules* related to CA, with spreading of diseases, and those associated with persons movement inside CA. Each one of rules are:
  1. *CA Rules:* If a cell is in state B (CA with absorbent borders), its state does not change at no time of the simulation.
  2. *Diffusion and Spread Rules:* A cell occupied by a person in time *t*, also will be occupied in time *t+1*. It can change its state according to:
     - If at time t, the central cell is susceptible *(S)* and, some of its adjacent cells are infected (I), the central cell will be incubating the influenza *(E)* at time *t+1*, but will not spread. The probability of state change is proportional to the number of adjacent cells *I (S ' E).*
     - If the cell is incubating flu *(E)* and, the time *t* is the end of the asymptomatic period, at time t + 1 the cell will be in infected state *I(E ' I).*
     - After 8 days, a cell infected *(I)* in time t will pass to state infected without spread *(W)* at time *t + 1 (I ' W).*

- In time t, a cell infected *(I or W)* could become recovered *(R)* or dead *(D)* state *(time t+1)*. The selection between two stages depends of a probability function *(I ' {R, D} | W ' {R, D})*.

3. *Rules of Person Motion:* A person can move to a neighbor cell if it is free (state *F*). All free neighbor cells have the same probability to be occupied. An important aspect to note is that the matrix does not necessarily rep-resent a spatial universe. The model represents the people interrelation, a movement in the surface is an abstraction, it can mean that a person moves to speak with other (for example an officemate) or, he/she goes to a business and relates with a vendor. When a person comes in a neighborhood of another, this means an interaction between two people. As we assume an uniform distribution of probabilities (a person can move equally to any free adjacent cells), this model simplifies the problem. The movement is not physical, it models the interaction between a person with its neighborhood.

Once defined all CA characteristics, in next section we explain the main issues of our proposal: a parallel SEIR-CA using shared memory.

4.2 Parallel Solution

To solve the SEIR-Flu in parallel, we consider the shared memory paradigm, in consequence it is necessary to consider synchronization mechanisms to access to memory or, other programming techniques. For the implementation, some characteristics are:

- OpenMP is selected as application programming interface (API).
- In each timestep, two CA are necessary, one represents the state in time *t* (input CA), and the other is the output (CA in time *t + 1*).
- Each thread takes a particular cell and works over it. When it finishes, takes the next cell and starts again.

The solution is structured in three stages, each stage does:

- *First Stage:* In this stage, we calculate for each cell: the next state and a list of movement intentions. For that we apply the CA rules. When this stage finishes, the new CA for time *t + 1* is obtained and each free cell has an intention list to be occupied by one of its neighbor.
- *Second Stage:* Taking into account the intentions list of each free cell, one of candidates is selected in random form. To improve the solution and save memory, we use a bitmap technique for intention lists. Each list is represented by 9-bits to the Moore neighborhoods plus the central cell (fifth bit). If some position is set to 1, this means that central cell wants to move there.
- *Third Stage:* After the second stage, the data structures of CA are inconsistent state. This stage carries to a safe state, all cells and their positions are adjusted. In time t + 1, the input and output CA exchange their roles.

Every stage is made in parallel, independent and sequential sections without using synchronization mechanisms.

Other characteristic is which the breakpoint of simulation is. It can end by two reasons, whichever first occurs, they are:

- *Maximum simulation time:* A time limit for simulation is established, if this is reached, the simulation ends.
- *Some stable or starvation state:* The simulation is based on spread from person to person, once the last sick person is cured or dies, none new u infection in the current population can be developed. For example, when population is small, the rate of spread is not enough for the disease persists over 4 months, usually a stable state is reached and the simulation ends before 120 days.

The next section some experimental results of this parallel SEIR-Flu are dis-played.

5. Experimental Results

In this section, we show and analyze the experimental results for the parallel solution of SEIR-Flu.

The environment of simulations, sequential and parallel, were in a multi-core computer whose characteristics are: 2 processors AMD Opteron 6272, 2.1GHz, 16 cores per processor, RAM memory of 64GB Memory (16x4GB), 1333MHz, OpenMP 4.0 and Debian 8.

In order to obtain the results, different scenarios of simulation are considered. Each of them is a combination of next parameters:

- *The maximun time of simulation:* 120 days. The stational flu has more propagation in winter, in others seasons its infection decreases drastically.
- *Virtual clock:* A timestep is equivalent to 1 hour.
- *Initial infection factor:* 1% of population has flu.
- *Population:* 50% of cells of grid.
- *Size of Square Grid:* 50, 100, 200, 400, 800, 1600 and 3200 by side. For lack of space, we only report the more representative results.

We consider a stationary type of population pyramid [13], each individual has an age between 1 to 90 years. The degree susceptibility to infection and the mortality rate is determined according of individual age[14, 12]. In all populations, there are three groups of person: Children (Up to 6 years old), Young-Adult (7 to 60 years) and Elderlies (Greater than 61 years), each of them has a susceptibility equal to 35%, 20% and 35% respectively.

For mortality, we recognize the following groups and each mortality rate: Under 3 years (8%), 4 to 10 years (5%), 11 to 18 years (2%), 19 to 50 years (0.5%), 51 to 60 years (2%), and more than 60 years (8%).

Each reported value is the average of 10 executions of Parallel SEIR-Flu. For parallel solution and each scenario, we use 2, 4, 8, 16, 32 and 64 threads. To sequential solution, we consider the same solution but for only one processor. In first time, we evaluate if our parallel proposal works well and its performance is better than sequential solution. The figure 2(a) shows the speedup reached for every simulation scenario and different numbers of threads. In majority of scenarios, we achieve speedup. Although this is not close to optimal, we can reduce significantly the computational time of simulation. Similar behaviors and earnings are observed in Figure 2(b) for the same size of population (50%) but the 10% of them are infected initially.

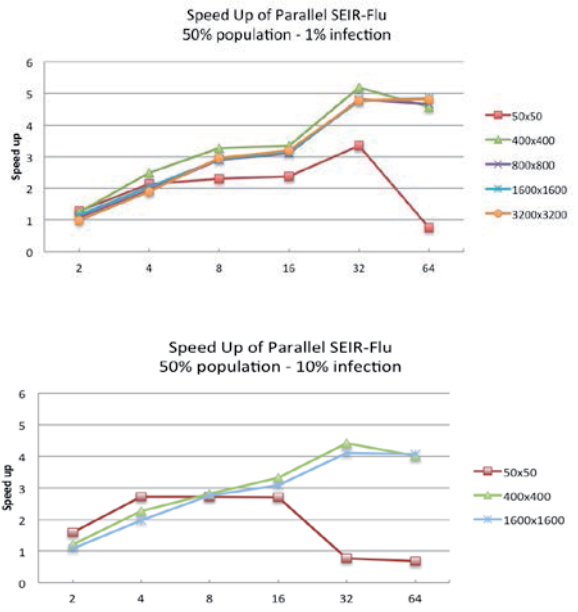


Fig. 2. Speedup of Parallel SEIR-Flu for 50% population with 1% and 10% initial infection

Besides of performance results, we can observe some behavior of the flu infection process. For example, a higher level of initial infection implies less time to reach a state free of u in the population of each grid, the Figure 3 shows this situation.

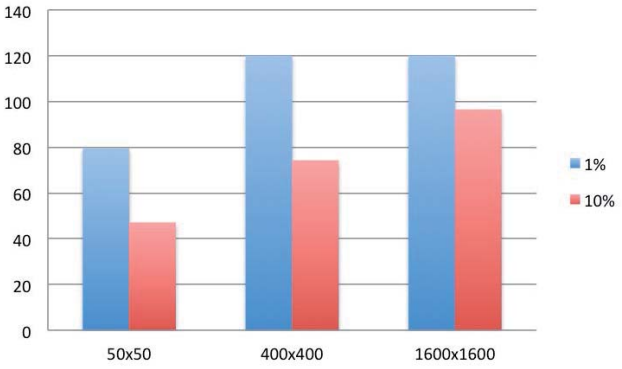


Fig. 3. Days of SEIR-Flu for 1% and 10% initial infection

In Figure 4, we sketch the percentage of infection (4(a)) and mortality (4(b)) for each populations group: Children, Young-Adults and Elderlies. The grid occupation is 50% and the 10% are infected.

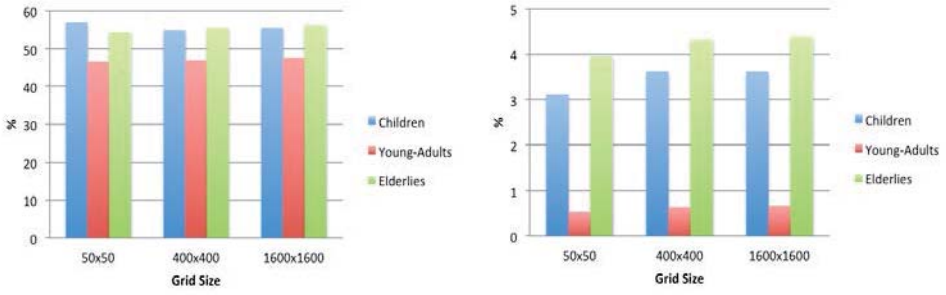


Fig. 4. Percentage of Infected (left) and Mortality (right) Individuals for 50% population and 10% initial infection

Figure 5 summarizes the speedups achieved for different levels of grid occupation when 1% of initial population is infected (Figures 5(a), 5(b) and 5(c)). Besides, Figure 5(d) displays how many days are necessary to reach a state free of flu.

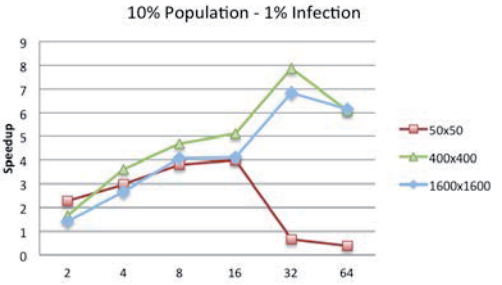


Fig. 5a. 10%

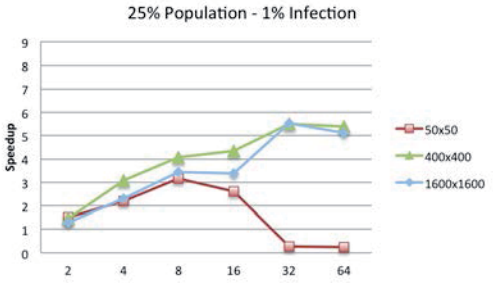


Fig. 5b. 25%

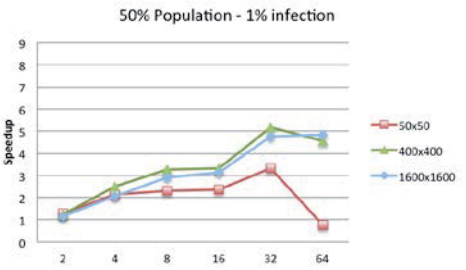


Fig. 5c. 50%

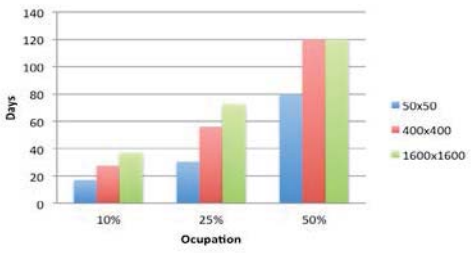


Fig. 5d. Days.

Fig. 5. Speedup and Days of SEIR-Flu for different levels of occupation with 1% of initial infection and occupation percentage.

6. Conclusions and Future Work

We presented a parallel model capable to simulate the propagation of seasonal u. The proposed model uses the CA and SEIR concepts to analyze the effects of seasonal influenza in a population with a certain territorial distribution modeling the people interaction.

One of objectives of this work is to develop and select techniques of high performance computing in order to perform and execute large-scale complex simulations of diffusion processes. While our proposal is not yet complete, the results of our multithreading implementation are encouraging and give us an initial framework for future works in the characterization of diffusion processes like signal propagation, avalanches, rumors spread, etc.

As future works, it is important to decide the optimal size of set of threads needed to solve the model taking into account the performance parameter to try characterize this kind of applications and this experience will be transferred to others developments of more complex parallel applications.

References

1. F. Chierichetti and A. Panconesi. Rumor spreading in social networks. Theoretical Computer Science, 412(24):2602{2610, 2011.
2. M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili. Theory of rumour spreading in complex social networks. Physica A: Statistical Mechanics and its Applications, 374(1):457{470, 2007.
3. S. White, A. del Rey, and G. Sanchez. Modeling epidemics using cellular automata. Applied Mathematics and Computation, 186(1):193{202, 2007.
4. Huppert and G. Katriel. Mathematical modelling and prediction in infectious disease epidemiology. Clinical Microbiology and Infection, 19(11):999{1005, 2013.
5. T. Johnson and B. McQuarrie. Mathematical modeling of diseases: Susceptible-infected-recovered (sir) model. In University of Minnesota, Morris, Math 4901 Senior Seminar, 2009.
6. D. Reyes. Descripción y aplicaciones de los automatas celulares, 2011. Technical Report, Autonoma de Puebla.
7. S. Kau man. Emergent properties in random complex automata. Physica D: Nonlinear Phenomena, 10(1):145-156, 1984.
8. S. Wolfram. Universality and complexity in cellular automata. Physica D: Non-linear Phenomena, 10(1):1-35, 1984.
9. J. Klüver and C. Klüver. On communication. An interdisciplinary and mathematical approach, volume 40.
10. B. Chapman, G. Jost, and R. van der Pas. Using OpenMP: Portable Shared Memory Parallel Programming. Number v. 10 in Scientific Computation Series. MIT Press, 2008.
11. C. Beauchemin, J. Samuel, and J. Tuszynski. Simple cellular automatons model for influenza viral infections. Journal of theoretical biology, 232(2):223-234, 2005.
12. World Health Organization. Influenza (seasonal), 2014. Fact sheet N211.
13. P. Hernandez. Demografía y antropología demográfica. CONACULTA-INAH, 2004.
14. Argentina Ministerio de Salud. Manual para el fortalecimiento de la vigilancia de la enfermedad tipo Influenza utilizando la estrategia de Unidades Centinela. 2011.

XV

Information Technology Applied to Education Workshop

Modeling Students through Analysis of Social Networks

María Emilia Charnelli[1], Laura Lanzarini[2], Javier Díaz[1]

[1] LINTI - Laboratory of Research in New Information Technologies
[2] III LIDI - Institute of Research in Computer Science LIDI
Computer Science School, National University of La Plata
mcharnelli@linti.unlp.edu.ar , laural@lidi.info.unlp.edu.ar , jdiaz@unlp.edu.ar

Abstract. Educational Data Mining gathers the multiple methods that allow new and useful information extraction from great volumes of data coming from the educational context. The goal of this article is to obtain a model of the students of the Computer Science School of the UNLP from their participation in Facebook. The work describes the process of extraction of latent topics in posts made in public groups related to the School, and the modeling of the students from the topics discovered. Additionally, it includes the preprocessing done to the collected data, which constitutes a fundamental stage since it strongly conditions the performance of the models to be obtained. Finally, obtained results are presented together with conclusions and future lines of work.

Keywords: Educational Data Mining, Learning Analytics, Topic Modeling, User Modeling, Recommender Systems.

1. Introduction

In the last few years, educational institutions have embarked on their own exploration of big data sets to increase retention rates and provide students with a customized and higher quality experience. Applying data mining techniques in education has allowed to characterize the actors involved in teaching and learning processes. Generally speaking, it is very difficult to exploit available information to create models that describe students objectively. In particular, extracting information from the web constitutes a significant challenge due to the unstructured nature of the data it contains. Social networks make up an environment external to the educational institution, however containing valuable information regarding the interests of the students. Current research focuses on methods for extracting implicit information on student behaviors from social platforms in order to obtain dynamic models that are capable of easily adapting to changes in information and contributing to decision-making processes in education.
Analyzing latent topics has emerged as one of the most efficient methods to classify, group and retrieve textual data, such as those found in social network posts. Many latent topic modeling methods have been developed and

studied extensively, such as PLSA[1] and LDA [2]. In these models, documents are modeled as mixtures of topics, where a topic is a probability distribution of all the possible words in the documents. Statistical techniques are used to learn the topics and the mixing coefficients for each document. Conventional topic models reveal latent topics by discovering word co-occurrence patterns in all documents [3] [4]. Applying traditional topic modeling techniques on short texts such as tweets, Facebook status updates and instant messaging may not yield optimal results since they lack rich contexts. The main reason is that topic modeling implicitly captures word co-occurrence patterns by document in order to discover topics, therefore there is a severe data dispersion in shorter documents. More specifically, word occurrence in short texts plays a less discriminatory role in comparison to longer documents where the model has a number of words that is sufficient to know how they are related [5]. Discovering topics in short texts is crucial for a wide range of topic analyzing tasks, such as characterizing content [6] [7], modeling user interest profiles [8][9] , and detecting latent or emerging topics [10]. BTM (Biterm Topic Model) [11] is an effective way to learn latent topics in short texts. BTM extracts underlying topics in a set of documents and a global distribution of each topic in each of them, through an analysis of the generation of biterms. Likewise, BTM has extensions to treat data flows and an incremental scheme for updating the model and giving more importance to the latest data collected.

This paper presents a method for obtaining a model of the students of the Computer Science School of the National University of La Plata (UNLP) through an analysis of the posts they make in public student groups in the Facebook social network.

This work is organized as follows: the second section describes the preprocessing effected on the collected data, the third section shows the extraction and modeling of latent topics, the fourth section shows the construction of student models through discovered topics, and the fifth section shows results obtained. Finally, the sixth section presents conclusions and future work.

2. Data Preparation

The information used in this work comes from Facebook posts in groups created by Computer Science School students. The data were collected through the Facebook Graph API and involve over 3000 posts and 1500 students that write, comment, share and ``like'' posts. The Graph API is the main way of consulting and collecting data in the Facebook platform. Posts, group information and public user information were gathered. The following information was obtained from each of the posts: id of the post, creation date, author, likes, comments, shares and unstructured text content. When operating with textual information, it is necessary to use Text Mining techniques, in order to represent each post in a vector of terms. This was achieved through a process comprising many stages. The first stage consisted

of the application of a stopwords filter, which filters the words that match any indicated stopword. Stopwords were filtered in Spanish and English, using words from the social networks context such as smileys, greetings, etc; words from the context of the group such as university, school, informatics, etc; words that refer to web page addresses; words that are between symbols, etc. Likewise, students post code fragments that constitute solutions to exercises included in courses. In general, the most exercises are from the first year and written in Pascal. Therefore, each sentence or signature of the language used was reduced to a single word in order to better identify them. Following, each word in the text was reduced to its root applying the Snowball [12] stemming algorithm. The importance of this process lies in that it eliminates syntactic variations related to gender, number and tense. The algorithm was applied for both the English and Spanish languages. Once the roots of each of the words were obtained, the frequency of occurrence was calculated for each of them, and words that appeared more than once were chosen.

3. Extraction of latent topics

BTM was used for extracting topics in Facebook groups, which is an unsupervised learning technique that discovers topics characterizing a set of brief documents. In this context, each post and comment is considered a document.

Let a set N_D of documents be called corpus and let W be the set of all the words in the corpus, a topic is defined as a probability distribution of W.Therefore, a topic may be characterized by its T most likely words. Given a number K of topics, the goal of BTM is to obtain the K distributions of each of the words.

A "biterm" is a pair of words that co-occur without a set order in a short document. In this case, two different words in a document constitute a biterm. Given a corpus of N_D documents and a W vocabulary of unique words, it is assumed to contain N_B $B = \{b_i\}_{i=1}^{N_B}$ biterms with $b_i = (w_{i,1} \in W, w_{i,2} \in W)$, and K topics expressed of W. Let $z \in [1, K]$ be a variable to indicate a topic. The $P(z)$ probability that a document in the corpus is of a z topic is defined as a $\theta = \{\theta_k\}_{k=1}^{K}$ K-dimensional multinomial distribution with $\theta_k = P(z = k)$ and $\sum_{k=1}^{K} \theta_k = 1$.

The $P(w|z)$ distribution of words per topic can be represented as a $\Phi \in R^{K \times W}$ matrix where the ϕ_k k-th row is a W-dimensional multinomial distribution with $\phi_{k,w} = P(w|z = k)$ entry and $\sum_{w=1}^{W} \phi_{k,w} = 1$.

Given the α and β parameters, the main assumption of the model is that each of the documents of the corpus was generated thus:

A $\theta \sim Dirichlet(\alpha)$ topic distribution is chosen for all the corpus

For each $k \in [1, K]$ topic

- A $\phi_k \sim Dirichlet(\beta)$ distribution of words is extracted for the topic

For each $b_i \in B$ biterm

- A $z_i \sim Multimonial(\theta)$ topic assignment is extracted
- Two $w_{i,1}, w_{i,2} \sim Multimonial(\phi_{z_i})$ words are extracted

Taking into accont the generation mechanism assumed by BTM, likelihood for all the corpus can be obtained given parameters α and β from the probability of each of the biterms:

$$P(B|\alpha, \beta) = \prod_{i=1}^{N_B} \int \int \sum_{k=1}^{K} P(w_{i,1}, w_{i,2}, z_i = k|\theta, \Phi) d\theta d\Phi \quad (1)$$

$$= \prod_{i=1}^{N_B} \int \int \sum_{k=1}^{K} \theta_k \phi_{k,w_{i,1}} \phi_{k,w_{i,2}} d\theta d\Phi$$

(2)

Obtaining exactly the θ and Φ parameters that maximize the likelihood of equation 2 is an untreatable problem. Following the proposals in [13], parameters θ and Φ can be approximated using Gibbs sampling [14].

3.1 Co-occurrence Matrix

In order to evaluate the quality of the topics obtained, the coherence metric proposed by Mimno et al. [15] is used. Given a z and its T most likely words $V^{(z)} = (v_1^{(z)}, ..., v_T^{(z)})$ where $v_i^{(z)} \in W$ for $i = 1...T$, the coherence score is defined as:

$$C(z; V^{(z)}) = \sum_{t=2}^{T} \sum_{l=1}^{t} \log \frac{D(v_t{}^{(z)}, v_l{}^{(z)}) + 1}{D(v_l{}^{(z)})}$$

where $D(v)$ is the frequency of word v in all documents, $D(v, v')$ is the number of documents where words v and v' co-occur. The Coherence metric is based on the idea that words that belong to one concept will tend to co-occur in the same documents. This is empirically demonstrable since the coherence score is highly correlated with human criteria. In order to evaluate

the general quality of a set of topics, the $\frac{1}{k}\sum_{k} C(z_k; V^{(z_k)})$ average of the coherence metric is calculated for each of the topics obtained. These results allow us to determine the amount of topics that best represent the entire corpus.

4. Modeling the Students

Once the K topics representing Facebook group posts collected were obtained, the students were modeled as vectors in a K-dimensional space. Each position of the vector represents the level of participation of the student in each of the topics. The level of participation is associated with possible actions performed by a user on the contents. Possible actions are creating content, commenting or liking a post. Given n users, $X \in R^{n,K}$ is defined as the matrix that contains in its rows the representation of each user in the new characteristics space. Let $A_{l,m} = \{a_1, a_2, ..., a_t\}$ be the set of actions of user l on posts classified in topic m. Then the l, m-th component of matrix X is defined as:

$$\text{X[l,m]} = \max\left(\sum_{i=1}^{t} w(a_i), 1\right), \text{ with l=1...n } and \text{ m=1...K}$$

where $w(a_i)$ is the weight associated with action a_i. $w(a_i)$ is defined as:

$$w(a_i) = \begin{cases} 1/20 \text{ if } a_i \text{ is a new post} \\ 1/40 \text{ if } a_i \text{ is a comment} \\ 1/50 \text{ if } a_i \text{ is a ``like''} \end{cases}$$

Cosine similarity is used in order to evaluate the similarity between two users in the new space of characteristics. Given $v_1, v_2 \in R^K$, the similarity function is defined as:

$$d(v_1, v_2) = \frac{v_1 v_2}{||v_1 v_2||} = cos(\theta)$$

5. Results

The model obtained with BTM was evaluated in the set of Facebook group posts. For each number of topics between 2 and 45, the average of the obtained coherence was calculated sampling the test and training set randomly in 1000 iterations. Figure 1 shows the average of the coherence of the topics model according to the number of topics extracted. What is interesting is the number of topics in which there is a breaking point in the

growth of the average coherence function. In this case, the optimal value is between 10 and 12 latent topics. Once the optimal number of topics was obtained, each of the Facebook group posts was classified into a topic according to the latent topics model that was obtained.

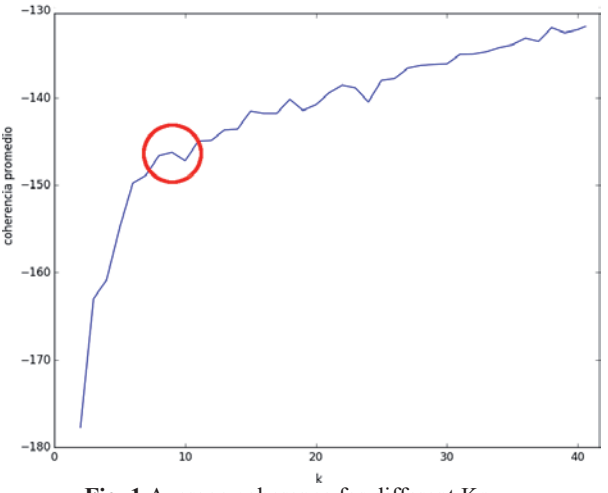


Fig. 1. Average coherence for different Ks

Table 1 shows topics obtained with K=10. For each of the topics, the six most important words are shown, i.e., those with the most likelihood of belonging to the topic. For example, topic 1 is about scholarships, topic 9 is about the first year programming subject where Pascal is used, and topic 10 contains posts on student accommodation search and offer.

Table 1. Model of the topics obtained with BTM

Topic	Most important words of the topic				
1	enrollment	students	scholarships	university	national
2	support	workshops	courses	student	matters
3	community	sharing	tools	hacking	security
4	classroom	midterm	final	date	adp
5	courses	enrollment	information	page	systems
6	work	experience	knowledge	development	java
7	file	commands	linux	ubuntu	text
8	tutoring	classes	study	question	mathematics
9	pascal	close	file	enter	reset
10	students	double	simple	rooms	foreigners

Afterwards, the actions performed by the users on each of the contents classified in the 10 topics were obtained which allowed for the generation of the users model described in section 4. Unsupervised learning techniques were applied to determine the underlying structure of student groups. In order to obtain the optimal number of user clusters automatically, the Silhouette [16] index was used, finding the value that optimized the criterion. Figure 2 shows the result of the evaluation of the model, generated by a hierarchical clustering algorithm with the average linkage criterion. It can be observed that the optimal value is implicitly imposed by the amount of topics obtained with BTM. In this context, it is interesting to model with greater detail the interests of the users. The following criterion-optimizing value is found in k=42.

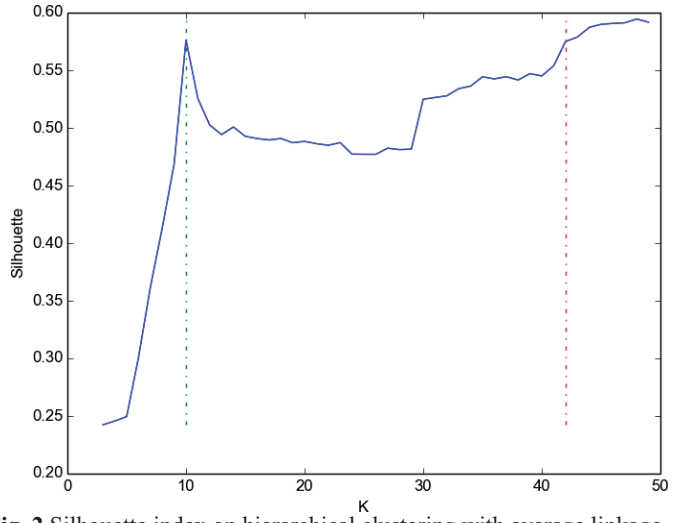


Fig. 2. Silhouette index on hierarchical clustering with average linkage

Figure 3 shows the dendrogram obtained by means of a hierarchical algorithm using the average linkage criterion with the cosine similarity metric, showing the obtained groupings and the correspondence with the cluster validity criterion.

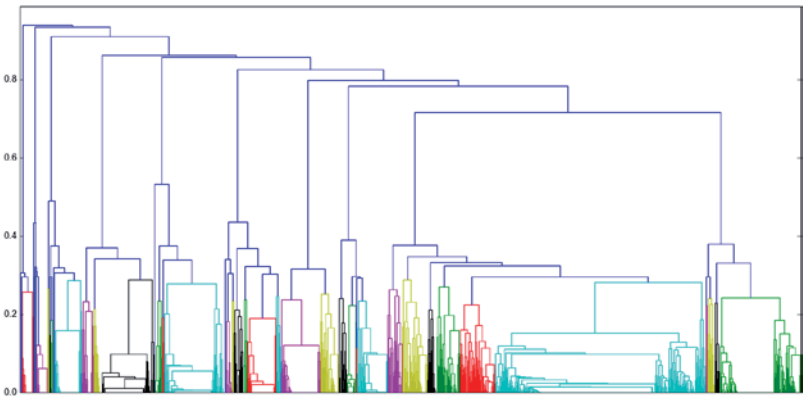


Fig. 3. Dendrogram obtained using average linkage and cosine distance

6. Conclusions and Future Work

This article presents a method for modeling the students of the Computer Science School of the UNLP through the detection of latent topics in their posts in public Facebook groups. This allows characterizing the students from a different context, knowing their topics of interest and how they relate to one another.

The methodology and the validation metrics used to obtain the user model presents satisfactory preliminary results. Topic coherence allows for automatic detection of the optimal number of topics and the user model presents compact groupings that characterize the behavior of the students in Facebook groups.

One of the future work points contemplates the implementation of an incremental model that allows updating the latent topics and therefore, the user modeling, from new posts in social networks.

The results of this work join those in [17] [18], which identifies features better characterizing students regarding their academic level through personal and academic information provided by the student management system. Results obtained will allow for the creation of an initial recommender system for the educational environment.

References

1. Hofmann, T.: Probabilistic latent semantic indexing. In: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, ACM (1999) 50–57.
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of machine Learning research 3 (2003) 993–1022.

3. Boyd-Graber, J.L., Blei, D.M.: Syntactic topic models. In: Advances in neural information processing systems. (2009) 185–192.
4. Wang, X., McCallum, A.: Topics over time: a non-markov continuous-time model of topical trends. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM (2006) 424–433.
5. Hong, L., Davison, B.D.: Empirical study of topic modeling in twitter. In: Proceedings of the first workshop on social media analytics, ACM (2010) 80–88
6. Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H., Li, X.: Comparing twitter and traditional media using topic models. In: European Conference on Information Retrieval, Springer (2011) 338–349.
7. Guo, J., Xu, G., Cheng, X., Li, H.: Named entity recognition in query. In: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, ACM (2009) 267–274.
8. Ramage, D., Dumais, S.T., Liebling, D.J.: Characterizing microblogs with topic models. International Conference on Weblogs and Social Media 5 (2010) 130–137.
9. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the third ACM international conference on Web search and data mining, ACM (2010) 261–270.
10. Lin, C.X., Zhao, B., Mei, Q., Han, J.: Pet: a statistical model for popular events tracking in social communities. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM (2010) 929–938.
11. Cheng, X., Yan, X., Lan, Y., Guo, J.: Btm: Topic modeling over short texts. IEEE Transactions on Knowledge and Data Engineering 26 (2014) 2928–2941.
12. Gupta, V., Lehal, G.S.: A survey of common stemming techniques and existing stemmers for indian languages. Journal of Emerging Technologies in Web Intelligence 5 (2013) 157–161.
13. Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proceedings of the National academy of Sciences 101 (2004) 5228–5235.
14. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on pattern analysis and machine intelligence (1984) 721–741.
15. Mimno, D., Wallach, H.M., Talley, E., Leenders, M., McCallum, A.: Optimizing semantic coherence in topic models. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics (2011) 262–272.
16. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20 (1987) 53–65.
17. Lanzarini, L., Charnelli, M.E., Baldino, G., Diaz, J.: Seleccion de atributos representativos del avance academico de los alumnos

universitarios usando técnicas de visualización: Un caso de estudio. Revista TE&ET (2015) 42–50.
18. Lanzarini, L., Charnelli, M.E., Diaz, J.: Academic performance of university students and its relation with employment. In: Computing Conference CLEI, 2015 Latin American. (2015) 1–6.

Learning Object Assembly Methodologies. In-Depth Analysis of the Underlying Concept of Learning Object.

ASTUDILLO GUSTAVO J.[1], SANZ CECILIA V.[2,3], SANTACRUZ-VALENCIA LILIANA P.[4]

[1] GrIDIE. Dpto. de Matemática, FCEyN, UNLPam, Av. Uruguay 151, La Pampa, Argentina, astudillo@exactas.unlpam.edu.ar,

[2] [3] III LIDI, Facultad de Informática, UNLP, Calle 50 y 120, La Plata, Argentina, Associate researcher, Scientific Research Agency of the Province of Buenos Aires (CICPBA), Argentina csanz@lidi.info.unlp.edu.ar,

[4] Escuela Técnica Superior de Ingeniería Informática (ETSII), Universidad Rey Juan Carlos, c/Tulipán s/n, Móstoles, Madrid, España, liliana.santacruz@urjc.es

Summary. In this article, we present an analysis of Learning Object (LO) Assembly Methodologies that focuses on the LO definition used by each methodology. This research is part of a more comprehensive study of this type of methodologies; however, this particular paper discusses the consequences of not agreeing on a definition of LO before starting the assembly process. To this end, a detailed analysis of 33 methodologies selected through bibliographic review is carried out. The results obtained point to the need of having an explicit definition of LO before working with one of these methodologies, as well as taking into account the main features of LOs as agreed by the academic community.

Key words: Assembly Methodologies, Learning Objects, Definition.

1. Introduction

Learning is a process that involves sequencing contents –and the strategies used to convey it–, and it is precisely this aspect of the educational process on which the assembly process focuses.

Assembly Methodologies (AMs) have the main purpose of defining a learning path based on a set of Digital Educational Materials (DEM) stored in repositories. The process involves three stages: definition of the topic, definition of the learning sequence, and incorporation of all relevant DEM.

This research work is part of a Master's thesis [1]. It included a Learning Object Assembly Methodology (LOAM) search and selection process. Thirty-three LOAMs were analyzed to define assessment criteria.

The analysis carried out, among other results [1], allowed corroborating that authors talk about the use of LOs by AMs to create learning path. However, the conceptualization and characterization of a LO differs from one methodology to another. This calls for an analysis focused on the definition and characterization of each AM before selecting one, and also before selecting the Assembler System (AS) used to implement it for use. In this sense, this paper provides information that will make the decision-making process easier, since it presents and sorts the definitions used by various authors, it uses assessment criteria that allow characterizing the type of educational material used by LOAMs, and discusses the importance of using a suitable definition.

This paper is organized as follows: Section 2 presents some background in relation to the concepts of Assembly Methodology, Assembler Systems and DEM used for assembly. Section 3 analyzes the various definitions of LO used in the AMs being studied; Section 4 discusses the criteria used to analyze LOAMs with particular focus on the criterion related to LO conceptualization, which is applied to carry out an analysis task based on this aspect. Finally, Sections 5 and 6 discuss the impact of not having a single definition of LO in Assembly Methodologies, and we present our conclusions.

2. Learning Objects and Assembly Methodologies

An assembly methodology is aimed at establishing a syntactic, structural and semantic correspondence between educational materials [2]. Thus, a learning path based on DEMs hosted in repositories could can be created. This action can be carried out with various automation levels, and it can take user profile into account (or not) [1].

ASs must use the DEMs that are available at local or remote stores. Materials must meet a set of requirements that will allow using them in the context of these systems. Basically, they must:

- include information that allows relating them to other materials
- include information that describes them (both technically and pedagogically)
- be hosted in repositories

As stated in [3] "[LOs] are seen as the technology of the future [...] due to their adaptability, reusability, and potential scalability". These advantages offered by the LO paradigm are supported by a type of educational material that, since its inception, proposes using the features of assembly, labeling through metadata, and design focused on reutilization [4, 5].

In order to maximize material reutilization (which is a distinctive feature of LOs), metadata must follow a standard and be hosted in repositories. Therefore, LOs meet, at least in theory, the requirements imposed by ASs on the educational materials used to create learning paths.

3. LO Definitions Used by LOAMs

For a preliminary analysis, the definitions used by LOAMs can be considered. It should be noted that only 20 out of the 33 methodologies analyzed define the concept of LO, the remaining 13 use it without proposing a definition.

The set of publications analyzed allows taking into account the variety and differences between the definitions of LO used by the different authors.

From all the publications that propose a definition of LO, some (8:20) opt for a generic concept in agreement with David Wiley's proposal [6]. Thus, a LO is defined as:

- "any digital document that can be used for learning [...and] are described by metadata" [7].
- "any piece of information (text, image, sound, etc.) that can be identified uniquely" [8].
- "raw content components or atomic learning objects (ALO) (text, figures, summary of text, keywords of text)" [9].
- "an element with recyclable digital documents with multimedia content that have a purpose and some use in teaching and learning, while meeting certain technological specifications" [10].
- "small curricular components that can be reused several times in different learning contexts" [11].
- "an atomic, self-contained learning object that is uniquely identifiable and addressable by an URI" [12].
- "teaching element based on the [OOP...] paradigm, oriented to the support of online learning, that is created only once and can be used many times more and in different contexts" [13].
- "an entity, digital or not, that can be used for learning, education or training" [14].

The following definitions (7:20) are characterized for attributing LOs the ability for combination or assembly, among other properties. Thus, LOs are:

- "is a unit of learning content with a specific leaming objective, often used as a building block in designing and assembling courses. The [LOs] that make up collections are tagged with metadata, such as topic or level of difficulty, and they are stored in a repository" [15].
- "the unit of assembly in larger teaching and learning environments constructed from smaller units" [16].
- "learning resources described by meta-data and organized in a multilayer structure, where the highest elements possess information about their associated knowledge that facilitates their assembly and reuse" [2].
- "small and easily reusable educational resources to be composed to allow personalized instruction and courseware creation" [17].
- self-contained learning materials that once developed can subsequently be exchanged, composed, and reused" [18].

- "structured, i.e., composed by one or more alternative sequences of other components, until reaching atomic LO" [19].
- "reusable units of educational content that can be sequenced into larger units to allow personalized learning" [20].

In the following five definitions, the authors take some of the basic features of a LO established by reference authors:

- "minimal reusable units of information [...that combine] three elements: contents, behavioural descriptions of the object, and a set of metadata which refer to the objects" [21].
- "fine granularity elements for knowledge transfer [… that] allow content engineers to design modular and self-contained learning units, and recompose them to new courses that can be offered in e-learning environments" [22].
- "Web accessible resources of pedagogical material can be designed for use and reuse in a variety of contexts, from remediation to lesson preparation and use of the lesson" [23].
- "fragments or parts of a course that can vary in size and complexity from a simple chart to the whole course". They must also be self-contained, independent, and reusable; they must have low granularity; and it must be possible to combine them, personalize them, and label them with metadata [3].
- "all the material structured in a significant way,and it must be related to a learning objective which must correspond to a digital resource that can be distributed and consulted online. A LO must also have [...] metadata that includes a list of attributes which not onlydescribes the possible attributes of an LO, but also allows to catalogue and exchange it" [24].

4. Assessment Criteria for the Analysis of LOAMs

4.1 Description of the Criteria

Based on the review of the selected publications that was carried out, different analysis focus were identified for LOAMs, and 17 criteria (see Fig. 1) were generated that allow both classifying the various assembly methodologies as well as selecting an AS based on how the methodology is to be used and the type of material that is available [1, 25].

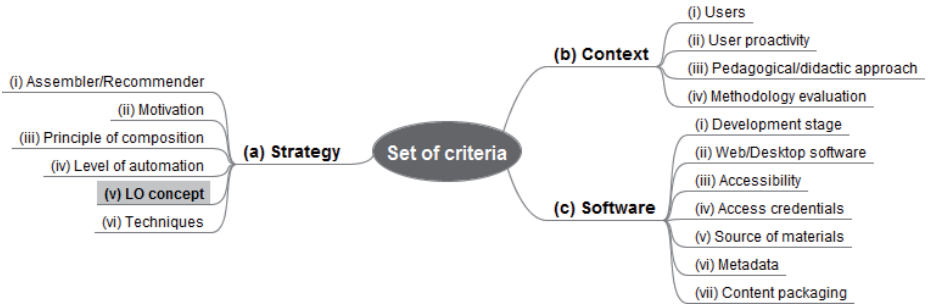


Fig. 1. Full list of criteria

Even though the publications selected work with LOs, not all of them have adopted the same definition for LO, nor they require the same features for an object to be considered a LO.

Out of all the criteria (Fig. 1), this paper focuses on *a.v Concept of LO*. This criterion allows assessing which of the features that are more generally considered by the specialized community to be required [26] are present (or not) in the concept of LO adopted by each AM. To that end, both the definition of LO used and the characterization of work materials were reviewed.

It should be noted that the feature "Assembled" is added, which is relevant for this work. Those cases that "Do Not Define" the concept of LO are also recorded (this does not affect the potential identification of some of the other features).

Since all LOAMs use digital material, this feature is present in all of the publications that were analyzed; therefore, it is not included as a specific label.

The labels used for this criterion are as follows[1]:

- Pedagogical Intent
- Internal Structure
- Metadata
- Reusable
- Self-Contained
- Interoperable
- Accessible
- Granularity
- Assembled
- Do Not Define

1 Due to space constraints, labels are not described in details; please see [1].

4.2 Criterion Application

When applying *a.v Concept of LO* (Fig. 2), it can be stated that there is no unanimous concept of LO. Additionally, even though there is no definition that is generally accepted by the experts, 13 of 33 AMs use the concept, but do not define it.

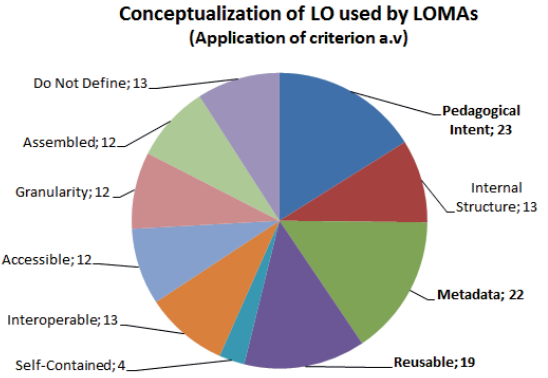


Fig. 2. Application of the criterion *a.v Concept of LO*

Three features that are used in most of the cases can be identified (Table 1): Pedagogical Intent (23 of 33), Metadata (22 of 33), and Reusable (19 of 33). A second group is formed by the features Interoperable (13 of 33), Assemblable (12 of 33), Accessible (12 of 33), With Definition of LO Structure (13 of 33), and Granularity (12 of 33). Only 4 of 27 require that the material be self-contained.

Table 1. Characterization of the LO proposed in the AMs that were analyzed.

ID*	Pedagogical Intent	Internal Structure	Metadata	Reusable	Self-Contained	Interoperable	Accessible	Granularity	Assembled	Do Not Define
S01	✓	✓	✓	✓				✓		
S03	✓		✓	✓					✓	
S04	✓			✓						
S05	✓		✓	✓		✓			✓	
S06	✓	✓	✓	✓		✓		✓	✓	
S07	✓		✓				✓		✓	
S08	✓	✓	✓	✓	✓	✓		✓	✓	
S13	✓	✓	✓			✓			✓	
S14			✓					✓	✓	
S15	✓		✓	✓			✓			✓
S16	✓	✓		✓				✓		✓

S17	✓			✓					✓	
S18	✓					✓				✓
S20	✓	✓	✓	✓		✓				✓
S21	✓	✓	✓			✓		✓	✓	
S22			✓	✓		✓	✓	✓		
S23	✓	✓	✓	✓	✓	✓	✓	✓	✓	
S24	✓	✓		✓		✓	✓			
S25							✓			✓
S27	✓	✓	✓				✓			
S28			✓	✓						✓
S30		✓	✓				✓	✓		✓
S31			✓			✓	✓			✓
S32	✓		✓	✓	✓		✓	✓		
S35	✓		✓	✓		✓	✓	✓		
S38	✓	✓	✓	✓	✓			✓	✓	
S39										✓
S40	✓									
S41						✓	✓			✓
S42	✓									✓
S43			✓							✓
S44				✓					✓	✓
S45	✓	✓	✓	✓						

** See Table 4.1, Chapter 4 in* [1] *for information about the authors.*

On the other hand, our analysis allowed identifying which methodologies require materials to be assemblable. As shown in Table 1, less than half (12 of 33) of the methodologies require this feature in LOs. Three of them (S03, S07 and S14) include the idea of building block in their definitions of LO, while the rest (S05, S06, S08, S13, S17, S21, S23, S38, S44) associate it to the possibility of assembling LOs to other LOs and create greater granularity DEMs.

5. Discussion

After more than 20 years[2], the definition of LO is still being discussed. Even though a large number of definitions have been proposed [26–28], the question "What is a LO?" still does not have a unique answer. This forces, or should force, those working with LOs to explicitly mention their definition of choice. Additionally, such definition should echo, at least, the main understandings reached by the community specialized in the paradigm. Otherwise, it is very likely that, when the time comes to actually use the LOs in a real context, resources or other types of educational materials will be used that should not be considered as a LO.

In the context of the LOAMs discussed in this article, we can see how the situation described in the previous paragraph comes to life. When analyzing

2 Wayne Hodgins introduces the concept for the first time in 1994.

the publications, three groups were identified based on how each LOAM represents LOs: (*i*) those that use a representation of the LOs through metadata, ontologies or web services (12 of 33); (*ii*) those that use packaging standards such as SCORM[3] or IMSCP[4] (10 of 33); and (*iii*) those that use other types of materials, such as slides (3 of 33), WebQuest[5] (1of 33), GLO[6] (1 of 33) or resources and activities available in the LAMS environment[7] (1 of 33). The remaining LOAMs (5 of 33) did not include any information about the type of materials used to represent/associate LOs.

By focusing on the definition of LO, LOs can be analyzed based on their characteristics required by the different LOAMs. Considering Figure 2, it could be stated that LOAMs in general define LOs as a reusable DEM labeled with metadata.

Similarly, an additional analysis of interest that can be done is comparing the definitions presented by LOAMs with a definition of the concept that takes into account the main characteristics on which the authors agree. In this article, the following definition of LO is used [30]:

"a type of DEM that is characterized, from a pedagogical point of view, for its orientation towards a specific learning objective, and having a series of contents that will allow presenting the topic related to the objective, activities that will allow students to put the contents presented into practice or consider related problems, and a self-evaluation that will allow students determine if they have understood the contents linked to the objective. From a technological standpoint, it is characterized for containing a set of standardized metadata used for search and retrieval operations, as well as for being integrated, using a standard-compliant packing model, which allows the interaction with different technology environments".

Table 2 proposes such comparison – due to space constraints, only those LOAMs with higher match rates with the characterization selected for reference are included. It can be seen that less than half the LOAMs use an appropriate concept for LO, and only two of them define the internal structure of the object, although in the case of those using SCORM/IMSCP, it could be assumed that they follow the structure proposed by the standard.

3 Sharable Content Object Referente Model (http://www.adlnet.org/scorm/).
4 IMS Content Packaging(https://www.imsglobal.org/content/packaging/index.html).
5 Lesson on how to conduct research using web information (http://www.webquest.org/).
6 Generative Learning Object [29].
7 Learning Activity Management System (http://lamsfoundation.org/).

Table 2. Comparison between the definition provided by [30] and the different concepts defined by LOAMs.

Sanz[30]	Pedagogical Aspects			Technological Aspects		
Criterion	Digital	Pedagogical Intent	Internal Structure	Metadata	Interoperable	SCORM IMSCP
S05	✓	✓		✓	✓	✓
S06	✓	✓	✓	✓	✓	
S20	✓	✓		✓	✓	✓
S23	✓	✓		✓	✓	✓
S35	✓	✓		✓	✓	✓
S01	✓	✓		✓		✓
S07	✓	✓		✓		✓
S08	✓	✓		✓	✓	
S13	✓	✓		✓	✓	
S18	✓	✓			✓	✓
S21	✓	✓		✓	✓	
S22	✓			✓	✓	✓
S24	✓	✓			✓	✓
S31	✓			✓	✓	✓
S45	✓	✓	✓	✓		

6. Conclusions

The criteria defined to carry out the analysis of LOAMs have allowed identifying the strengths and weaknesses of the methodologies reviewed. One of the main weaknesses is related to the aspects tackled in this paper, where it can be seen that the types of materials used, even though they are referred to as LO, in reality refer to different characterizations and formats.

When embarking on a comparative analysis of LOAMs, the lack of a preset definition and general consensus in relation to the concept of LO hinders our task. Especially so when in some cases, the publication used as source for some of the methodologies used is not explicitly mentioned.

Similarly, most of the definitions used by LOAM authors do not take into account the main characteristics that are generally accepted by the specialized community.

All of this leads to conclude that, in order to adopt this type of methodologies, and the ASs that implement them, the materials available should be analyzed in detail or *ad-hoc* materials adapted to those proposed by LOAMs should be designed.

DEM design, and LO design in particular, is a costly process, the same as their adaptation to a new format. Therefore, this would limit the adoption of LOAMs.

We believe that this type of research should continue to find clear and consensual answers to the question "What is a LO?" This would be useful both for designing and producing LOs, as well as for reusing LOs and adopting/creating new methodological/technological approaches that use and reuse this type of DEM, particularly for LOAMs.

References

1. Astudillo, G.J., Sanz, C.V., Santacruz Valencia, L.P.: Estrategias de diseño y ensamblaje de Objetos de Aprendizaje, http://sedici.unlp.edu.ar/handle/10915/53442, (2016).
2. Santacruz-Valencia, L.P., Delgado Kloos, C., Cuevas Aedo, I.: Automatización de los procesos para la generación ensamblaje y reutilización de Objetos de Aprendizaje, www.lite.etsii.urjc.es/liliana/Defensa_Tesis_LPSV.pdf, (2005).
3. St•nic•, J.L., Cri•an, D.A.: Dynamic Development And Assembly Of Learning Objects In A Math Learning Environment. J. Inf. Syst. Oper. Manag. 6, 29–40 (2012).
4. Wiley, D.: The post-LEGO learning object, http://opencontent.org/docs/post-lego.pdf, (1999).
5. Wiley, D.: The learning objects literature. In: Handbook of research on educational communications and technology. pp. 345–353. Taylor & Francis, New York/London (2007).
6. Wiley, D.: Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In: The Instructional Use of Learning Objects (2000).
7. Bouzeghoub, A., Selmi, M.: Authoring Tool for Structural and Semantic Coherence Validation of Composed Learning Objects. In: Advanced Learning Technologies, 2009. ICALT 2009. pp. 175–177. , Riga, Letonia (2009).
8. Rigaux, P., Spyratos, N.: Selene report: Metadata management and learning object composition in a self elearning network. Last Accessed Sept. (2007).
9. Schreurs, J., Dalle, R., Sammour, G.N.: Authoring Systems Delivering Reusable Learning Objects | Schreurs | International Journal of Emerging Technologies in Learning (iJET). Int. J. Emerg. Technol. Learn. IJET. 4, 37–42 (2009).
10. Menéndez Domínguez, V.H., Castellanos Bolaños, M.E., Zapata González, A., Prieto Méndez, M.E.: Generación de objetos de aprendizaje empleando un enfoque asistido. Pixel-Bit Rev. Medios Educ. 141–153 (2010).
11. Roig Vila, R.: Diseño de materiales curriculares electrónicos a través de Objetos de Aprendizaje. RED Rev. Educ. Distancia. 1–9 (2005).
12. Ullrich, C., Melis, E.: Pedagogically founded courseware generation based on HTN-planning. Expert Syst. Appl. 36, 9319–9332 (2009).
13. López, M.G., Miguel, V., Montaño, N.E.: Sistema Generador de AMBientes de Enseñanza-ApRendizaje Constructivistas basados en Objetos de Aprendizaje (AMBAR): la Interdisciplinariedad en los ambientes de aprendizaje en línea. Rev. Educ. Distancia. 1–14 (2008).
14. Becerra, C., Astudillo, H., Mendoza, M.: Improving learning objects recommendation processes by using domain description models. Conf. LACLO. 3, (2012).
15. Farrell, R.: Dynamic Assembly of Learning Materials in a Corporate Context. Educ. Technol. 46, 70–73 (2006).

16. Pahl, C., Barrett, R.: A web services architecture for learning object discovery and assembly. In: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters. pp. 446–447. ACM, New York, NY, USA (2004).
17. Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F.M., Ragone, A.: Semantic-based automated composition of distributed learning objects for personalized e-learning. In: The Semantic Web: Research and Applications. pp. 633–648. Springer (2005).
18. Li, Y., Huang, R.: Dynamic composition of curriculum for personalized e-learning. IOS Press. 151, 569–576 (2006).
19. Lopes Gançarski, A., Bouzeghoub, A., Defude, B., Lecocq, C.: Iterative search of composite learning objects. In: IADIS Internationnal Conference WWW/Internet. pp. 8–12. , Vila Real, Portugal (2007).
20. Karam, N., Linckels, S., Meinel, C.: Semantic Composition of Lecture Subparts for a Personalized e-Learning. In: Franconi, E., Kifer, M., and May, W. (eds.) The Semantic Web: Research and Applications. pp. 716–728. Springer Berlin Heidelberg (2007).
21. Sarasa, A., Piquer, J., Arriola, R., Iglesia, S.: LOMEditor: Composition and Classification of Learning Objects. In: Mendes, A., Pereira, I., and Costa, R. (eds.) Computers and Education. pp. 241–249. Springer London (2008).
22. Wetzlinger, W., Auinger, A., Stary, C.: Ad-hoc Composition of Distributed Learning Objects using Active XML. Int. J. Emerg. Technol. Learn. IJET. 3, 33–39 (2008).
23. Kellar, M., Stern, H., Watters, C., Shepherd, M.: An Information architecture to support dynamic composition of interactive lessons and reuse of learning objects. In: System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on System Sciences. p. 10 pp. IEEE, Hawaii (2004).
24. Torres, I.-D., Guzmán-Luna, J.A.: Composition of Learning Routes Using Automatic Planning and Web Semantics. In: Sobh, T. and Elleithy, K. (eds.) Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering. pp. 321–328. Springer International Publishing, Online (2015).
25. Astudillo, G.J., Sanz, C.V., Santacruz Valencia, L.P.: Criterios para evaluar metodologías de ensamblaje de objetos de aprendizaje. In: 2015 International Symposium on Computers in Education (SIIE). In Print, Spain (2016).
26. Astudillo, G., Sanz, C., Willging, P.: Análisis del estado del arte de los objetos de aprendizaje. Revisión de su definición y sus posibilidades, http://sedici.unlp.edu.ar/handle/10915/4212, (2011).
27. Callejas Cuervo, M., Hernández Niño, E.J., Pinzón Villamil, J.N.: Objetos de aprendizaje, un estado del arte. Entramado. 7, 176–189 (2011).
28. Santacruz Valencia, L.P.: Objetos de aprendizaje: estado de la cuestión. In: Actas del II Seminario de Investigación en Tecnologías de la Información: SITIAE 2008. pp. 67–79. Universidad Rey Juan Carlos (2009).
29. Damaaevi•ius, R., ` tuikys, V.: On the Technological Aspects of Generative Learning Object Development. In: Mittermeir, R.T. and Sysło, M.M. (eds.) Informatics Education - Supporting Computational Thinking: Third International Conference on Informatics in Secondary Schools - Evolution and Perspectives, ISSEP 2008 Torun Poland, July 1-4, 2008 Proceedings. pp. 337–348. Springer Berlin Heidelberg, Berlin, Heidelberg (2008).
30. Sanz, C.V.: Los objetos de aprendizaje, un debate abierto y necesario. Bit Byte. 1, 33–35 (2015).

XIV

Graphic Computation, Images and Visualization Workshop

Software tools for detecting and tracking people on video cameras

LEONARDO D. DOMINGUEZ[1,2], ALEJANDRO J. PEREZ[1], ALDO J. RUBIALES[1,3], JUAN P. D'AMATO[1,2], ROSANA BARBUZZA[1,3].

[1]Instituto PLADEMA, Universidad Nacional
del Centro de la Provincia de Buenos Aires,
Campus Universitario S/N, 7000 Tandil, Argentina

[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

[3]Comisión Nacional de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA), Argentina

{jpdamato, crgarcia, mlazo, cifuente}@exa.unicen.edu.ar

Abstract. Insecurity is a problem that affects all the cities of the world. Most digitalized cities make use of video surveillance to deal with it, mounting monitoring centers handling hundreds and even thousands of cameras. In many of this centers, trained employees perform the observation task, however, the actual technology gives us the possibility to automatize many of these daily tasks. In this paper, a video analysis platform developed in UNCPBA to easy people tracking through different cameras is presented. This platform uses image techniques to project the moving detected points from the different cameras to a single georeferenced space. Different algorithms were evaluated, and general problems that generally appears in this type of systems are explained. Some preliminary obtained results in multi-tracking are presented.

Keywords: Videosurveillance, Tracking multi-camara, Security

1. Introduction

Security is becoming one of the fundamental concern of society. The high rates of violence and insecurity [7] in countries such as Argentina, motivate the study of new technologies that allow people to live more peacefully, especially in big cities, where these dangerous situations are more frequent.
Between 2000 and 2008 [10], the rate of penitentiary population in Latin America grew about 42%, and according to trends in WEB searchers, in the last 10 years the interest of people in terms such as "Theft" and, "Insecurity" are constantly increasing.
In contrast, countries like United States have agencies to attend this concern. In particular, the DARPA agency has more than 60 active projects, of which we can mention Satellite Remote Listening System and Combat Zones That See, projects that aim to record everything that is moving through videocameras [12]. Another paradigmatic case is the United Kingdom, which according to

[1], they estimated that in London any individual could be captured daily from approximately 300 cameras.
In general, current systems are mostly oriented as monitoring tools, also known as CCTV (Closed Circuit Television), which store the images in a device for future visualization or to be used as evidence. At the same time, surveillance cameras manufacturers offer simple tools which simply connect and record what the cameras observe. However, thinking about a system with certain automations tasks, this should be able to monitor an area, register alarms, classify and count people and even be should be able to follow their movements. Even more, operators should only focus on important events so that they can apply the corresponding reaction protocols.
With a similar proposal, the project [14] has been started, which intends from an extensive network of sensors and wireless devices, to control large metropolitan areas in an intelligent way.
Undoubtedly, with the rise of technology in terms of image capture and video, it is possible to acquire cameras with good definition at a low cost, which allows a tool like this can be used not only in security, but also in crowded places that require analyzing patterns of daily mobility, traders could set up marketing strategies with more arguments [13].
Understanding that security is a very important issue, in this work, we will present a distributed platform for the automation of tasks of detecting and tracking people and objects in different video formats. For this task it is necessary to unify the different points of view of each camera, to take it to a unique geo-referenced space. With this objective, calibration tools have been developed to obtain the transformations that allow a person to be placed on a map. One of the main algorithms is the one that performs the motion detection, analogous to the background subtraction (BGS) and objects detection. What is innovative about the platform is that it is prepared to have different algorithms in a simple way and it is distributed architecture allows it to scale to support many cameras without affecting the overall performance.
The paper is presented as follows. In section 2, the existing tools are presented with similar purposes and a summary of the state of art. Subsequently, Section 3 lists two of the detectors used and how they are mapped to the plane. Section 4 presents some preliminary results of the routes of a person along with the problems encountered, and in section 5 the conclusions and future works are presented.

2. State of the Art

Currently there are a plenty of video surveillance systems. Regarding people tracking topic, there is a lot of research papers and it is studied in recent years from many groups, thanks to the improvement of computing capabilities.
Research works such as [3], deals with the most common algorithms for detecting and tracking objects. In particular in [5], an interesting comparison between the different objects detectors is made. It is concluded that the combination of detectors can be very useful for decreasing the rate of false

positives while maintaining the time and rate of true positives.
There are also commercial solutions such as like Blue Iris at a cost of U$S60 with support up to 64 cameras or EyeLine at a cost of U$S250 without connectivity limits. On the other hand, there are Open Source platforms like ZoneMinder or Ispy which have highly evolved although in the last years, but they are limited to operate in a single computer. The hardware features recommended by all the solutions are: Intel Core i7, 8 gb RAM or more, SSD disk, Nvidia accelerator board for hardware decoding and Windows 8 or higher.

2.1 Argentinian legislation

Argentinian authorities' are doing a great effort to provide better security facilities for people. On the other hand, it is being debated how the privacy is been violated as the amount of surveillance systems increases. We still need laws to cover these issues.
According to article [2] published in JAIIO 2015, 87.5% of Argentinian provinces do not have adequate or complete regulation, and even worse, they are very different. The authors emphasize the great disparity in these regulations, for example, in how many days video images should be stored before they can be deleted. On one hand, Santa Fe province defines 30 days, and Corrientes and San Luis provinces defines 2 years. In addition, most of the regulations do not conform to the principles established by DNPDP provision 10/2015 [15].

3. People detection and tracking

3.1 Movement detection

The first step for tracking people is to be able to determine whether or not there was movement from a series of images or video. This problem, analogous to the background subtraction problem, which has been very studied. Multiple solutions already exist to calculate the variation between consecutive images [9] [11].
The methods vary in computational complexity and efficiency of the result. When cameras are outside, some factors occur that affects detection, such as the time of day, cloudiness or other weather conditions. To counteract these effects, movement windows are usually used as mentioned in [8], whose objective is to apply a logical operation, to process only the movements that are detected within its dimensions and to discard all those that are produced by outside. In general, the output of this step is a set of pixels (grouped or not) that have been marked as objects moving.
For the platform development, the capabilities of the AForge package (open source) were used, with the detectors included to make the first tests of the system. Observing that all report the level of motion detected in the video with values between 0 and 1, which allows the programmer to set a threshold to

compare the movement and define different types of alarms, which is one of the objectives of the project. Today they are being tested with other algorithms, such as VIBE [4]

- Two Frame Difference (TF): It is the simplest and fastest detector. It is based on finding the difference between two consecutive frames.

- Simple Background Modeling (BM): In contrast to the above, this detector is based on finding the difference between the current frame and a defined frame representing the background. There are also techniques for updating the background as time progresses.

Fig. 1: Left: object with shadow detection. Right: mapping zone

One of the most problematic situation that affect the result of a detection is the shadow. To minimize the error caused by the shadow, it is proposed to use the central pixel of the detected object. Other works propose to use the minimum coordinate "Y" of the detected object (it is considered that is the location of the feet), but this does not always correspond at all points of observation. It is intended in the future to improve these algorithms, applying filters that first eliminate the shadow, from the analysis of the variation of lighting.

3.2 Tracking over a map

In order to map the detected people in each camera onto a satellite map of the observed area, it is necessary to calibrate each camera. The output of this process is a homographic matrix, which allows the mapping of each pixel in a geo-referenced point. This matrix is obtained manually, using a some tools that allow a user to mark several reference points from the image and the map and infer the corresponding matrix, called HS. The points are chosen over a rectangular planar region.
This matrix is then stored in the Database and could be accessible from each camera. This process is error-sensitive, so it is repeated several times, calculating the matrix using the average of selected points. In real time, for

each new detection in the space of the image, the transformation is applied on the most representative moving points.
Figure 2 shows the path of the pipeline of detecting and mapping a person in a monitoring center in a simplified form. First we used the object detection obtained and then a spatial transformation is applied on a map.

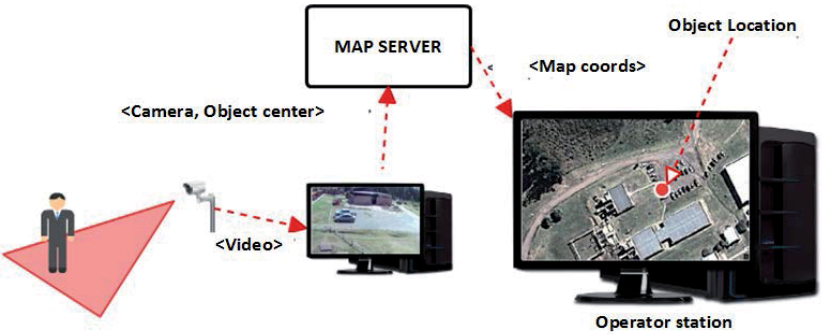


Fig. 2: From detection to visualization

3.3 Calibration

Since an Hs matrix relates points of two images, and taking into account that a different one is required to calibrate each camera with each satellite map, a tool was implemented that allows the operator to load two images (a map and a frame of one Camera) and establish a reference between them from a same square that can be observed in them, so that the perspective matrix necessary to adapt them can be calculated.
The process of the same is divided into 3 flaps (Figure 3), Map, Camara and Result. In the first, it is possible to load the map and in the second the image of the camera. Then clicking on both images you can define the vertices of the square. The tool will also allow to define the point where the camera is positioned and calculate the measurements of the square (width and height) and its angle of rotation α with respect to the X axis of the satellite provided by the map, with the purpose of generating the homographic matrix of the camera, and a matrix that contemplates all the necessary transformations between both images.

Fig. 3: Calibration tool. From Left to Right: Map, Camera and Result

3.4 Data integration

The developed platform is designed to support multiple cameras, so it works in a distributed way. In this sense, based on [6], it was thought to have several CPU machines (desktop computers or micro-PCs), all connected to a central server. Each computer connects with one or more IP cameras and carries out the image processing task. The server coordinates the motion detection of all the cameras, and performs the transformation operations of the received points from the stations to the maps. The server registers everything in a database, which can then be accessed by different means. A third module visualize on a map. This module can run on some of the computers and even on the server.

The basic stages of this process are as follows:

- A computer 1 establishes a connection with an IP camera and applies an SF algorithm to the images, also indicating its current state to a server S.
- A computer 2 connects to the server with the display functionality, subscribing to the receipt of points for a map that covers a particular satellite region.
- When the device 1 detects movement, it sends to the server S a message containing the identifier of the processed camera and the coordinates (x, y).
- When S receives the message, it verifies if any equipment is subscribed to the messages of the area that watches this camera. If there is at least one client, S looks for the calibration matrix and transforms the pixel (x, y). If it is active, it applies post-processing.

In this way, with this small message scheme, a server will act as a hub and will know at all times which computers are processing video signals and which devices want to receive the transformed coordinates for a map.

3.5 Overlapping of multi-camera and false positives reduce method

The observed areas by the cameras often overlap with each other, so a moving point can be sent by several devices. At the same time, false positives can be generated, resulting from a movement of an object.

To unify these points, in ord be treated as one and reduce false positives, a second processing of the same should be applied.
One of the proposals of this work is to order observations by time intervals. For doing this correctly, all the devices should be synchronized with a unique clock, so the server clock is used. Then, for each set of points projected within a range, they are averaged (an interval was chosen for each second), and then the euclidean distance of the group is calculated with respect to the next interval. When the distance is greater than a certain tolerance, the points are discarded. These thresholds were based on [16] where it is expressed that on average a person walks at 5 kilometers per hour (or equal to 1.38 meters per second).

4. Results and discussions

The tracking of a person includes two stages that are considered critical, as the detection of the object in the image of the camera and the mapping of the same to the plane. For the tests used a virtual PC of 4 cores and 2 GB of RAM with 6 connected cameras, the videos counted with a resolution of 640x480 pixels. The chambers were located as shown in Figure refConfiguration.

Fig. 4: Definition zone of the camera

In this configuration, there were several areas of occlusion where the person could not be observed. Zones of interest were defined for each case to obtain the relevant area of analysis. The points where a person passed are marked in "white".
To perform the tests, a point filtering is applied to each set, where it is grouped per second, by averaging the x and y coordinates. Once the sets containing a single mark per second were defined, the measurement was calculated and the distances obtained were averaged. Filtration was applied as explained in section 3.5. Knowing the scale of the map, maximum movement thresholds were defined in a second between 1.38mts and 11.04mts, which correspond to four times the average speed when walking [16].

On the other hand, in order to compare the efficiency of the detectors, the distance between the calculated points and a reference path was calculated. To measure this distance, we used the indicator eq. 1 which measures as the average of the minimum distances from all points of the original path to the consecutive segments.

$$\sum_{j=1}^{n} \frac{1}{n} (Min_{pb}(Distance(pA, pB(j), pB(j+1)))) \quad (1)$$

where pA $\in$ Initial points ; pB $\in$ Detected points In turn, a gain filter of (1 % to 99 %) was applied to the difference image, prior to applying a binary threshold and obtaining the motion pixels of the object of interest. In these cases, a lower gain would remove pixels from the object and a higher gain would increase the amount. In any case, the geometric center of the object, later used in the projection, was modified. The table was generated in (Figure 5) to analyze these combinations, where the number of points were counted.

DETECTOR	Threshold = 1.38	Th. = 2.76	Th. = 5.52	Th. = 11.04
B.M. con G=1	1,615	1,598	1,729	1,712
B.M. con G=10	1,610	1,497	1,837	1,765
B.M. con G=30	1,448	1,549	1,678	1,702
B.M. con G=99	1,515	1,747	1,690	1,667
T.F. con G=1	1,561	1,435	1,894	2,204
T.F. con G=10	2,087	2,182	2,357	2,243
T.F. con G=30	1,429	1,914	2,666	2,639
T.F. con G=99	1,301	1,582	2,180	2,086

Fig. 5: Continuity (Filtered points)

where the distance of the path obtained by each method from the reference was measured for both detectors
In all cases, the influence of the shadow was observed, displacing the center of the object. It is observed that when the threshold is less than 5 mts, the average distance is smaller (since far false positives are discarded). In other case, false points appear that do not belong to the person and the the average distance increase.
The line resulting before and after the filter were those shown in Figure 5.

Fig. 6: (left) Initial points (rigth) Filtered points

It was observed that a great number of samples were filtered, many coming from false movements; and the observed displacements have been due in large part to the projection of the shadow.

5. Conclusions

In this work, a distributed platform for the management of multiple cameras with support for people tracking was presented. This platform allows the application of different background subtraction techniques for detecting objects. Once an object is detected, a map projection of the points is applied in order to visualize it in a georeferenced space in a more user-friendly way. The platform is configurable, allowing to distribute the processing algorithms, visualization or monitoring in any computer or device.
Some data filtering techniques were also tested, in order to improve precision in movement tracking. In this case, a reference path (either manually or by GPS) can be used and compared to those obtained from the automatic analysis. The first results were not entirely satisfactory, since in an ideal case they should be close to 0. What it was also observed is how shadow affects analysis, generating disturbances in the images.
In a future work, we intend to perfect the subtraction techniques that contemplate the appearance of shadows, in order to obtain more precise measures. Also the idea is to work on tracking many people simultaneously, while integrating with other Deep-Learning object recognition techniques; in order to optimize the objects classification.

References

[1] Watch Big Brother. The price of privacy: How local authorities spent £ 515m on cctv in four years. *A Big Brother Watch report, February*, 2012.

[2] Cejas E. B. and González C. C. Estado de la normativa sobre video vigilancia en argentina y su relación con la protección de datos personales. *44 JAIIO*, 2015.

[3] Legua C. C. Seguimiento automático de objetos en sistemas con múltiples cámaras. 2013.
[4] Bouwmans D., Porikli F., B. Höferlin, and Vacavant A. *Background Modeling and Foreground Detection for Video Surveillance*. CRC Press, 2014.
[5] Hall D., Nascimento J., Ribeiro P., et al. Comparison of target detection algorithms using adaptive background models. In *Visual Surveillance and Performance Evaluation of Tracking , 2nd Joint IEEE Int. Work.*, pages 113–120. IEEE, 2005.
[6] Foresti G.L., Mähönen P., and Regazzoni C. S. *Multimedia video-based surveillance systems: Requirements, Issues and Solutions*, volume 573. Springer Science & Business Media, 2012.
[7] Kessler G. *El sentimiento de inseguridad: sociología del temor al delito*. Siglo Veintiuno Editores, 2009.
[8] Kruegle H. *CCTV Surveillance: Video practices and technology*. Butterworth-Heinemann, 2011.
[9] Shaikh S. H., Saeed K., and Chaki N. *Moving Object Detection Using Background Subtraction*. Springer, 2014.
[10] Dammert L., Salazar F., Montt C., and González P. Crimen e inseguridad: indicadores para las américas. *FLACSO-Chile/Banco Interamericano de Desarrollo (BID)*, 2010.
[11] Piccardi M. Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on*, volume 4, pages 3099–3104. IEEE, 2004.
[12] Shachtman N. Big brother gets a brain. *Village Voice*, 48(28):40, 2003.
[13] Chandon P., Hutchinson J., Bradlow E., and Young S. H. Measuring the value of point-of-purchase marketing with commercial eye-tracking data. *INSEAD Business School Research Paper*, (2007/22), 2006.
[14] Celtic Telecommunication Solutions. Husims- human situation monitoring system, 2012. https://www.celticplus.eu/wp-content/uploads/2014/09/HuSIMS-leaflet_lq.pdf.
[15] Dirección nacional de protección de datos personales-disposición 10/2015. http://www.infoleg.gob.ar/infolegInternet/anexos/240000-244999/243335/norma.htm.
[16] Velocidad promedio del desplazamiento humano. https://es.wikipedia.org/wiki/Kil

ARENA Simulation Model of a Conversational Character's Speech System

YOSELIE ALVARADO[1], CLAUDIA GATICA[2], VERONICA GIL COSTA[2], ROBERTO GUERRERO[1]

1 Laboratorio de Computación Gráfica,
2 Laboratorio de Investigación y Desarrollo en Inteligencia Computacional,
Universidad Nacional de San Luis, Ejército de los Andes 950,
Tel: 02664 420823, San Luis, Argentina
{ymalvarado, crgatica, vgcosta, rag}@unsl.edu.ar

Abstract. Currently, there are many applications like human-computer interactions in which speech technology plays an important role. In particular, embodied conversational character interfaces research has produced widely divergent results and it has tended to focus on the character's dialog capabilities associated with reasoning. Conversely, the present work attempts to evaluate a conversational character from interaction-related aspects. This paper describes an analysis of a conversational character spoken dialogue system using a discrete event simulator. The simulation model was implemented in the ARENA traditional simulator. Simulation results show that sometimes the response time from the speech conversational character can be too long to user, as well as environment noise is an important aspect of the system to be improved.

Keywords: Virtual Reality (VR), Virtual Humans, Conversational Characters, Natural User Interface (NUI), Speech Interface, Conversational Interaction, ARENA.

1. Introduction

Virtual reality technology has become a very popular technology, which embodies the newest research achievements in the fields of computer technology, computer graphics, sensor technology, ergonomics and human-computer interaction theory. Virtual reality and interactive technology have attracted a great deal of attention and much active research is currently being carried out in an effort to investigate their possible benefits in several areas. It is not a secret: virtual reality technology is an important technology to be paid attention, which will bring huge impact to our life and work [1-4].
New developments in the fields of speech recognition, natural language processing, and computer graphics have given rise to the emergence of more sophisticated computer interfaces with multimodal interaction [5].
On one hand, the evolution in speech technologies allow the development of new systems with several purposes, including voice search, personal digital

assistant, gaming, living room interaction systems, and in-vehicle infotainment systems as the most popular applications in this category [6].

On the other hand, embodied conversational character emerged as an specific type of multimodal interface, where the system is represented as a person conveying information to human users via multiple modalities such as voice and hand gestures, and the internal representation is modality-independent, both propositional and non-propositional. Embodied conversational character answers questions and performs tasks through interaction in natural language-style dialogs with users contrasting the traditional view of computers. Many people believe that such interfaces have great potential to be beneficial in human-computer interaction for a number of reasons. Conversational character could act as smart assistants, much like travel agents or investment advisors [7-10].

Conversational character applies rich style of communication that characterizes a human conversation. Researchers have built embodied multimodal interfaces that add dialogue and discourse knowledge to produce more natural conversational characters. For example, *Peedy the parrot* is an embodied character that allows users to verbally command it to play different music tracks. While most research on embodied conversational characters has concentrated on the graphical representation and conversational capabilities of the virtual character, others investigated the question of whether auditory embodiment can provide cues that influence user behaviors and ultimately affects the learning performance of users interacting with a virtual character to learn about biology, mathematics, geography, literature, etc. [11-14].

Over last decade, virtual character has been increasingly used for educational purposes. The rationale behind this emerging trend is the belief that information technology can be utilized as a powerful means to assist learners with the acquisition of general knowledge, literacy, narrative competence, social skills like teamwork and negotiation capabilities, logical and spatial reasoning, eye-hand coordination and fine motor control, among others [15-17].

Conversational character must allow the user (speaker) to watch for feedback and turn requests, while the character (listener) can send these at any time through various modalities. The interface should be flexible enough to track these different threads of communication in the appropriate way to each thread. Different threads have different response time requirements; some, such as feedback and interruption occur on a sub-second timescale [18].

It is evident that human-computer communication is extremely complex just as human-human communication. Its analysis implies to turn to several disciplines like as psychology, sociology, biology, among others for foundation of computational model. These disciplines provide very precious qualitative and quantitative information that are indispensable to consider. Moreover, the evaluation of single modalities from conversational character often can be useful for system performance analysis. In this case, simulation models are a powerful tool for both understand the system behavior and to detect possible bottlenecks [19-22].

Thus, as the conversational interaction is a vital aspect for an embodied conversational character, in the present work we present a simulation approach as a real-time modeling for the speech modality.

2. Conversational Character System

The spoken dialogue subsystem of a semantic conversational character is the system simulated. The *character* feature means that system's interface has a visual embodiment and the *conversational* feature means that the system simulates a conversational skill like a human being so the system shows listening and speaking ability. Finally, the conversational character is *semantic* because it is able to perform reasoning through a query to a *Dbpedia*'s query module for users' answers [23, 24].
More specifically, the implemented approach is a question-answering virtual character driving responses to user's inquiries (See fig. 1). The system, named CAVE-VOX, is thought to be used in inquiry applications through the analysis of the user's keyword natural language utterance and the generation of the appropriate response; previous works describe the whole system [25, 26].

Fig. 1. CAVE-VOX system.

System's human-computer interaction involves real-time synchronization of several aspects like visualization, speech recognition, keywords detection, speech synthesis, among others. Thus a dialogue system between user and virtual character that exploits knowledge provided by structured data (ontologies) in order to help users in specific information search was built.

Then, this question-answer character has a collection of responses relevant to a particular topic. The driven query must be specific query about a previous chosen topic. The character plays an appropriate role according with the selected topic and answers to user's queries through data stored in a *Dbpedia* data base.

2.1 Speech system

In this work, the spoken dialogue subsystem of a semantic conversational character was simulated. As the real-time interaction is important for this analysis the conversational module was only considered, this system's modality involves simulating inherent human skills like as listening and talks. The interaction between the character and the user is performed in real-time through user's voice (by speech to text conversion) and virtual character's voice (by text to speech translation), adding simple gestures and expressions, and lip synchronization.
In order to make the evaluation and considering the original system, the main involved components are:

- *Speech recognition (listening)*: while interacting, a human user talks to the system, which transforms the user's speech to a textual representation by using an **Automatic Speech Recognition** (ASR) module. The module takes as input the user's speech utterance that comes from microphone and gives a resultant text from parsing the word string produced by speech recognition and forms an internal semantic representation based in keywords contained into a grammar.
  Each word said by the user is analyzed for the recognizer and compared through a phonetic transcriber. All the picked up words by the microphone are contrasted with words defined by rules in a grammar. According with the grammar, possible sentences are analyzed and compared with the received sentence, then every word gets a confidence value. The ASR module takes this confidence value to determine if a word is accepted or rejected.
  If the word is accepted, then it is stored as a valid keyword. If the word is rejected, then the system must inform about the failure to user. From previous analysis with users, it is known that the rejection rate is approximately 50%.
  When all the words are accepted it allows for a query as a text string consisting of keywords.
  Because the process performed by the ASR is completely computational and is performed in parallel, it is not possible for the user to know the time required executing this operation and it is usually represented in nanoseconds. Therefore, for the simulations performed in this work, we consider that the ASR processing time is 0 seconds.

- *Reasoning (thinking):* this module generates a response based on the input, the current state of the conversation and the dialog history. For this, it extracts the meaning of the utterance from keywords, manages the dialog flow and produces the appropriate actions for the target domain in on-line *Dbpedia*.
  This approach gives complete control over the virtual persona's knowledge and expressions to the scriptwriter who creates the responses. It allows the writer to specify the character of the virtual persona, what information it can deliver and the form of that delivery. When an interactor comes up to the conversational character and asks it a question, the system driving the character analyzes the interactor's question and selects the appropriate response from *Dbpedia* collection.
  Finding an answer takes between 10 and 120 seconds according with the user's query. If this process faults, then the system notifies to the user about the fault. From previous analysis, it is known that the fault rate is approximately 10%.

- *Speech synthesis (talking):* It is a **Text-To-Speech** (TTS) module carrying out the generation of the synthetic output voice from the text that comes as a response from the Reasoning Module. This module is performed on three main situations: if one word is rejected, then the conversational character says "*Excuse me, I didn't hear you very well*" which takes about 3 seconds, if the *Dbpedia* query faults, then character says "*Sorry. It's not possible to retrieval the elements you requested*" which takes about 7 seconds, and if the conversational character has an answer from *Dbpedia*, then the delay is between 40 and 300 seconds according with its length.

3. Modeling and Simulation with ARENA

Modeling and simulation provide the basis for efficient solving of various problems related to the operation of complex systems like analysis, optimization and management, industry problems (like mining process, etc.). Simulation is considered to be one of the most effective technologies for the analysis and planning of logistics systems [27].
ARENA simulation software is a general propose simulation tool enabling the construction of models over a series of modules or basic components organized hierarchically. The ARENA simulation software has high level of modeling supporting graphical design. It also includes a lower level of modeling including specific details as arrival times, service time, scheduling of processes, etc. [28].
A model is developed using modules that are part of the basic processes. In ARENA, modules are the flowchart and data objects that define the process to be simulated. All information required to simulate a process is stored in modules. The dynamics associated with the processes can be viewed as nodes

in a network by which entities circulate causing a change in the system state. The entities with attributes and variables compete for the services provided by the resources. Entities are items (like trucks, mineral, etc.) that are being served or produced [29].
Figure 2 shows a simple model build with ARENA. The CREATE module is the starting point for entities in a simulation model. Entities are created using a schedule or based on a time between arrivals. Entities then leave the module to begin processing through the system. The entity type is specified in this module. The PROCESS modules are intended as the main processing method in the simulation. They include the resource by which entities compete. A resource retained by an entity must be released at some point in the model. Otherwise, a deadlock can occur. The DECIDE module allows for decision-making processes in the system. Finally, the DISPOSE modules are ending point for entities in a simulation model [30].

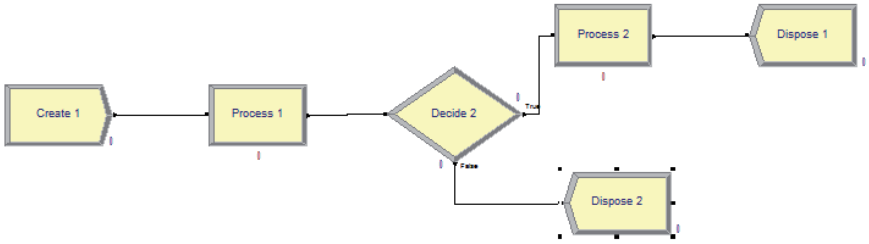


Fig. 2. Basic modules used in ARENA.

After the model is built, we can run simulations to obtain different metrics and statistics like resources utilization, waiting time, etc.

4. Conversational Character Modeling

The simulation model covers a subset of a conversational character. ARENA v.10 was used for modeling the real-time simulation of the spoken dialogue system. As ARENA is an entity-driven application, only the tasks that directly impact the entities are modeled. In the model, an entity represents a query made by the user.

Our resulting model contains the three basic components of a conversational system:

- *Speech recognition (listening):* this system's stage modeling uses a CREATE module which creates entities (queries based on a time between arrivals), a PROCESS module intended to represent the process performed by ASR resource (because it is not possible for

the user to know the time required to execute this operation then the simulated time-delay is zero), and a DECIDE module representing the process which decide as to whether or not something heard by the recognizer is a valid word (according with this problem the success rate is approximately 50%).

- *Reasoning (thinking*): this component is modeled with a PROCESS module representing keyword searching in the *Dbpedia* resource (with a uniform delay between 10 and 120 seconds), and a DECIDE module determining as to whether or not a keyword is in *Dbpedia* (according with this problem the success rate is approximately 90%).

- *Speech synthesis (talking):* it is represented by means of three different PROCESS modules according with the three main established situations (3 seconds, 7 seconds and a uniform delay between 40 and 300 seconds respectively). All of these processes use TTS resource and all of them complete at DISPOSE module.

Considering the system's components described before, our simulation model is showed in figure 3.

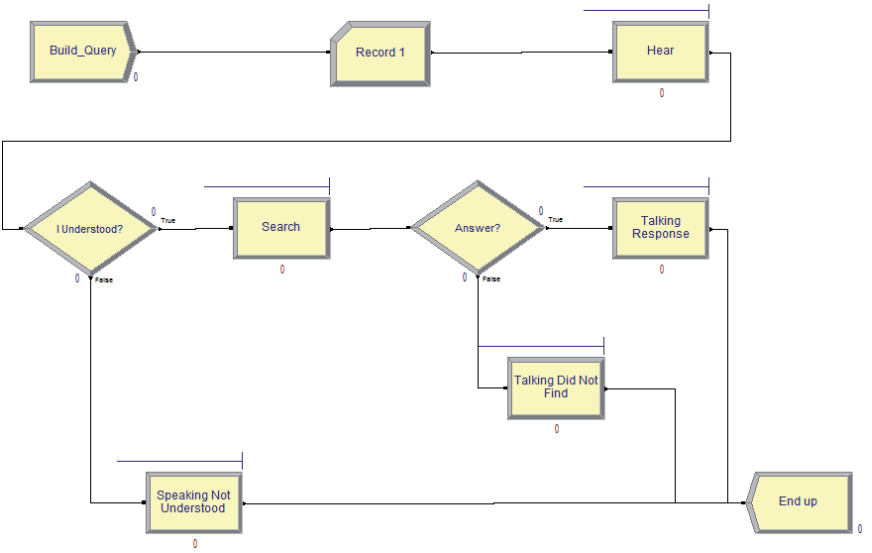


Fig. 3. Conversational Character modeling in ARENA.

5. System Evaluation

Once the simulation model of the conversational character has been built and verified, it can be used to analyze the system performance. After the

simulation's running, the ARENA simulator recorded several data related to the model parameters. ARENA reports include category by replication, entity times, entity number, queue times, resource usage, among others.
In this case, 20 replications were performed. From every replication, information about entities times and resources usage was collected: entity parameter represents the query from user and resources parameters are ASR, TTS and *Dbpedia* tools.
The simulation model generates outputs including the performance measure used in the users' experiments considering the process realized by the conversational character from the user's utterance to the appropriate response synthesized by the subsystem for a particular situation.
Query is the unique entity in the simulated system. In this system the query analysis is vital because a query time allows learning about response times of the system. The analyzed time for the entities (query) is from the query creation to the query output. An average time from all repeats was considered to obtain minimum, maximum and average times of entity. Table 1 shows the resulting values in seconds.

	Minimum	Maximum	Average
Value	3,0	359,2	103,4

Table 1. Times entity: Query.

From these values, it is concluded that the processing time necessary to attend the demand of a user is 103 seconds, meaning near 2 minutes in average, and the maximum time is 359 seconds meaning near 6 minutes. Minimum value represents the case when speech recognition fails.
In analysis of resources, resource usage was considered. ASR and TTS resources are used in each query only once, then their constant values are not important for system's performance analysis. This is different from *Dbpedia* resource where its analysis requires to be considered. According with ARENA outputs, *Dbpedia* usage is about 30% as a minimum, 80% as a maximum and 42,3% in average. It is important to emphasize that if *Dbpedia* resource is not used this is due to speech recognition failure.

6. Conclusions and Discussion

Latest research had been oriented to create embodied computer-animated characters that produce accurate auditory and visible speech, as well as realistic facial expressions, emotions, and gestures. The invention of such characters has a tremendous potential to benefit virtually all individuals in natural user interface. Conversational interaction is an important aspect to enable and improve the human-computer interaction.

This paper involved the evaluation of the spoken dialogue system of a conversational character with a discrete event simulator ARENA.
We described character's speech system and its simulation model that was developed to study conversational interaction performance. We introduced the general and most important issues one has to take care of when starting to simulate with simulator ARENA. We discussed the modules and parameters in detail with the main idea to obtain quantitative results from simulation.
At the moment, analysis is limited to entities and resources parameters. Even though, the analysis does not cover all the aspects of the user query, we believe it has several elements to improve. For example, from entity point of view the simulation results show that sometimes the extensiveness of the speech from the conversational character can be too long to users. A more exhaustive study is need for determining if a long answer is produced by reasoning or something else.
For the analysis of resources, *Dbpedia* usage is not ideal because sometimes speech recognition fails due to many factors influence in a correct recognition of speech as: speaker's clarity, microphone's quality, user's accent, environment noise, etc. Several of these aspects can be improved by upgrading microphone's quality.

In this work, we focused on human-computer conversational interaction in real-time and how to do it more natural and intuitive to users allowing proficient user-interactivity in real-time, meaningful feedback and learning through an interface.

References

1. XinXing Tang, editor. *Virtual Reality - Human Computer Interaction*. InTech, 2012.
2. M. Chan. *Virtual Reality: Representations in Contemporary Media*. Bloomsbury Publishing, 2014.
3. T. Parisi. *Learning Virtual Reality: Developing Immersive Experiences and Applications for Desktop, Web, and Mobile*. O'Reilly Media, Incorporated, 2015.
4. Kurt Squire. Changing the game: What happens when video games enter the classroom? *Innovate: Journal of Online Education*, 1(6), August 2005.
5. Cecilia Sik Lanyi, editor. *The Thousand Faces of Virtual Reality*. InTech, 2014.
6. Dong Yu and Li Deng. *Automatic Speech Recognition: A Deep Learning Approach*. Springer Publishing Company, Incorporated, 2014.
7. J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. Embodied conversational agents. chapter Human Conversation As a System Framework: Designing Embodied Conversational Agents, pages 29-63. MIT Press, Cambridge, MA, USA, 2000.
8. M. Mancini. *Multimodal Distinctive Behavior for Expressive Embodied Conversational Agents*. Universal Publishers, 2008.
9. Q. Chen, P. Torroni, S. Villata, J. Hsu, and A. Omicini. *PRIMA 2015: Principles and Practice of Multi-Agent Systems: 18th International Conference,* Bertinoro, Italy, October 26-30, 2015, Proceedings. Lecture Notes in Computer Science. Springer International Publishing, 2015.

10. B. Endrass. *Cultural Diversity for Virtual Characters: Investigating Behavioral Aspects across Cultures*. EBL-Schweitzer. Springer Fachmedien Wiesbaden, 2014.
11. Niels Ole Bernsen and Laila Dybkjr. *Multimodal Usability*. Springer Publishing Company, Incorporated, 1st edition, 2009.
12. Gene Ball, Dan Ling, David Kurlander, John Miller, David Pugh, Tim Skelly, Andy Stankosky, David Thiel, Maarten V Dantzich, and Trace Wax. Life-like computer characters: The persona project at microsoft research. *Software agents*, pages 191-222, 1997.
13. Sharon Oviatt, Courtney Darves, and Rachel Coulston. Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Trans. Comput.-Hum. Interact.*, 11(3):300-328, September 2004.
14. R.J.S. Sloan. *Virtual Character Design for Games and Interactive Media*. CRC Press, 2015.
15. Andrea Corradini, Klaus Robering, and Manish Mehta. *Conversational characters that support interactive play and learning for children*. INTECH Open Access Publisher, 2009.
16. David Griol, José M Molina, Zoraida Callejas, and Ramón López-Cózar. La plataforma educagent: agentes conversacionales inteligentes y entornos virtuales aplicados a la docencia. 2011.
17. Marissa Milne, Martin Luerssen, Trent Lewis, Richard Leibbrandt, and David Powers. Embodied conversational agents for education in autism. *A comprehensive Book on Autism Spectrum Disorders*, page 387, 2011.
18. Dominic W Massaro, Ying Liu, Trevor H Chen, and Charles Perfetti. A multilingual embodied conversational agent for tutoring speech and language learning. In INTERSPEECH, 2006.
19. J.O. Turner, M. Nixon, U. Bernardet, and S. DiPaola. *Integrating Cognitive Architectures into Virtual Character Design*. Advances in Computational Intelligence and Robotics. IGI Global, 2016.
20. Y. Zhang, P. Zhang, and D.F. Galletta. *Human-computer Interaction and Management Information Systems: Foundations*. Taylor & Francis, 2015.
21. Benjamin Weiss, Ina Wechsung, Christine Kühnel, and Sebastian Möller. Evaluating embodied conversational agents in multimodal interfaces. *Computational Cognitive Science*, 1(1):1, 2015.
22. Lee W Lacy. *Interchanging Discrete event simulation Process Interaction Models using the Web Ontology Language-OWL*. PhD thesis, University of Central Florida Orlando, Florida, 2006.
23. Yu-Lin Chu and Tsai-Yen Li. *Realizing semantic virtual environments with ontology and pluggable procedures.* INTECH Open Access Publisher, 2012.
24. Mohamed Morsey, Jens Lehmann, Sören Auer, Claus Stadler, and Sebastian Hellmann. DBpedia and the Live Extraction of Structured Data from Wikipedia. *Program: electronic library and information systems*, 46:27, 2012.
25. Y. Alvarado, N. Moyano, D. Quiroga, J. Fernández, and R. Guerrero. *Augmented Virtual Realities for Social Developments. Experiences between Europe and Latin America*, chapter A Virtual Reality Computing Platform for Real Time 3D Visualization, pages 214-231. Universidad de Belgrano, 2014.
26. N. Jofré, G. Rodríguez, Y. Alvarado, J. Fernández, and R. Guerrero. Virtual humans: Conversational characters for a cave-like environment. In *XX Congreso argentino de ciencias de la computación*, pages 937-946. Universidad Nacional de La Matanza, Octubre 2014.
27. Mustafa Rawat and Steven Vaccaro. Implementing a distributed logistics simulation using arena and hla. In *Proceedings of the 2009 Spring Simulation Multiconference*, page 62. Society for Computer Simulation International, 2009.

28. T. Altiok and B. Melamed. *Simulation Modeling and Analysis with ARENA*. Elsevier Science, 2010.
29. M.D. Rossetti. *Simulation Modeling and Arena*. Wiley Series in Modeling and Simulation Series. Wiley, 2015.
30. W.D. Kelton, R.P. Sadowski, and N.B. Swets. *Simulation with Arena*. McGraw-Hill Education, 2015.

XIII

Software Engineering Workshop

An HCI quality attributes taxonomy for an impact analysis to interactive systems design and improvement

FERNANDO PINCIROLI

Research Institute, Faculty of Informatics and Design, Champagnat University.
Godoy Cruz, Argentina
pincirolifernando@uch.edu.ar

Abstract. In the interaction between users and systems, software quality attributes are mainly involved. When designing interfaces for human-computer interaction different alternatives can be considered in order to obtain the highest quality in an interactive system. However, quality attributes have positive and negative contribution relationships among each other, so that a change in one of them can cause a higher improvement than expected or an unwanted degradation of the system. This is the reason why in this paper we propose a taxonomy of non-functional requirements that can be assigned quality properties susceptible to be measured to propose alternatives that achieve a better quality for the systems. Quality that can be obtained by taking into account the contribution relationships among quality attributes, in order to select those alternatives that provide the biggest gain of system quality for the design and improvement of systems and software interfaces.

Keywords: HCI, quality attributes, quality attributes taxonomy, contribution relationships among quality attributes, quality metrics.

1. Introduction

The design and improvement of human-computer interaction (HCI) is a delicate task and requires a lot of effort. In both cases, for the design and the improvement, alternative solutions are analyzed and one of them is chosen to be finally implemented. The overall quality of the interface will have a value that can be measured by different techniques offered in the literature. However, the greatest difficulty is not in the measurement techniques, but on what should be measured, and we believe that this is due to two main factors: the lack of a taxonomy for quality attributes involved in HCI and the lack of analysis of the positive and negative contribution relationships existing among them.

In this work, we offer a taxonomy of quality attributes for HCI and a perspective of how to analyze the impact of the positive and negative contribution relationships among quality attributes that could enhance an

alternative solution or could prevent the degradation in the overall quality of an interactive system after the efforts to improve it.
In section 2, we will present how to measure the quality properties of a set of non-functional requirements that are structured as a taxonomic classification. In section 3, we will review how to analyze the positive and negative contributions that occur among non-functional requirements. Finally, in section 4 we will offer a taxonomy of requirements for the design and improvement of interactive systems that can be used to determine its quality measurements and to analyze the cross-impact among them.

2. Taxonomy of quality attributes to measure software systems quality

Taxonomy is the "science that deals with the principles, methods and purposes of classification"[1], normally used to present information classified in a hierarchical and systematic way. Villegas et al. [1] present a list of the main taxonomies used in the disciplines of Software Engineering and HCI. Other authors offer a good overview of the major classifications over time of quality attributes, also called non-functional requirements [2] [3]. In the particular case of quality attributes for quality in use of software products, which are of particular interest in the area of HCI, one of the most widespread classifications is the proposal by ISO/IEC 25010 [4] standard:

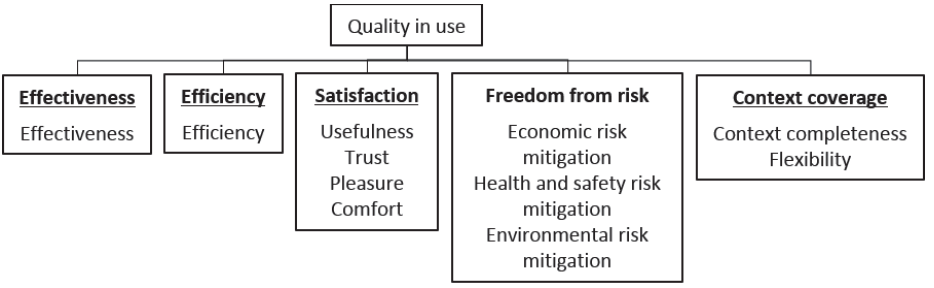


Fig. 1. Quality in use model in the ISO/IEC 25010 standard.

This model depicts a first level of five characteristics and a second level of its main subcharacteristics, which is usually extended with a new level of quality properties, which are the attributes that quality measurements are assigned to when the quality of a software product is objectively measured.

[1] Translated from the Real Academia Espanola's definition.

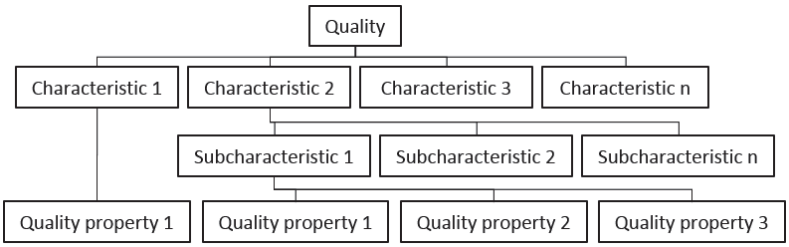


Fig. 2. Structure used to measure the quality.

Mairiza et al. [5] offer a broad classification of quality attributes with the same hierarchical structure, in which they call "non-functional requirement", "definition" and "attribute" to ISO's "characteristic", "subcharacteristic" and "quality property" respectively.

Taking all things considered, a breakdown structure of quality attributes is performed, where the last level has detailed the properties that will be measured to determine the level of the quality attribute and therefore go on aggregating the results in the higher nodes up to the root node, wherein the overall quality of the system is obtained.

There are different techniques for measuring the properties of the quality attributes of a system that can provide a quantitative assessment for each of them. One of these proposals is GOCAME [6], which stands for Goal-Oriented Context-Aware Measurement and Evaluation. As it can be seen in figure 3, GOCAME is applied to measure the quality of two consecutive versions of a software product. There, the values obtained for the properties of subcharacteristics are combined, then these results are combined among them to get the quality of each characteristic, and finally the same thing is repeated to obtain the overall quality of the software product.

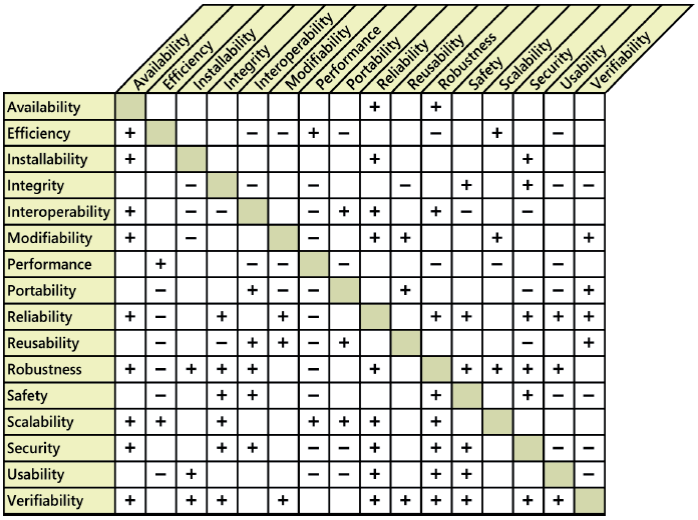
	JIRA v.1		JIRA v.1.1	
Characterisitcs and Attributes	EI	P/G I	EI	P/G I
1. Actual Usability		53.3%		67.0%
1.1. Effectiveness		73.2%		86.7%
1.1.1. *Sub-Task Correctness*	86.4%		91.9%	
1.1.2. *Sub-Task Completeness*	87.9%		95.5%	
1.1.3. *Task Successfulness*	45.5%		72.7%	
1.2. Efficiency		29.3%		42.8%
1.2.1. *Sub-Task Correctness Efficiency*	37.4%		44.3%	
1.2.2. *Sub-Task Completeness Efficiency*	37.5%		47.3%	
1.2.3. *Task Successfulness Efficiency*	13.1%		36.8%	
1.3. Learnability in use		57.3%		71.6%
1.3.1. *Sub-Task Correctness Learnability*	78.8%		75.1%	
1.3.2. *Sub-Task Completeness Learnability*	26.4%		77.3%	
1.3.3. *Task Successfulness Learnability*	66.7%		62.5%	

Fig. 3. Measurement of Jira's usability in versions 1.0 and 1.1 (taken from [6]).

However, as shown in figure 3, although the usability of the system improved from 53.3% to 67.0% of a Jira's[2] version to the following one, some subcharacteristics suffered degradation, such as those related to system learnability. Even though it is difficult to ensure, we could think that the increment of the functionality from one version to the next one could have resulted in a greater difficulty to learn them. This is well known as contribution relationships among quality attributes contribution, which can be positive or negative, occurring the latter in the mentioned example.

3. Contribution relationships among quality attributes

Quality attributes have positive and negative contribution relationships among each other. This fact, taken into account by the lead authors on requirements, is of enormous importance in establishing solutions to improve the quality of software products [7] [8] [9] [10] [11] [12].



	Availability	Efficiency	Installability	Integrity	Interoperability	Modifiability	Performance	Portability	Reliability	Reusability	Robustness	Safety	Scalability	Security	Usability	Verifiability
Availability									+		+					
Efficiency	+				−	−	+	−			−		+		−	
Installability	+								+					+		
Integrity			−		−		−			−		+		+	−	−
Interoperability	+		−	−			−	+	+		+	−		−		
Modifiability	+		−				−		+	+			+			+
Performance		+			−	−		−			−		−		−	
Portability		−			+	−	−			+				−	−	+
Reliability	+	−		+		+	−				+	+		+	+	+
Reusability		−		−	+	+	−	+						−		+
Robustness	+	−	+	+	+		−		+			+	+	+	+	
Safety		−		+	+		−				+			+	−	−
Scalability	+	+		+			+	+	+		+					
Security	+			+	+		−	−	+		+	+			−	−
Usability		−	+				−	−	+		+	+				−
Verifiability	+		+	+		+			+	+	+	+		+	+	

Fig. 4. Contribution relationships among quality attributes (taken from [7]).

When a need for improvement is detected in any of the quality attributes, it is a mistake to seek improvement solutions only for one attribute regardless of its relationship with the remaining ones. This is due to the fact that the solution implemented to improve one attribute could negatively affect other attributes and degrade them in such a way that a lower overall quality of the product would be obtained.

[2] Jira is one of the most popular applications in the world for collaborative work, developed by the company Atlassian; http://www.atlassian.com

Thus, when we look for an improvement in the quality of a software product, the combination of positive and negative contributions of the different solutions must be measured in order to choose the one offering the greatest gain in overall quality. We have recently published an article presenting a technique to analyze and to measure the impact of solutions on the overall software quality, to guide the selection of the best alternative in this regard [13].
Our goal is to develop a matrix of contribution relationships to the quality attributes involved in interactive systems, but we found some difficulties on this that we will try to resolve.

4. Taxonomy of quality attributes for HCI

It is difficult to set a taxonomy for quality attributes related to HCI. Different authors, many of them recognized in the field, have different opinions and sometimes even contradictory. One of the difficulties is to set a classification of quality attributes. In the research conducted by Mairiza et al. [5] along the existing literature on quality attributes, they obtained an exhaustive list of them. But at the same time, they found that some authors present as quality attributes what other authors consider properties of quality attributes.
Another difficulty is the diversity of criteria in establishing the quality attributes that influence the acceptability of interactive systems. For example, the ISO/IEC 25010 standard [4] details the quality attributes and their respective subcharacteristic to a system's quality in use. It also indicates which quality properties of software products and computer systems influence on quality in use for primary users of the system. As shown in figure 5, ISO considers that *"functional suitability"*, *"performance efficiency"*, *"usability"*, *"reliability"* and *"security"* of the product have influence on quality in use, while according to that standard *"compatibility"* (besides *"maintainability"* and *"portability"*) does not.

Software product properties	Computer system properties	Product quality characteristic	Influence on quality in use for primary users	Influence on quality in use for maintenance tasks	Information system quality concerns of other stakeholders
↳	↳	Functional suitability	*		
↳	↳	Performance efficiency	*		*
↳	↳	Compatibility		*	
↳	↳	Usability	*		
↳	↳	Reliability	*		*
↳	↳	Security	*		*
↳	↳	Maintainability		*	
↳	↳	Portability		*	

Fig. 5. Product and system quality that influence on quality in use (taken from [4]).

However, if we look at the known Nielsen's classification of acceptability [14], we will find a different opinion.

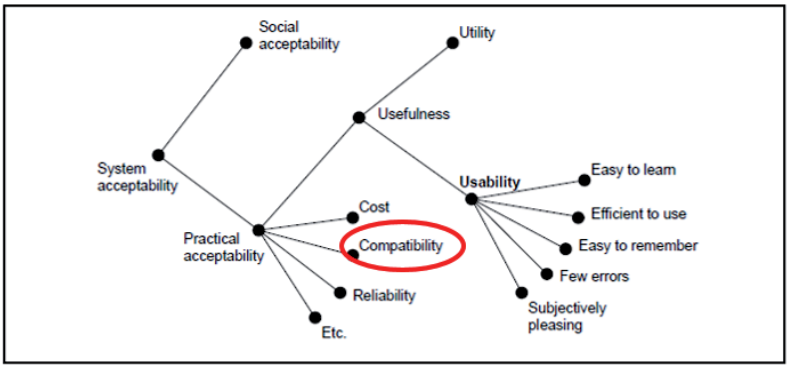


Fig. 6. Nielsen's system acceptability model.

In the classification presented in figure 6, it can be seen that *"compatibility"* is part of the system acceptability, contrary what is indicated above for quality in use by ISO 25010 standard. For all the aspects previously considered is that we dare to propose a taxonomy of quality attributes for this area. To develop this taxonomy we start from Nielsen's model due to its widespread use and acceptance, but we extend it with a new level of quality subcharacteristics obtained from the following sources:

1. Compilations of the most comprehensive lists of quality attributes presented by different authors, from which we have taken only those attributes that influence the HCI [2] [3] [5] [15].
2. Within the Nielsen's category of "Few errors" we incorporate the *"antifragility"* quality attribute which is a new concept that considers that the systems must not only be robust, but even strengthened and improved from the impacts they receive [16] [17] [18].
3. We incorporate other HCI specific quality attributes from the material of the PhD course "Design of interactive systems from a user-centered approach", taught by Dr. César Collazos [19].

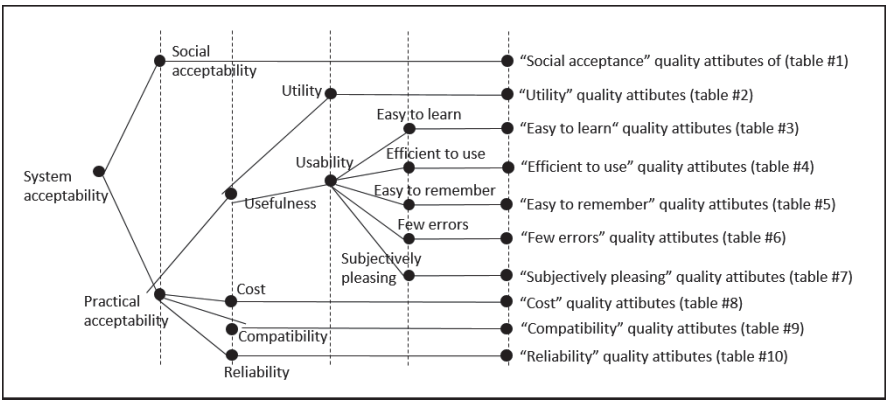


Fig. 7. Extended Nielsen's acceptability model.

We have extended the Nielsen's model with a new level of quality attributes, as shown in figure 7. The quality attributes of the new level mentioned in the figure above are detailed in the following explanatory tables.

Table 1. "Social acceptability" quality attributes.

Attributes	
Diffusion	Status

Table 2. "Utility" quality attributes.

Attributes	
Completeness	Functionality
Comprehensiveness	Maturity
Coverage	Service quality
Effectiveness	Suitability

Table 3. "Easy to learn" quality attributes.

Attributes	
Affordability	Natural relationship between controls and their functions
Auto-description	Observability
Available helps	Perception
Communicativeness	Predictability
Comprehensibility	Presentability
Conciseness	Readability
Consistency	Simple design
Content and interaction metaphor	Structuredness
Cultural level	Trainability
Decreased cognitive load	Universality
Degree of technology knowledge	Use of analogies
Easy to use	Use of icons

Expressiveness	Use of standards
Language and communication	Visibility

Table 4. "Easy to use" quality attributes.

Attributes	
Adaptation to different types of environments	Latency
Agility	Layout and organization of controls
Alternatives of use	Layout of the most important information
Availability	Manageability
Comparability	Multiuser architecture
Complexity of interaction	Navigation among windows
Configurability	Navigation inside the windows
Conformance	Operability
Controllability	Physical size of the equipment
Customizability	Reconfigurability
Design for users with cognitive decreases	Remote operation
Design for users with hearing decreases	Repeatability
Design for users with motion decreases	Replicability
Design for users with visual decreases	Response time
Dialogue techniques	Simplicity
Flexibility	Use of system resources
Handling of attention	Use of user resources
Installability	WYSIWYG

Table 5. "Easy to remember" quality attributes.

Attributes	
Experience gain	Mechanisms to help remember where you are
Mechanisms to help remember actions	Memory of use

Table 6. "Few errors" quality attributes.

Attributes	
Access control	Possibility of correcting an error
Accuracy	Possibility of reversing an error
Antifragility	Presentation of correct messages
Correctness	Recoverability
Demonstrability	Restriction indication
Distinction of colors	Stability
Error highlighting	Traceability
Error protection	Uniformity
Fault tolerance	Use of codes besides colors
Feedback of results of actions	Verifiability
Feedback of the actions taken	Visibility of system status
Immunity	Visual structure of information
Integrity	

Table 7. "Subjectively pleasing" quality attributes.

Attributes	
Actualization of technology	Esthetic
Appropriate combination of colors	Fatigue and health
Attractiveness	Formality
Comfort	Gamification
Emotionality	Motivation
Ergonomics	Social context

Table 8. "Cost" quality attributes.

Attributes	
Acquisition cost	Training cost
Cost of consumables	Update cost
Maintenance cost	Upgrade cost

Table 9. "Compatibility" quality attributes.

Attributes	
Coexistence	Mobility
Standardizability	Portability
Generality	Replaceability
Integratability	Transferability
Interoperability	

Table 10. "Reliability" quality attributes.

Attributes	
Anonymity	Privacy
Auditability	Protection
Certainty	Responsibility
Confidentiality	Security
Dependability	Supportability
Non-repudiation	Trustability

Quality attributes presented in the tables above are, then, the basis of quality measurement of interactive systems and of confrontation of their respective contribution relationships in order to determine the best alternative for the design and improvement of these systems.

5. Conclusions and future work

We consider this work as a starting point for the design and improvement of HCI from a taxonomy of the involved quality attributes (tables 1 to 10) that can have applied measurement techniques to obtain a comparable measure of the resulting overall system quality (figure 3) and considering the positive

and negative contribution relationships among quality attributes (figure 4) in order to select the design or improvement alternatives providing the best measure of the system quality.
To achieve this there is still much work to be done. The first is to achieve an adequate taxonomy for HCI quality attributes. Different taxonomies and classifications of quality attributes that can be found in the literature are not uniform. In many cases the quality attributes are presented clearly, with sufficient definitions, but other times it is difficult to ensure the concept that aims to convey the author. Moreover, these taxonomies can present different quality attributes groupings, where sometimes some of them are grouped into each other and in other opportunities they are presented at the same level. This is why we presented our proposal in figure 7.
The second step consists on proposing a matrix of contribution relationships among quality attributes brought from taxonomy obtained in the previous point.
Finally, the definition of the form of measurement and the metrics associated with each quality attribute as the proposal presented in section 2 of this work [6] is pending.

References

1. Villegas, M. et al. "Activity theory as a framework for accommodating cultural factors in HCI studies," IEEE Lat. Am. Trans., vol. 14, no. 2, pp. 844–857, 2016.
2. Afreen, N. et al. "A Taxonomy of Software's Non-functional Requirements". Proceedings of the 2nd Intl. Conference on Computer and Communication Technologies, 2015, pp. 47–54.
3. Odeh, Y. and M. Odeh, "A New Classification of Non-Functional Requirements for Service-Oriented Software Enineering," Nauss.Edu.Sa, pp. 1–7, 2009.
4. ISO/IEC, "Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models", vol. 2011, 2011.
5. Mairiza, D. et al. "An Investigation into the Notion of Non-Functional Requirements". Proceedings of the ACM Symp. on Applied Computing, 2010, pp. 311–317.
6. Olsina, L., P. Lew, A. Dieser, and B. Rivera, "Using web quality models and a strategy for purpose-oriented evaluations," J. Web Eng., vol. 10, no. 4, pp. 316–352, 2011.
7. Wiegers, K. and J. Beatty, Software Requirements. Microsoft Press, 2013.
8. Chung, L., B. Nixon, and E. Yu, "Using Non-Functional Requirements to Systematically Select Among Alternatives in Architectural Design". 1st Intl. Work. Archit. Softw. Syst. - Coop. with 17th Int. Conf. Softw. Eng. ICSE 1995, pp. 31–43, 1995.
9. Chung, L., B. Nixon, E. Yu, and J. Mylopoulos, "The NFR Framework in Action," Non-Functional Requirements, pp. 15–45, 2000.
10. Mairiza, D. and D. Zowghi, "Constructing a Catalogue of Conflicts among Non-functional Requirements," Commun. Comput. Inf. Sci., vol. 230, pp. 31–44, 2011.
11. Barbacci, M., et al. "Quality Attributes" CMU/SEI-95-TR-021. 1995.
12. Hines, M. and A. Goerner, "Software quality: attributes and modalities". Trans. Inf. Commun. Technol., vol. 11, 1995.

13. Pinciroli, F. “Improving software applications quality by considering the contribution relationship among quality attributes”. 3rd Int. Work. Comput. Antifragility Antifragile Eng. (ANTIFRAGILE 2016). Elsevier, Procedia Computer Science, vol. 83, pp. 970-975, 2016.
14. Nielsen, J. Usability engineering. San Francisco: Morgan Kaufmann Publishers Inc., 1993.
15. Masip, L. et al. “User experience specification through quality attributes,” Lect. Notes Comput. Sci., vol. 6949 LNCS, no. PART 4, pp. 656–660, 2011.
16. Taleb, N. Antifragile. Things that gain from disorder. New York: Random House, 2012.
17. De Florio, V. “Antifragility=Elasticity+Resilience+Machine learning: Models and algorithms for open system fidelity,” Procedia Comput. Sci., vol. 32, pp. 834–841, 2014.
18. Jones, K. “Engineering antifragile systems: A change in design philosophy,” Procedia Comput. Sci., vol. 32, pp. 870–875, 2014.
19. Collazos, C. “Diseño de sistemas interactivos desde un enfoque centrado en el usuario”. Material of the PhD on Computer Sciences course, Univ. Nacional de San Juan, May 2016.

Quality Evaluation in Agile Process: A First Approach

NOELIA PINTO[1], CÉSAR J. ACUÑA[1], LILIANA CUENCA PLETSCH[1]

[1]Grupo de Investigación en Ingeniería y Calidad del Software (GICS),
Facultad Regional Resistencia, Universidad Tecnológica Nacional, French 414,
Resistencia, Chaco, Argentina
{ns.pinto, csr.acn, lilianacp]@gmail.com

Abstract. In recent years, it has been given much importance to the use of models and quality standards on software development processes. Because these are those that facilitate continuous improvement and enable companies to provide higher quality products to its customers by increasing their competitive level.
Today, software development is based on agile processes that allow production characterized by its changing requirements and the need for continuous customer deliveries environments. Thus it is imperative to provide companies with tools for assessing the quality of these cycles agile processes.
QuAM is presented in this article, an approach to design a model of quality, integrated and flexible, that assesses the quality development cycles based on the principles and practices of the agile approach.

Keywords: Agility, Quality Software, Software Engineering, Agile methodologies.

1. Introduction

The final product quality, low cost and timely deliveries become key elements for the benefit of domestic sales and international projection of the Software Industry. In this sense, and in order to increase the quality and capability of their processes and, consequently, the quality of its products and services, software process improvement [1] becomes the differentiating element that companies in the sector need to increase their competitiveness.
In the case of Argentina, Software Industry consists mainly of PYMES, companies that represent 80% of the sector, according to the latest report of the Permanent Observatory of Software Industry and Information Services (OPSSI) [2]. Thus, taking into account this reality, it is important to note that several authors [3] [4] [5] agree on the difficulty that means for SMEs to implement programs of Software Process Improvement (SPI) mainly due to

the lack of monitoring action plans and implementation due to the high cost involved. Thus, the parameters of development time and cost of solutions will directly affect the work done, resulting the quality as the first variable adjustment available.
However it is not correct to consider the study or the impact of those elements that associate the quality of the final product but is also necessary to adjust the parameters associated with the processes that have facilitated obtaining it. Depending on this, there are numerous methodological proposals that guide the software development cycle and affecting different dimensions of the process. The more traditional methodologies are especially focused on a rigorous definition of roles, the activities involved, artifacts to be produced, and the tools and notations that will be used [6].
But these approaches are not the most suitable for many projects related to current scenarios where the system environment is changing and where it is demanded to drastically reduce development time without neglecting high levels of quality. So agile methodologies, pursue principles such as incremental delivery of new functionality to the client, which are prioritized according to business value added (so the software product evolves in different deliveries), favoring the continuous improvement and emphasis on close collaboration between the development team and business experts [7].
Previously it has been presented [8] a study on SMEs of Software companies in the North East of Argentina (NEA), in which the adoption of agile methodologies such as life cycle in their development processes was analyzed. From there the objective of this work is clear, that is to present QuAM, one first approach to the design of a model that allows quality assessment of agile processes, contemplating two perspectives: Process and Product. The article is structured as follows: in section 2 the state of the art is described through the presentation of related work. Then, in section 3, the characteristics of the proposed model by the design quality of experience for the same validation are presented. And finally, conclusions and future work are to be developed from this line of work are exposed.

2. Related Work

There are several models in the literature to assess the quality of software, quality trying break into a category of simpler characteristics and from two perspectives: the product and the process.
Among quality models that enable evaluation of the software product, the Model Mc Call, created by Jim Mc Call in 1977 [9] stands out. This defines 3 perspectives (Operability Product, Product Review and Transition Product) for the analysis of the quality of software, together with associated factors and criteria. The proposed metrics are questions that apply a numerical weighting to a particular software product attribute. After obtaining the values for all metrics specific criteria, the average of these is the value for that criterion. Another model worth mentioning is FURPS [10], developed by

Hewlett-Packard in 1987, in which a set of quality factors of software are described: Functionality (Functionality) Usability (Usability), Reliability (Reliability), Performance (Performance) and capacity support (Supportability). These elements can be used to establish quality metrics for all software process activities.

Among the international quality standards associated with the most relevant software is ISO / IEC 9126 [11], based on a hierarchical model with three levels: Features, Sub-characteristics and Metrics. The first level has six characteristics: Functionality, Reliability, Efficiency, Maintainability, Portability and Ease of Use. These characteristics (factors) are composed in turn by sub-characteristics (sub- factors) related to external quality, and sub-characteristics related to the internal quality.

There are also quality models that evaluate processes for obtaining software product. One of them, based on agile methodologies, is the AGIS (combination of AGILE and ISO) [12] which establishes a mechanism for measuring the degree of agility of software development processes. The ISO model supplemented with 10 dimensions; this configuration is oriented to measure the degree of implementation of the values of the agile manifesto [13] in the areas of engineering knowledge. AGIS aims to meet two needs: one focuses on companies, since this model can achieve differentiation from other companies that have only certified quality through ISO 9001: 2008. Furthermore, AGIS provides a report with suggestions for improvement based on the assessment of the size proposed evaluate.

Another model, similar to the above, is the AGIT (Agile Software Development) [14] which suggests that the best performance is achieved when the goals of all stakeholders are met. This requires an approach considering the views of different stakeholders, for which appropriate indicators are defined for each. AGIT considers four different views for stakeholders: the IT Manager is the actor concerned with traditional aspects of performance considering SW development time, cost and quality; the second actor is represented by team members whose goal is "job satisfaction"; the Scrum Master whose main goal is the "efficient resolution of impediments". Finally, the main objective seeking customers, the fourth stakeholder is its own satisfaction. This model suggests evaluate the quality of development processes considering the views of the different stakeholders involved, describing the indicators that are appropriate to each of these profiles.

On the other hand, it is also available the COBIT Model (Control Objectives for Information and Related Technology) [15], a tool that represents a particular collection of documents which can be classified, generally accepted as the best practices for the management, control and IT security. In COBIT, these domains is called: Plan and Organization (PO); Acquisition and implementation (AI); Delivery and Support (DS); Monitoring and Evaluation (ME). Through these four domains, COBIT has identified 34 IT processes. Through these four domains, COBIT has identified 34 IT processes and for each of these domains it defines goals and metrics to

determine and measure their results and performances, based on the principles of balanced scorecard business (BSC).
Finally, among the models that apply to processes of software development it is important to emphasize the CMMI (Capability Maturity Model Integration) [16], a model for the improvement and evaluation of processes for development, maintenance and operation of software systems. CMMI has four disciplines to choose from: Systems Engineering (SE), Software Engineering (SW), Processes and Products Development (IPPD) and Distribution (SS). The model itself has two representations. One of them is the staged representation in which it is centered a set of process areas, which are organized by maturity levels (1-5), while in the continuous representation, each process area are classified in terms of capacity levels (0-5). CMMI and agile methods have also been compared in several studies [17] [18] [19], for example, Paulk [20] suggests that the use of stories XP, at the customer's premises and continuous integration can comply with the requirements management objectives CMMI-SW. Moreover, in his study, Turner and Jain [21] determined that several of the components of CMMI and agile methods were in conflict, most of them related to organizational processes.
Among the presented models, it is observed that there is no proposal which allows quality assessment of agile processes themselves. Therefore, it is presented below the QuAM design, an approach which aims to provide a method of evaluation to determine the quality of software development processes based on Agile and its resulting products.

3. QuAM: Quality Evaluation Model of Agile processes

3.1 Design of the proposal

QuAM defines a scheme of components to set up a quality assessment model that provides an objective measure of the quality of the agile process implemented in any given project, allowing to obtain the agile profile associated with it. In this first instance, the model structure is presented taking into account only the dimension to the process level. Thus, the proposal that has been called Quam Level 1 provides a metric tree (Mi,i=1... 4) composed by measurable attributes (Ai) through a series of criteria with associated measures. Then, according to the scope of this article, the dimension of the quality model presented here will evaluate the following components based on:

- Metrics 1 - Election of Life Cycle: The life cycle of a software project defines the order of the process activities. Quam will consider better Iterative life cycles and incremental over others. Focus will be placed on the implementation of the process, not in the documentation generated. The attributes and criteria to be evaluated are presented in Table 1.

Table 1. Attributes and Metrics Criteria: Election of Life Cycle

Positive Attributes		*Negative Attributes*	
A1.1 - Value to Iterative and Incremental Life Cycle		*A1.2 - Value to Waterfall Life Cycle*	
Not complete iterations are performed, but new features are added to the product	0	The project is divided into strictly sequential steps.	-2

Table 1 (cont.). Attributes and Metrics Criteria: Election of Life Cycle

In each iteration, the product is revised and improved through refactoring.	1	Phases are executed simultaneously.	-1
At each iteration, not only the functionality are improved, but also new are added to the product.	2	At the end of each phase, it is possible to make a backtracking and improve the defined in the previous stage.	0

- Metrics 2 - Assessment of Team: The human component of the project to assess must have adequate skills to the agile philosophy, and the company must have the means to achieve it. For QuAM will be important to evaluate the flow of communication between team members and the ability to face the same agile practices. The attributes and associated criteria to this metric are presented in Table 2.

Table 2. Attributes and criteria associated with Metric: Assessment of Team

Positive Attributes		*Negative Attributes*	
A2.1 - Value to team meetings		*A2.2 - Value of compliance schedule*	
No meetings are held in all iterations.	0	The schedule is adapted according to the changes and needs that arise throughout the project.	-3
In each iteration a meeting is done virtually at least.	1	Control milestones are set out in the schedule and changes can be defined on the scheduled dates.	-1
In each iteration, at least one meeting is held with the physical presence of the entire team.	3	The schedule established by stages is strict and does not allow changes.	0

A2.3 - Value to the definition of roles.		*A2.4 - Value to the process by over the team.*	
No roles are defined for individuals.	0	Activities, deliverables and development and management tools are defined for the project.	-3
A clear definition of roles is performed on individuals team.	1	Activities and project deliverables are defined.	-2
A clear definition of roles and responsibilities is done between team members.	2	Activities for each iteration are defined in the project.	-1

Table 2 (cont.). Attributes and criteria associated with Metric: Assessment of Team

A clear definition of roles, responsibilities and interaction between team members are made.	3	Activities for the project are defined but not at the level of each iteration	0

- Metric 3 - Production capacity of deliverables: QuAM will evaluate the frequency with which the project produces deliverables versions from the product to the customer. In this component, it will take into account the compliance with the lead time and the validity of each deliverable, favoring those projects whose validation has been automated. The change management process will be also measured the change management process will be also measured about the product and the verification process implementation and validation of them. In Table 3, the attributes and criteria that are considered for this metric are included.

Table 3. Attributes and criteria associated with Metric: Production capacity of deliverables

Positive Attributes		*Negative Attributes*	
A3.1 - Value of the use of change management tools.		*A3.2 - Value to requirements management and requisites.*	
There is a single project in the change management tool with a single workflow shared by all team members.	0	The Document of Software Requirements Specification (SRS) is updated simultaneously with the software	-3
There is a unique project in the	1	SRS is updated only if new	-1

change management tool but not all team members have their workflow (branch).		requirements are added to Software	
There is a unique project in the change management tool used, and workflows (branches) are administered by each team member involved	3	SRS can not be upgraded, and must be strictly enforced.	0
A3.3 - The value to functional product.		*A3.4 - Value to documentation.*	
Generate deliverable upon project completion without making testing.	0	It requires detailed documentation at project start.	-3
Generate deliverable with manual testing after each iteration.	1	It requires only documentation needed at the beginning of each iteration.	-1

Table 3 (cont.). Attributes and criteria associated with Metric: Production capacity of deliverables

Generate deliverable with testing automated and integrated with other functions after each iteration.	3	No documentation is required to begin implementing the functionality included in one iteration.	0

- Metrics 4 - Customer communication: The quality model proposed will propitiate the incorporation of the client, as an active member in all stages of the project. Thus, this metric will assess implementation of regular communication mechanisms between the client and the team.

Table 4. Attributes and criteria associated with Metric: Customer Communication

Positive Attributes		*Negative Attributes*	
A4.1 - Assess collaboration with the customer.		A4.2 - *Assess contract negotiation.*	
Customer collaborates to team demand.	0	There is detailed recruitment at the beginning and no changes accepted.	-3
Customer is part of the team, answers queries and plans iterations.	1	The contract requires consider changes during the project.	-1

Customer is part of the team. He responds consultations, planning iterations, and collaborates on writing and testing requirements	3	The contract exists but does not affect the level project development process.	0

It is worth mentioning that for the design of this proposal, and taking into account defined in [22], positive attributes were considered (those who try to emphasize), and negative attributes (those who try to belittle). Thus, the positive attribute is measured on a scale from 0 to 3, and the negative attribute on a scale from -3 to 0. Thus, each metric could obtain a measure of -3, in the case that both attributes take the worst value (-3 for negative and the positive attribute 0), and 3, in the case that both attributes take the best value (0 negative attribute and 3 positive attribute). If a zero or near zero value is obtained, it means that the measurement values are not significantly above the positive negative.
Therefore, and taking into account the details of the associated criteria for the final value of each metric, it must consider both the corresponding measure to positive as the associated to the negative and the sum of its values, shown in (1):

$$Mi = M(Ai.1) + M(Ai.2) \quad i=1..4 \tag{1}$$

For example, the metric 1 (M1) - Election of the life cycle is measured by adding the measure of the value that the process gives the cycle of iterative and incremental life (A1.1) over the life cycle cascade (A1. 2).

3.2 Validation: Experience design

It is also necessary, design the process of validation of information that QuAM provides with the tools necessary to do so. To do this, in principle, the SMEs companies of the NEA Software Industry are convened, to assist in the validation of QuAM with real production environments to detect successes or issues to be improved in the model definition.
One of the instruments to be used will be an online survey with closed questions referred to the SW development processes of these companies. The same, it was designed and implemented through Google Forms, to facilitate dissemination among participants of experience and maximum reliability in the process of gathering information. The target population is made up of a group of 15 NEA companies, characterized by work on development projects web applications.
Currently the validation process has started with 25% of companies invited to take part of it. And it is expected that once the process of data collection is completed, an analysis of the obtained will be conducted to generate partial

reports to determine the level of quality associated with agile business processes. The process will not end there, but the information obtained must be filed with the involved to achieve the feedback to determine if the concluded from the use of the proposed model is approximately or not, the reality perceived by the companies.

4. Conclusions and Future Work

The main contribution of this work is the preliminary definition of the components that form part of a quality model that helps to assess the quality of agile processes in SMEs dedicated to software development. There are several papers in the literature with the aim of improve the quality of the development process, submit proposals to adapt norms and standards to the philosophy of agile methodology. However, they are not specifically focused on the evaluation of the results obtained by processes under the agile philosophy. Thus, in principle, the QuAM presentation as a new approach to quality model will allow to start the cycle of quality assessment in real software projects guided through streamlined processes.
As future work, it is intended to obtain results of the validation experience presented in this article. And based on that, start defining a framework, including guidelines and practical guides, whose objective is the automation of measuring the quality of software projects based on agile processes.

5. Acknowledgements

This work was done under the accredited research project "Assessment Framework for Software Quality" EIUTIRE0002205TC of UTN. It should also be noted that the article is part of the activities planned in the Project on Technology and Social (PDTS) Development presented, "Contribution to the competitiveness of Software development NEA companies", IP253, evaluated and approved by the National Council of Scientific and Technical Research of Argentina (CONICET).
It is necessary to acknowledge the assistance of researchers Engineer Nicolás Tortosa, Blas Cabas Geat and Maximiliano Ulibarrie and collaboration of the translator Miryam González de Pinto on gramatical English review of this paper.

6. References

1. Navarro, J. M., & Garzás, J. (2010). Experiencia en la implantación de CMMI-DEV v1. 2 en una micropyme con metodologías ágiles y software libre. REICIS. Revista Española de Innovación, Calidad e Ingeniería del Software, 6(1), 6-15.
2. Reporte anual sobre el Sector de Software y Servicios Informáticos de la República Argentina. OPSSI, ABril 2016. Disponible en

http://www.cessi.org.ar/descarga-institucionales-2007/documento2-130347cd83ae771a9f3db3da5407269a
3. Mas A., Amengual E. (2005). “Las mejoras de los procesos de Software en las pequeñas y medianas empresas (pymes). Un nuevo modelo y su aplicación a un caso real”. Revista Española de Innovación, Calidad e Ingeniería del Software, Vol.1, No. 2.
4. Pasini, A. C., Esponda, S., Bertone, R. A., & Pesado, P. (2008). “Aseguramiento de Calidad en PYMES que desarrollan software.” XIV Congreso Argentino de Ciencias de la Computación.
5. Pflegger, S. (2002) “Ingeniería de Software. Teoría y Práctica.” Pearson Education.
6. Letelier, P., Penadés, P. (2006) “Metodologías ágiles para el desarrollo de software: eXtreme Programming (XP)” Técnica Administrativa, Buenos Aires. ISSN 1666-1680
7. Alliance, A. (2001). “Agile manifesto”. Disponible en http://www. agilemanifesto. org
8. Rujana, M., Romero Franco, N., Tortosa, N., Tomaselli, G., & Pinto, N. (2016, May). Análisis sobre adopción de metodologías ágiles en los equipos de desarrollo en pymes del NEA. In *XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina)*.
9. Córdoba, J., Cachero, C., Calero, C., Genero, M., & Marhuenda, Y. (2007, October). Modelo de Calidad para Portales Bancarios. In *XXXIII Conferencia Latinoamericana de Informática (CLEI'07)*.
10. Behkamal, B., Kahani, M., & Akbari, M. K. (2009). Customizing ISO 9126 quality model for evaluation of B2B applications. Information and software technology, 51(3), 599-609.
11. ISO/IEC 9126: “Software Engineering - Product quality”, International Organization for Standardization, 2000.
12. Matalonga, S., & Rivedieu, G. (2015). AGIS: hacia una herramienta basada en ISO9001 para la medición de procesos ágiles. Computación y Sistemas, 19(1), 163-175. Disponible en http://www.agilemanifesto.org/iso/es/ Último acceso 06/2016
13. International Organization for Standardization. (2000). ISO 9001: 2008: Quality Management Systems-Requirements. International Organization for Standardization.
14. Cohen, D., Lindvall, M., & Costa, P. (2003). Agile software development. DACS SOAR Report, 11.
15. Paulk, M. C. (2001). Extreme programming from a CMM perspective. Software, IEEE, 18(6), 19-26.
16. Piattini, Oktaba, Orozco, “COMPETISOFT. Mejora de procesos software para pequeñas y medianas empresas”, Editorial Ra-Ma, Año 2008.
17. D. Kane and S. Ornburn, "Agile Development: Weed or Wildflower?" CrossTalk, The Journal of Defense Software Engineering, http://www.stsc.hill.af.mil/crosstalk/2002/10/kane.html, 2002. (1.3.2006)
18. J. Nawrocki, W. Bartosz, and A. Wojciechowski, "Toward Maturity Model for eXtreme Programming," In proceedings of the 27th Euromicro Conference, pp. 233-239, 2001.
19. M. C. Paulk, "Extreme Programming from a CMM Perspective," Software, vol. 18, issue 6, pp. 19-26, 2001

20. R. Turner and A. Jain, "Agile Meets CMMI: Culture Clash or Common Cause". In proceedings of the Second XP Universe and First Agile Universe Conference on Extreme Programming and Agile Methods - XP/Agile Universe, pp. 153-165, 2002.
21. Turner, R., & Jain, A. (2002, August). Agile meets CMMI: Culture clash or common cause? In *Conference on Extreme Programming and Agile Methods* (pp. 153-165). Springer Berlin Heidelberg.
22. Mendes Calo, K., Estevez, E. C., & Fillottrani, P. R. (2009). Un framework para evaluación de metodologías ágiles. In *XV Congreso Argentino de Ciencias de la Computación*.

Web Applications Requirements: A Taxonomy

SILVIA SÁNCHEZ-ZUAÍN[1], ELENA DURÁN[1,2]

[1]Facultad de Matemática Aplicada
Universidad Católica de Santiago del Estero (UCSE)
szuain@ucse.edu.ar; elena.duran@ucse.edu.ar
[2]Facultad de Ciencias Exactas y Tecnologías
Universidad Nacional de Santiago del Estero (UNSE)
eduran@unse.edu.ar

Summary: Web applications provide a variety of functionalities with different features and requirements that allow for varied classifications. Categorize Web applications is useful for understanding their needs and developing and implementing Web-based systems. To define requirements considering web application category may be beneficial. In this work, a taxonomy for web applications requirements is proposed out of the analysis of different categories and of the identification of the applications requirements. On this basis, a three-level taxonomy was built; at the first the general requirements were differentiated from the specific ones; at the second, for the specific requirements, their web application type is considered; and eventually, at the third level, the requirements are classified into functional and non-functional. The resulting taxonomy will help web-developers to minimize errors along the developing stage, contribute to generate highest quality web-systems, so that delays affecting both the development team and the client will be avoided.

Key words: Web applications, Requirement Engineering, Software requirement classification.

1. Introduction

Web applications offer today a variety of functionalities with different features and requirements that allow for varied classifications. Their scopes and uses have expanded making the daily world strongly dependent. Also, they capturing the user characteristics by incorporating customizing techniques. This leads to the web application categorization and to categories likes semantic or ubiquity, where software can process its content, reason with it, combine it and make logical deductions to solve daily problems automatically from everywhere and whichever device. To categorize web applications is useful for understanding their needs and to develop and implement web-based systems and applications [1].

Along the process of developing an application, the working team faces the problem of identifying requirements. This stage aims to analyze and

document the functional needs that must be supported by the system to be developed. The set of requirements must be as much complete, consistent and correct as possible. The web applications special features must be considered at the phase of requirement specification [2]. In this way, web-developers' work will be easier and more simple and consequently, the client will be given a high quality and functionality solution.
In this work, a taxonomy of requirements is presented. It was made from the different web applications categories surveyed; since identifying to what category a web application belongs, the types of requirements that at least this must satisfy, can be established.
The paper is organized as follows: section two reviews background knowledge on web applications; section three reviews software requirements classifications. Then, in section four, the taxonomy and the process carried out to developing it is explained, and eventually the conclusions drawn from the work are stated.

2. Background information on web applications categories

To select the web applications categories that might serve as the basis for defining the Requirements Taxonomy, different classifications proposed by the following authors were analyzed. Ginige and Murugesan [4] propose one that illustrates the evolution of web applications. This is useful for comprehending their needs, developing and implementing web systems. Kappel et al. [3] include into this classification, new types such as the ubiquitous and semantic web applications. They also make it evident the relationship between the development chronology and complexity. Cohelo et al [5] categorize starting out of the analysis of the HTTP traffic by which a client asks the server for a web page, and the server searches and sends the client the page through the net. Softaculous [6] has made a list considering that also exist in the cloud, web application installing services with a wide range of applications available to them. Pressman [7] introduces a category based on software process activities. Rossi G. et al. [8] define a category meeting the main characteristics and the technology used for its creation.
Out of these classifications, a table showing the categories appearing in most of the works analyzed was designed. To do so, those with the same meaning were unified despite their authors named them differently. Additionally, some types of web applications were grouped into the more general categories in terms of their main service, use or functionality so that the number of domains was reduced as much as possible for making them manageable at the time to identified the requirements. The resulting web applications categorization used for constructing the requirements taxonomy is shown in Table 1.
Next, the web application types included in Table 1 are defined according to the authors consulted.

1. **Interactive applications:** they interact with the user; their content is constantly changing and is generated according the user needs. The use of HTML forms and of the Common Gateway Interface (CGI)

provide new ways and aspects enhancing the interface and menus. For example: online games applications.

Table 1. Synthesis of the categories resulting from the analysis

Web Applications Categories	Authors					
	Kappel et al.	Ginige and Murugesan	Cohelo et al	Softaculous	Pressman	Rossi G. et al
1. Interactives	x	x		x	x	x
2. Transactional	x	x		x		
3. Collaborative	x	x		x		
4. Web sites	x	x		x	x	
5. Social web		x		x		
6. Service-oriented		x	x	x	x	
7. Ubiquitous (mobile)	x					x
8. Semantic	x					x

2. **Transactional applications:** they were created to give the user higher interactivity by allowing them not only reading but also site-content updating actions. For example: electronic buying.
3. **Collaborative applications**: they are used for collaborative objectives. They have a constructive and active sense, involve a developing process as a strategy inviting the user to interact by exchanging knowledge and capabilities. For example, distributed auditing applications.
4. **Web sites:** they offer a single access point to independent, potentially heterogeneous sources of information and services. For example, www.educar.ar
5. **Social web**: they allow people to communicate one to another and work across organizational and physical limits, creating what is known as "communities". They are based on the links existing among their users. There are several types of social webs: general, professional and vertical or thematic. For example, www.facebook.com
6. **Service-oriented applications:** this type of applications attempts to provide a specialized service, include an endless number of online tools, software and platforms to offer the final user added-value services. For example, Postal Tracking applications.

7. **Ubiquitous applications:** they provide customized services. The user can access to the information from everywhere, at any time, regardless they type of device used. For example, Tourist assistance applications.
8. **Semantic Applications:** they supply information not only for human understanding but also the proper system that manipulate them. This would facilitate knowledge management on the web. For example, semantic web browsers.

3. Background information on software requirements classification

Various authors supply different classifications and definitions of diverse requirements. In this work, the better-known classifications are presented: Sommerville [9], Roman and Fairley [10], Grünbacher [11], Escalona and Koch [12] and Standard IEEE [13]. Since these authors called requirements with the same meaning differently, they were grouped in such a way that their identification and comprehension become easier. Following, Table 2 presents a synthesis of the categories analyzed.
In general, within Software Engineering settings, requirements are classified into two groups: functional and non-functional. However, other elements of interest must be considered when considering web applications developments. Thus, requirements of content, system environment, user interfaces and evolution of the application are met. In addition, they must meet such quality requirements as performance, usability, scalability, maintenance and accessibility among others.

Table 2. Synthesis of the requirements categories analyzed

Requirements			Authors				
Types	Categories	Sub-Categories	Sommerville	Roman and Fairley	Grünbacher	Escalona, and Koch	Standard IEEE
Functional			x	x	x	x	x
Non-Functional	Quality	Functionality			x		
		Usability	x		x	x	
		Efficiency	x	x	x		x
		Trustability	x	x	x		x
		Maintenability		x	x	x	x
		Mobility		x	x	x	x
		Security	x	x			x
	Organizational	Setting	x		x		
		Operational	x	x			
		Development	x	x		x	x
	External	Ethic	x				x
		Legal	x	x			
		Economic		x			
	Interface	User		x	x	x	x
		System		x	x		x
	Of Evolution				x		
	Of Content				x	x	x
	Of Navigation					x	
	Of Personalization					x	

4. Development of the Taxonomy

A taxonomy is a scientific process (or a system) for categorizing entities, that is, for organizing them into groups [14]. A taxonomic system must be clear, consistent, flexible, comprehensive and practical. Through the taxonomy, categories can be set up within the classification, whether relations of likeness (principle of interaction) or interdependence (principle of duality) are to be determined. A structural unit is fixed that in turn will be the terminal unit termed a Taxon [15].

4.1 Features of the Web Application Requirements Taxonomy.

In accordance with Argudo and Centelles [16] the basic features of the taxonomy were defined as follows:

- ✓ *Scope:* the taxonomy is within the Requirement Engineering and in particularly it is for web systems. The main users will be web engineers and web developers who take part in the designing process of the Web System (analysts, clients, users, graphic designers, multimedia and security experts, etc.). The taxonomy will exclusively contain web application requirements. It will be written in Spanish, exception made to those in English with no translation into Spanish.

- ✓ *Uses and functions:* it will be used to support the development of web applications, as a tool for Requirement Engineers and web developers, to guide the creation of a web application along the requirement specification stage, like a checking list for determining whether all types of requirements have been used considering the type of web application to be developed.
- ✓ *Taxonomy type*: visible to the user.
- ✓ *Categorization complexity degree:* it will grow steadily in as much as other new categories of web applications appear.
- ✓ *Taxonomy objective:* "To provide a structure for classifying requirements related to various kinds of web applications, that serves as support for developing web systems.
- ✓ *Existing tolls utilized for defining the Taxonomy:* the resources used for defining the taxonomy were: the categories of web applications introduced and summarized in Table 1; the categories of requirements introduced and summarized in Table 2.

4.2 Taxonomy Construction Process

In the construction of a taxonomy, four necessary stages can be distinguished [16], which are described below for the proposed Taxonomy

4.2.1 Faceting criteria identification

Facets stands for the aspects, properties or features of a specific reality clearly defined, mutually exclusive and, altogether, exhaustive. Each facet is decomposed in categories of different levels of specificity. For their correct identification, the results of the context, audience and content analysis must be considered.
For the Requirements Taxonomy of Web Applications, the taxon, i.e. the classification unit, is the web requirements. Even though software requirements express the needs and restrictions of a software product or a service, they have neither the structural simplicity nor homogeneity proposed by most of the authors consulted. On the other hand, software requirements are defined in different settings and at different level of detail. That is why, in this work, three large criteria defined below have been considered.

Criterion N° 1: in terms of requirement granularity, it is considered:

- **General Requirements:** they are independent of the type of web application to be developed. Encompass the basic requirements will be considered at the time the software prototype starts being designed and reflects thus the client's requirement. Additionally, they will help the entire team have a clear objective to be achieved in designing the web application. Therefore, it is important in an initial stage to have the general requirements well defined at the time the development of a web application is faced. These are:

- ✓ *To know the target user*: to define what kind of public the web application will be serving (students, workers in a company or any other type of user)
- ✓ *To define the web context:* all the developers must understand "why" and "what for" the web application is being created to avoid adding things that are estrange to the context desired.
- ✓ *To define the kind of web application* that will be created in terms of the context and of the targeted users.
- ✓ *To consider each actor (participant)* that will be part of the team facing web application design and development: to consider, in based on the type of application and its basic requirements, what the required profiles are during the application development to distributed the responsibilities (requirement engineers, designers, architectures, etc.).

- **Specific requirements.** These are identified and defined considering the web application type, that is, a web system that does not meet any of the requirements presented here, will not be accepted.

Criterion N° 2: Consider the web application category established in this work.

Criterion N° 3: Define requirements for each web application type.

- **Functional:** a description of "what" the system does. These requirements answer the question *"what can be done in the web application?* i.e. what it is possible to be done using web application information and possible functionalities must be defined [18].

- **Nonfunctional:** a description of "how" the system does it, which involves defining how the system has to behave or what limits upon the solution are imposed by environment. In any web application, several needs appear that can be catalogized as nonfunctional requirements for web applications. The following can be included: quality (usability, efficiency, reliability, security, maintainability, portability); organizational (environment, operational, developments); external (regulatory, ethic, legal/political, economic); interface (user, system. This las one includes hardware, software and communication); performance; Design Restrictions; content requirements; evolution requirements; conceptual requirements; personalization requirements; adaptation requirements.

4.2.2 Lexis Extraction

This phase aims to identify all the terms and categories denominating the proper concepts of the domain. Each category must link, at least, with one of the facets set in the previous stage.
Most of the lexis extraction process was carried when the requirements tables were constructed in Section 3. The terms and categories identified within the domain are: Web Applications Requirements, General Requirements, Specific Requirements, Web Application Types, Functional Requirements, Non-functional Requirements. The entire set of denominations is technically

referred to as the domain lexis. Each denomination is called a term, or a category.

4.2.3 Lexis Control

In the previous stage, it may have happened that the same concept was identified with different denominations (i.e. synonyms or quasi-synonyms) and/or the same denomination showed different forms (grammatical, orthographic, etc.). For example: Interface Requirements (to the user) are also called interaction requirement or user in some proposals. To be effective is recommended that a taxonomy has a preferable term to represent each one of the concepts that make up its domain. Upon controlling the lexis, the process performed when constructing Table 2, the requirements for web applications would be listed and grouped into this table.

4.2.4 Taxonomy Structure

This stage aims to identify and establish two kinds of relationships among the categories of the taxonomy: the hierarchical and the associative relationships. The result will be a controlled vocabulary like a conceptual structure. The hierarchical relationship is based on higher degrees or levels and subordination where a general term represents a whole or class and the subordinate terms correspond to its members, parts or instances. All the categories of a facet must be connected one to another by hierarchical relationships. The associative relationship connects categories from different facets out of idea associations that may occur between: an action and its result or product; a concept and any of its properties; a product and the material from which it is made and so on.
Traditionally, two techniques for developing the taxonomy structure have been distinguished [17]

- *Up to down technique* supposes the initial identification of a given number of higher categories and grouping the remaining categories in successive levels of subordination up to reaching the more specific categories. This technique can be applied to both a hierarchical (and/or tree) structural and a faceted model.
- *Down to up technique* is based on the initial identification of the more specific categories that later are grouped into successive levels of super-ordination up to the higher categories level is reached. Generally, this technique has essentially been applied to a hierarchical (and/or tree) structural model though as in the previous case, it can facilitate the analysis to decide which structural model that is suitable to apply.

In this work, the *Down to up technique* was chosen. With this technique, a hierarchical classification will be obtained, through a lower-to-higher order that will give a sense of collectivity and generality.

4.2.5 Determining the Taxonomy levels

Based on the taxonomic criteria defined in Section 4.2, three levels were determined as shown in Figure 1.

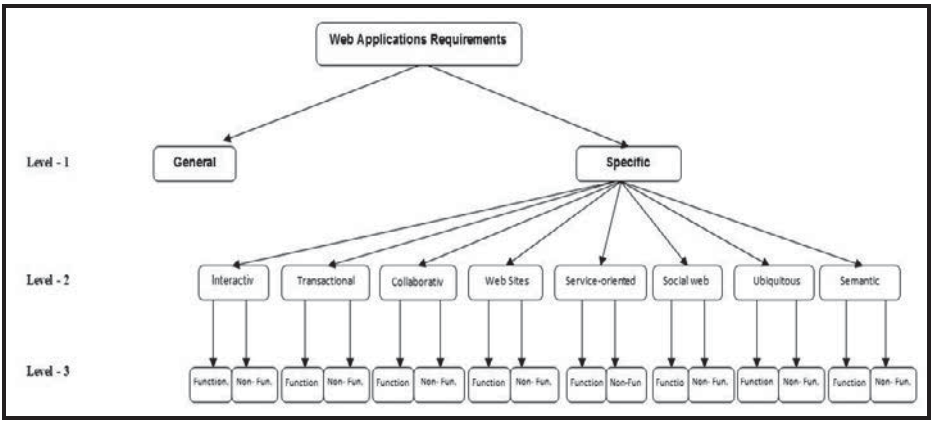


Fig. 1: The Taxonomy structure

Additionally, to consider in the Taxonomy structure what was defined in criterion 3, Section 4.2.1, it was identified which of the requirements included in Table 2 must be defined in each type of web application introduced in Table 1. From this, Table 3 was constructed to summarize what requirement must be defined for each web application type.

5. Conclusions and future Works

The proposed taxonomy was designed in order to define a requirements classification structure, related to the different types of web applications, that support the development of web systems. Out of the analysis of the different categories of the better-known web applications a taxonomy of web applications requirements was posited. This work lays the conceptual foundations for the construction of any requirements taxonomy, including one refining the one presented here, since the processes carried out are explained: faceting criteria identification, lexis extraction, lexis control and the taxonomy structure development.

The proposal introduced in this work will be useful during the Requirement Engineering Stage which is primordial in software production process since it focuses on defining what will be produced. With the taxonomy will be possible to accelerate the requirement identification stage provided that both the client's objectives and the type of web application to be developed are already identified. Counting on a Requirements Taxonomy allows to determine a web software project needs by attending its more relevant

aspects. The resulting Requirements Taxonomy of this work becomes a supporting tool for every actor involved in RE and will allow in the eliciting process to obtain and specify requirements bearing in mind the type of web application to be developed.

Table 3. Requirements related to the web application type.

Requirements			Web Applications							
Types	Categories	Sub-Categories	Interactives	Transactional	Collaborative	Web Sites	Social Web	Service-oriented	Ubiquitous (mobile)	Semantic
Functional			x	x	x	x	x	x	x	x
Non-functional	Quality	Functionality	x	x	x		x	x	x	
		Usability	x	x	x	x	x	x	x	x
		Efficiency	x	x	x		x			
		Trustability	x	x	x		x		x	x
		Security	x	x	x		x	x	x	
		Maintenability	x	x	x	x	x	x		
		Mobility	x	x	x		x	x		
	Organizational	Setting	x	x	x		x			
		Operational	x	x	x		x	x	x	
		Development	x	x	x	x	x	x	x	x
	External	Ethic	x	x	x	x	x	x	x	x
		Legal	x	x	x	x	x	x	x	x
		Economic	x	x	x	x	x	x	x	x
	Interface	User	x	x	x	x	x	x	x	x
		System	x	x	x	x	x	x	x	x
	Of Evolution	Chance			x					
		Increase		x	x		x	x	x	
	Of Content		x	x	x		x			x
	Of Navigation					x			x	x
	Of Personalization		x	x	x		x		x	x

As future work and aiming at drawing consistent conclusions in time, the taxonomy will be applied to real cases to compare the requirements identified using the traditional development against those out of the taxonomy proposed in this paper. On the other hand, it might also be possible to face the construction of a taxonomy based on system architecture views. To represent complex systems, it is better to divide the various though related points of views of involved ones (final users, developers or project directors), each one describing an aspect of the system architecture. Altogether, the views describe the entire system. Each view represents a specific behavior of the system. These views attempt to model the requirements in accordance with the function they perform.

6. References

1. Woojong, S.: "*Web engineering: principles and techniques"* Idea Group Inc., 2005.
2. Escalona, M. J. and Koch, N.: "*Requirements Engineering for Web Applications a Comparative Study,"* J. Web Eng., vol. 2, N° 3, 200e4.
3. Kappel, G.; Pröll, B.; Reich, S. and Retschitzegger, W.: "*Web Engineering. The Discipline of Systematic Development of Web Applications"*. G. Kappel, B. Pröll, S. Reich, & W. Retschitzegger (eds). John Wiley & Sons Inc. 2006.
4. Ginige, A. and Murugesan, S.: "*Web Engineering: AnIntroduction"*, IEEE Multimedia, Jan-Mar 2001.
5. Coelho, N.; Salvador, P. and Nogueira, A.: *"Differentiation of HTTP Applications based on Multiscale Analysis."* Bentham Science Publisher, vol. 2, N° 1, pp.12-25. 2013.
6. Softaculous Ltd. (s.f.). Obtenido de https://www.softaculous.com/apps.
7. Pressman, R.: "*Ingenieria del Software. 6ª Ed*." Mcgraw-Hill.n. 2005.
8. Rossi, O.; Pastor, O.; Schwabe, D. and Olsina L.: "*Web Engineering: Modelling and Implementing Web Applications"*. (Eds), Springer. 2007.
9. Sommerville, I. "*Ingeniería del Software"*, 9ª ed. Addison Wesley, 2011.
10. Roman, G.-C.: "*A Taxonomy of Current Issues in Requirements Engineering,"* IEEE Computer, Vol. 18, No.4, pp.14-22. April 1985
11. Fairley, R. E.: "*Software Engineering Concepts"*, McGraw-Hill, 1985.
12. Escalona, M. J. and Koch, N.: "*Metamodeling the Requirements of Web Systems"*. 2nd International Conference on WebIST, Setubal, Portugal, 2006.
13. Standards Association, IEEE Recommended Practice for Software Requirements Specifications. IEEE Std 830 1998. Disponible en: http://www.ieee.org/publications_standards/index.html
14. Abed Gregio, A.R.; Barbato, L.G.C.; Duarte, L.O.; Montes, A.; Hoepers, C.; Stedding-Jessen K. *"Taxonomías de Vulnerabilidades: Situação Atual"* Disponible en: http://www.lbd.dcc.ufmg.br/colecoes/sbseg/2005/009.pdf consultado el 8/04/2016.
15. Currás, E.: "*Ontología, Taxonomía y Tesauros. Manual de Construcción y Uso"*. Ediciones TREA S.L. Tercera Edición 2005.
16. Argudo, S.; Centelles, M.: *"Metodología para el diseño de taxonomías corporativas"*. Investigación Bibliotecológica, vol. 19, núm. 39, 2005.
17. Centelles, M.: "*Taxonomías para la categorización y la Organización de la información en sitios"*. 2005, disponible en http://www.hipertext.net, consulta realizada el 8/04/2016.
18. Escalona, M.J.; Mejías M., Torres J; Reina A.M.: *"The NDT Development Process."* Proceedings of IV ICWE 2003.

A Methodology for Assessing the Maturity Level of University Services

ARIEL PASINI[1], ELSA ESTÉVEZ[2], PATRICIA PESADO[1], MARCOS BORACCHIA[1]

[1] Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática –UNLP,
50 y 120, La Plata, Buenos Aires, Argentina
[2] Departamento de Ciencias e Ingeniería de la Computación,
Universidad Nacional del Sur
Av. Alem 1253, (8000) Bahía Blanca, Argentina
{apasini,ppesado,marcosb}@lidi.info.unlp.edu.ar, ecestevez@gmail.com

Abstract. This paper introduces the concepts of University Electronic Government and University Service, and a methodology for assessing the maturity level of such services by applying e-government concepts. Based on the identification of 25 basic services provided by academic units (AUs), that can be parameterized according to their context, a maximum achievable maturity level is defined for each service as a function of service automation. The proposed methodology enables to assess and plan potential improvements to the provision of the identified basic services by universities to their community members.

Keywords: University Electronic Government, University Public Services, Maturity Model, Quality Assessment

1. Introduction

The terms e-Government and Digital Government commonly refer to services delivered by government structures, including municipal, provincial, or national governments. However, there are some public institutions, not part of government, that have their own government structures with extensive institutional autonomy. Usually, the national or the provincial government, through laws, establishes the basic government structure of such units and assigns a budget so that they can achieve certain objectives, but it does not participate in the selection of authorities or the execution of the budget. These public institutions include national and provincial universities in Argentina.
Even though the state of the art in e-government has progressed significantly in the past two decades, thanks to the contributions of many research and practice works, little progress has been achieved in the application of the

concepts, methods, tools and practices of e-government to the university scope, as explained in Section 4.
In this paper, concepts related to university e-government (EGOV-U) are introduced, and a methodology to assess the maturity level of university services is proposed. The main contributions of this paper include the introduction of the concepts of University e-Government and University Service, and the proposal of a methodology to define a maturity model for providing such services.
The remaining sections of this paper are organized as follows. Section 2 provides background information and introduces the concepts used later. Section 3 discusses related works. Section 4 explains the methodology proposed to assess the maturity level of university public services and summarizes its application at the School of Computer Science of the National University of La Plata (UNLP) in Argentina. Finally, conclusions and future work are discussed in Section 5.

2. Background and Concepts

The following sections provide background information, describe the university context, introduce the concepts of university e-government and university service, and present a maturity model for such services.

2.1 Background

The university education system in Argentina currently consists of 129 educational institutions, 53 national universities, 3 provincial universities and 7 state tertiary education institutes [1]. There are also 50 private universities, 14 private tertiary education institutes, 1 international university and 1 foreign university.
The Argentinean Higher Education Act [2] establishes, in its Article 29, that university institutions shall have academic and institutional autonomy. Among its responsibilities, they must produce their own by-laws and government body. In particular, its Article 52 establishes that the by-laws of national universities must plan their government bodies, including a group of professional and individuals, as well as their members and responsibilities. Professional associations shall have basically the same general regulatory, policy definition and control attributions in its respective scopes, while individuals shall have an executive role.
The group of professionals is called “High Council”. For most universities, the Act itself defines their composition. For instance, for the National University of La Plata (UNLP), more than 50% of the members of the High Council must be lecturers, and the remaining members must be students with at least 30% of their courses approved, representatives of the administrative staff and, optionally, graduated students[3] . The representatives of the

government body are chosen by their respective academic units. The person that is part of the Council as an individual is commonly referred to as the President or Rector. Elections can be through direct vote or election by the Higher Council, as established in the by-laws of each University.

2.2 University e-Government and University Services

As already mentioned, university government bodies include lecturers, administrative staff, students, and graduated students. As a whole, they represent the university community and carry out their activities in accordance to the regulations set forth by such government bodies. To enforce its regulations, each university offers a set of services to its community. Currently, several of these services are provided through Information and Communication Technologies (ICTs).

The concept of University e-Government (EGOV-U) is defined as "the use of ICTs as a tool to improve processes and services provided by a university to the members of its community."

A Public Service is defined as "the result of a process carried out by a public entity, or an entity, specialized or not, that is controlled and regulated by a public entity, aimed at meeting the needs of the public.*"* Adapting this definition to the university context, a University Service is defined as "the result of a process carried out by a university, or by another organization, specialized or not, that is controlled and regulated by the university, aimed at meeting the needs of the members of the university community". In particular, the members of the university community include students, graduated students, lecturers, and administrative staff.

For instance, university governments provide public services to students when they provide information about the academic calendar; offer post-graduate courses to graduated students; provide public examination assessment for their positions to lecturers; and keep attendance records to administrative staff.

Services are provided through various channels. Delivery channels are classified as follows: 1) traditional channels, such as counter, phone, or fax; and 2) electronic channels, such as a web site, e-mail, mobile devices, and social networks. Given the alternatives in relation to channels, it should be noted that not all services can be delivered through all channels, and the selection of a given channel affects service delivery cost, response time for service delivery, and service user satisfaction.

2.3 Maturity Level of University Services

The maturity level of public services can be assessed based on the automation level achieved for delivering the service and the level of support provided by

ICTs to the business process delivering the service. The United Nations proposes a four-level maturity model[1]:

1) *Emerging* – Government websites provide information about public policies, governance, laws, regulations, and the types of public services that are provided. Links to ministries, departments and other government branches are present. Citizens can obtain updated information about the national government and its ministries, and they can click on links to access information that has been filed.
2) *Enhanced* – Government websites deliver on-line, one-way, improved or on-line services, simple two-way communication between the government and citizens is feasible, such as downloadable forms for government services and applications. Websites include audio and video resources, and offer information in more than one language. Some limited electronic services allow citizens to send requests for non-digital forms or personal information.
3) *Transactional* – Government websites involve citizens in two-way communication, including requesting and receiving information on government policies, programs, regulations, etc. Some form of electronic authentication is required to verify citizen identity for successfully completing the exchange. Government websites process non-financial transactions, such as filling in tax forms, requesting certificates, and applying for licenses and permits. Additionally, financial transactions are also processed, i.e., services for transferring money through a secure network are provided.
4) *Integrated* – Government websites change the way in which governments communicate with their citizens. They are proactive in requesting information and feedback from citizens using Web 2.0 tools or other interactive tools. Electronic services seamlessly cross department and ministry boundaries; information, data and knowledge are transferred between government agencies through integrated applications. Governments have transitioned from a "government-centric" approach to a "citizen-centric" approach, where electronic services are targeted to citizens through life-cycle events and segmented groups to provide tailored services.

By adapting the definitions listed above to the scope of EGOV-U services, the following maturity model is defined:

1) *Emerging* – University websites provide information about university policies, governance, regulations, and the types of university services that are provided. Links to academic units, administrative units and other universities are present. Users can obtain updated information about the

[1] https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2014

university and its dependencies, and they can click on links to access information that has been filed.

2) *Enhanced* – University websites deliver on-line, one-way, improved or on-line services, simple, two-way communication between the university and the members of its community is feasible, such as downloadable forms for university services and applications. Websites include audio and video resources, and offer information in more than one language. Some limited electronic services allow the members of the university community to send requests for non-digital forms or personal information.
3) *Transactional* – University websites involve the members of the university community in two-way communication, including requesting and receiving information on university policies, academic programs, regulations, etc. Some form of electronic authentication is required to verify the identity of the person for successfully completing the exchange. University websites process non-financial transactions, such as filling in enrollment forms, requesting certificates, and applying for any kind of permits. Additionally, financial transactions are also processed, such as paying some university fees through a secure network.
4) *Integrated* – University websites change the way in which the university communicates with the members of its community. They are proactive in requesting information and feedback from the members of the university community using Web 2.0 tools or other interactive tools. Electronic services seamlessly cross over academic and administrative units; information, data and knowledge are transferred between dependencies through integrated applications. Universities have transitioned from an "institution-centric" approach to a "user-centric" approach, where electronic services are targeted to the members of the university community through academic-cycle events – such as enrolment at the university, employed by the university, etc.: and segmented groups of recipients to provide customized services.

3. Related Work

In [4], the state of the art about e-government and its application to the university context is described. The study is based on the analysis of scientific publications included in the Scopus database. To this end, a search was done using the following keywords - "e-government," "electronic government," and "digital government." The results involve publications including such words in their title, abstract, or key words, limited to articles published between 2010 and 2015. Results were analyzed taking into account distribution by year, country, and region in particular, and those that were related to Computer Science and software quality.

Based on the results of the conducted searches, it was observed that publications reached a peak in years 2010-2011, with a bit more than 1,000 publications a year, and then decreased uniformly up to 2015, which has just above 600 publications. The countries with the highest number of publications on the topic are China and the United States; while in Latin America, Brazil is leading the list. From all publications, 63.9% are produced by Computer Science researchers, which shows that there is a strong relation between the concept of e-government and automated IT services.

When searching in quality models and assessment models, only 51 publications were found discussing the relation between e-government and quality models, and 33 establishing a relation between e-government and assessment models. Only 20 of these publications refer to Computer Science topics.

All the models proposed by the different authors [5]–[11] define the assessment feature based on different groupings of the classic McCall criteria [12]. Only [9] and [12] make reference to the use of ISO standards as quality model, and [13], [14] mention other pre-existing models. In [15], the authors propose the use of fuzzy logic for the assessment process. Eight of the publications [5]–[9], [13], [14], [16] mention that the main objective of quality is gaining citizen satisfaction, and some authors even describe the most relevant concepts to achieve this objective. The main target when analyzing the application of quality models have been government entities, such as public entities [15][17][18]. Interestingly, there are no case studies involving the university context.

Our findings show that there is scarce research work related to e-government applied to universities. This poses an important challenge when carrying out an assessment study on university services quality and maturity.

4. EGOV-U Services

In this section we discuss the methodology used to assess the maturity level of university services. First, the services provided by an academic unit (AU) to each type of users are identified. Then, the identified services are analyzed and explained in details. Finally, the application of the methodology at the School of Computer Science of the UNLP is discussed.

4.1 Service Identification

The university AUs provide a significant number of services to the various members of the university community. To assess the maturity level of these services, for each of the potential type of users, we identify a subset of the services that their provision is perceived as an essential goal of the AU. For instance, the identified services for students include providing information about the academic calendar, enrolling students into the various courses, and

choosing student representatives for the AU. Figure 1 shows the full list of services identified for each type of user.

4.2 Service Analysis

The steps followed to carry out the analysis include: 1) service identification, 2) service description, 3) service documentation, 4) quality assessment, and 5) maturity level assessment. Each stage is defined below.

Service Identification
In the context of university services, an AU is considered to be a university-dependent unit (university, school, high school, department, etc.) that has autonomy to regulate its own policies and the services that will deliver.
The goal of this activity is identifying the services that each AU offers to the different types of members in its community. Each AU must select, from a list of services, those that it is currently providing and those that are considered to be relevant for its community.

Service Description
AUs provide a brief description of each of the services selected from the list in the previous step, and the name of the person responsible for delivering the service. The information provided in this stage allows identifying the services to be analyzed for each AU and the individuals that will be able to provide more detailed information about each service. Table 1 shows the description of services 1-2 and 8 provided to students by the School of Computer Science of the UNLP.

Service Documentation
For each of the services described by the AU, the individual responsible for the service indicates all delivery channels used for the service. If the delivery process has some level of automation, the software application used for this is mentioned. If the process is manual, the type of support used, if any, is mentioned. Additionally, any regulations applicable to the provision of the service are listed.

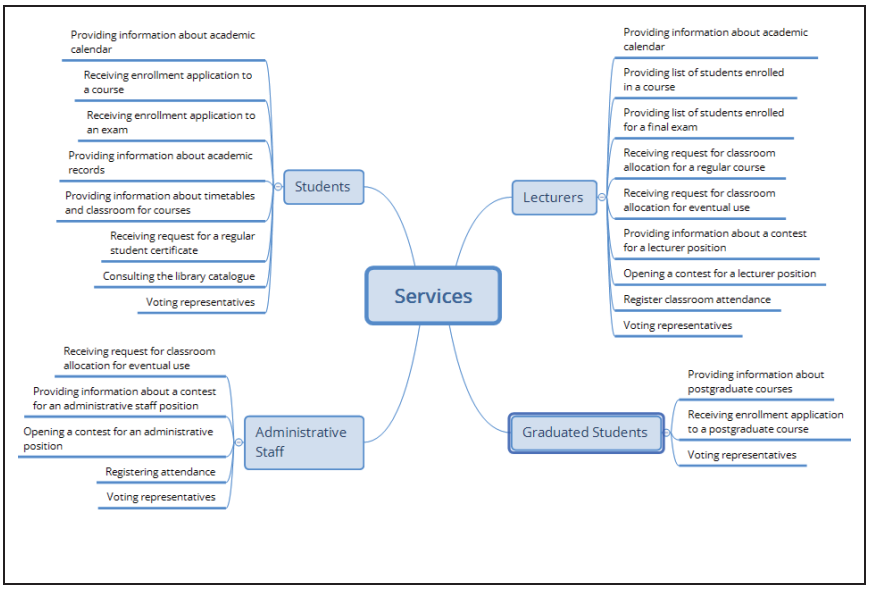


Fig. 1. Services identified for each type of members of the university community

Table 1. Example of service description

	Service	Description	Responsible
1	Providing information about academic calendar	The AU posts academic calendar information for students, with information about the academic year approved by the Honorable Board of Directors (HBD).	Academic Office
2	Receiving enrollment application to a course	The AU receives a student enrollment application for any given subject, verifies if the student meets requirements, and accepts or rejects the application. The AU informs the student the result of this process.	Students Office
8	Voting for representatives	The AU provides information about election periods and list of candidates. It organizes the electoral proceedings in accordance to regulations and publishes the results.	Electoral Board

The information provided in this stage allows establishing the maximum level of automation achievable by the service being considered, which may or may not be equal to the current level of automation used to provide the service. In some cases, the level of automation main be mainly constrained by compliance with legal regulations applicable to the provision of the service. For instance, Table 2 shows the documentation collected for the service "Providing information about the academic calendar."

Table 2. Service Documentation – Providing information about the academic calendar

1 – Providing information about the academic calendar		
Description	The AU posts academic calendar information for students, with information about the academic year approved by the Honorable Board of Directors (HBD).	
Regulation	Academic Calendar	
Channel	**Means of Access**	**Description**
AU	Notice board	
Web	http://www.info.unlp.edu.ar/	A detailed calendar is posted
e-mail	No	
IM	No	
APP	InfoAPP	Using the app, students can check the academic calendar
Facebook	Facultad de Informática UNLP	Published at the start of the year
Twitter	@informaticaUNLP	Published at the start of the year

Quality Assessment of the Service Delivery

The user of the service fills in a form to assess service delivery, indicating the channel used to receive the service, if the service was delivered immediately upon request or if there was a pending response after the first contact and, if so, how long it took for receiving the response, and the user's general level of satisfaction with the quality of the received service. Table 3 shows a form to assess the quality of service delivery.

Establishing Service Maturity Level

Based on the information gathered in the previous steps, this stage focuses on establishing both the current and the desired (maximum) maturity level for each service, considering the level of automation achieved in all delivery channels. Levels are determined based on the classification discussed in Section 2. For example, Table 4 illustrates the maturity level established for three services.

4.3 Applying the Methodology

The methodology proposed was applied at the School of Computer Science of the UNLP. A total of 25 services were assessed - 8 for students, 3 for graduated students, 9 for lecturers, and 5 for administrative staff. Based on the analysis carried out, 15 services could achieve a maximum level of *Enhanced* and 10 could achieve the *Transactional* level. After assessing the services, it was concluded that 17 of them had achieved their corresponding

maximum level, 6 were at the *Enhanced* level with a desired level of *Transactional*; 1 was at the Emerging level with a desired level of *Enhanced*, and 1 that did not use electronic channels for service delivery and has a desired level of *Enhanced*.

1 – Information about the academic calendar							
Academic Unit: School of Computer Science - UNLP	**Recipients**: students						
1 -How did you find out information about the academic calendar?	At AU	Web	Mail	IM	APP	Fb	Tw
1 .1 Which was the most useful channel for receiving information about the academic calendar?	At AU	Web	Mail	IM	APP	Fb	Tw
2 – How satisfied or dissatisfied are you with the selected options?	**1 – 2 – 3 – 4 – 5** 1 Bad 2 Fair 3 Good 4 Very good 5 Excellent						
3 – Was the service delivered immediately (the answer was received automatically)?	**Yes - No**						
3.1 – In case the service was not delivered immediately, please indicate how much time it took the receive the required information and the channel used	Delay: ___**Hs** or ___ **Days**						
	At AU	Web	Mail	IM	APP	Fb	Tw
4 – Overall, please, classify your level of satisfaction with the received service	**1 – 2 – 3 – 4 – 5** 1 Bad 2 Fair 3 Good 4 Very good 5 Excellent						

Table 4. Establishing Service Maturity Level

	Service	Desired Level	Current Level
1	Providing information about the academic calendar	**Enhanced** The goal of the service is providing updated information through various channels. No transactions are required for this service.	**Enhanced** It provides updated information through different channels, information can be downloaded, e-mail contact is available, and a search service is available.
2	Receiving enrollment application to a course	**Transactional** The goal of this service is keeping records of all applications received and analyzing them to either accept or reject them.	**Transactional** The system allows a student to enroll in a course, based on course prerequisites. Acceptance or rejection is automated.

5. Conclusions

EGOV-U applies e-government concepts to the university scope. In this paper, we have introduced concepts related to EGOV-U and a methodology that allows assessing the maturity level of the services provided by universities, in the context of EGOV-U. The methodology was applied at the School of Computer Science of the UNLP. For the analysis, 25 services considered as essential for the School were selected. The AUs documented their services based on the applicable regulations, defined the maximum automation level that each service can achieve, and identified the actual maturity level (current) of each of these services.

As a whole, 60% of the services provided by the School could achieve a maximum level of *Enhanced*, and 52% of the services were assessed to be at that level. Forty percent of the service had a maximum achievable level of *Transactional*, and only 40% of the identified services achieved that level. The remaining services are at a maturity level that is below the desired one. Future work includes analyzing the services provided by administrative units and studying innovation mechanisms for providing university services.

References

[1] S. de P. U. SPU, "Síntesis de Información Estadísticas Universitarias Argentina." 2015.

[2] Republica Argentina, "Ley Nro 24521 - Ley de Educacion Superior," 1995.

[3] UNLP, "Estatuto de la UNLP," 2008.

[4] A. Pasini and P. Pesado, "Quality Model for e-Government Processes at the University Level: a Literature Review," *Proc. 9th Int. Conf. Theory Pract. Electron. Gov.*, pp. 436–439, 2016.

[5] D. Bhattacharya, U. Gulla, and M. P. Gupta, "E-service quality model for Indian government portals: citizens' perspective," *J. Enterp. Inf. Manag.*, vol. 25, no. 3, pp. 246–271, 2012.

[6] X. Papadomichelaki and G. Mentzas, "Analysing e-government service quality in Greece," *Electron. Gov. an Int. J.*, vol. 8, no. 4, p. 290, 2011.

[7] S. Funilkul, W. Chutimaskul, and V. Chongsuphajaisiddhi, *Electronic Government and the Information Systems Perspective*, vol. 6866. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[8] P. Saha, A. Nath, and E. Salehi-Sangari, *Electronic Government*, vol. 6228. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[9] M. A. Alanezi, A. K. Mahmood, and S. Basri, "A proposed model for assessing e-government service quality: An E-S-QUAL approach," in *2012 International Conference on Computer & Information Science (ICCIS)*, 2012, vol. 1, pp. 130–135.

[10] A. Sivaji, N. Abdollah, S. S. Tzuaan, C. N. Khean, Z. M. Nor, S. H. Rasidi, and Y. S. Wai, "Measuring public value UX-based on ISO/IEC 25010 quality attributes: Case study on e-Government website," in *2014 3rd*

International Conference on User Science and Engineering (i-USEr), 2014, pp. 56–61.

[11] E. Ziemba, T. Papaj, and D. Descours, "Assessing the quality of e-government portals – the Polish experience," in *2014 Federated Conference on Computer Science and Information Systems, FedCSIS 2014*, 2014, pp. 1259–1267.

[12] J. a. McCall, P. K. Richards, and G. F. Walters, "Factors in Software Quality," *Nat'l Tech. Inf. Serv.*, vol. 1, 2 and 3, no. ADA049055, 1977.

[13] S. C. Misra and J. Chatterjee, "Applying Gap Model for Bringing Effectiveness to e-Government Services:," *Int. J. Electron. Gov. Res.*, vol. 9, no. 3, pp. 43–57, Jan. 2013.

[14] S. W. Liang, H. P. Lu, and T. K. Kuo, "A study on using the kano two-dimensional quality model to evaluate the service quality of government websites," *J. Internet Technol.*, vol. 15, no. 2, pp. 149–162, 2014.

[15] G. Yucel and A. F. Ozok, "Quantifying ergonomic quality of governmental websites," *Electron. Gov. an Int. J.*, vol. 7, no. 3, p. 233, 2010.

[16] F. Sa, A. Rocha, and M. P. Cota, "Quality models for online e-Government services," in *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*, 2014, pp. 1–5.

[17] E. Loukis, K. Pazalos, and A. Salagara, "Transforming e-services evaluation data into business analytics using value models," *Electron. Commer. Res. Appl.*, vol. 11, no. 2, pp. 129–141, Mar. 2012.

[18] E. Loukis, K. Pazalos, and A. Salagara, "Structuring e-services evaluation based on multi-level value flow models," in *Proceedings of the European, Mediterranean and Middle Eastern Conference on Information Systems - Informing Responsible Management: Sustainability in Emerging Economies, EMCIS 2011*, 2011, pp. 552–570.

XIII

Database and Data Mining Workshop

Discovery Process of Co-Localization Patterns around Reference Event Types

GIOVANNI DAIÁN ROTTOLI[1,2,3], HERNÁN MERLINO[3], RAMÓN GARCÍA-MARTINEZ[3,4]

[1] PhD Program on Computer Sciences. National University of La Plata. Argentina.
[2] PhD Scholarship Program. National University of Technology. Argentina
[3] Information Systems Research Group. National University of Lanús. Argentina
[4] Scierntific Researchs Comission of Buenos Aires - CIC. Argentina
gd.rottoli@gmail.com,hmerlino@gmail.com,rgm1960@yahoo.com

Abstract. The discovery of co-localization patterns reveals subsets of types of spatial events whose instances occur frequently adjacent to each other. Many algorithms and methods have been developed over the years for this purpose. However, when it is required to find these patterns around particular types of spatial events, the existing alternatives are incomplete and incorrect. In our work, a process of spatial information mining is presented for the discovery of co-localization patterns around types of spatial reference events that uses maximum clicks and TDIDT algorithms to provide a solution to this problem. A proof of concept of the proposed process is presented.

Keywords. Co-Localization Patterns, TDIDT, Maximum Clicks, Spatial Information Mining.

1. Introduction

Given a set of Boolean spatial event types and a neighborhood relationship, the discovery of co-localization patterns allows us to find subsets of these types of events whose instances are often located adjacent to each other [Shekhar & Huang, 2001]. In this context, a spatial event is an event that occurs in a given space and, a spatial event type makes reference to the class of the occurring event. A spatial event type is also called boolean spatial feature.

Many algorithms and methods have been proposed for co-location pattern discovery based on association analysis. These algorithms generate transactional data from spatial objects neighborhoods and, based on that, they can be categorized into two classes: (i) transaction-free algorithms, which exploit the association analysis algorithm internally, e.g., the Aprori-like algorithms [Agrawal & Srikant, 1994], but none of them generates or uses a transaction-type dataset externally; and (ii), transaction-based algorithms, which exploit association analysis methods after explicitly generating a

transaction-type dataset. This last approach result more efficient than the first one[Shekhar & Huang, 2001; Shekhar et al., 2011, Kim et al., 2014].
In both options, it is necessary to choose a model to generate the transactional data. This model describes the way in which the transactions are created. There are three different approaches [Shekhar & Huang, 2001; Xiong et al., 2004]. The first way is called Event-Centric Model and is used when there are many types of spatial event types, like in ecology, to find subsets of spatial event types likely to occur in a neighborhood around instances of given subsets of event types. This way was used in works such as [Agrawal & Srikant, 1994; Shekhar & Hung, 2001; Huang et al., 2003, 2006; Yoo et al., 2004; Xiong et al., 2004; Yoo & Shekhar, 2006; Celik et al., 2007; Eick, 2008; Adilmagambetov et al., 2013; Kim et al., 2011, 2014].
The second way is called Window-Centric Model and serves to discover patterns into subdivisions of the given data space, called windows. This model is used in areas such as mining, where each window corresponds to land parcels.
For last, the third model is called Reference Based Model, and serves to find co-location patterns generating transactions around spatial events of a given type. In this case, the transactions are created by "materializing" the neighborhood of the instances of the reference spatial feature.
It is important to say mention that the reference based model has not been implemented in too many algorithms, and those algorithms that implement it gives incorrect and incomplete results [Adilmagambetov et al., 2013; Kim et al., 2014].
The remainder of this paper is organized as follows: In Section 2, a problem derived from the analysis of the state-of-the-art is presented. In Section 3, we present a Knowledge Discovery Process to solve that problem. In Section 4, experimental results are presented. Finally, conclusions derived from the research are outlined in Section 5.

2. Problem Definition

When facing a co-location discovery problem, if there are many boolean spatial features to be considered for co-location pattern discovery, the Event-Centric Model may be expensive in terms of time and resources.
With the presence of a spatial feature that is interesting in a particular problem domain, using a Reference Based Model is a more suitable alternative. This model determines the neighborhoods in a special manner: first, a reference feature is selected and then, for each instance object of that feature, all spatial objects located within a pre-specified distance are selected, and transaction-type data generated [Shekhar & Huang, 2001; Xiong et al., 2004].
This approach, however, cannot be used to generate correct or complete transactions, as it does not ensure that all objects in the transaction are neighbors; moreover, some neighborhoods may be lost [Adilmagambetov et al., 2013; Kim et al., 2014].
For this reason, it is necessary to develop a solution that serves to discover correct and complete spatial co-location patterns around reference features.

In this work, we develop a Knowledge Discovery Process [Britos, 2008] to give a solution to this problem using an Event-Centric Model to generate transaction-based data, and induction of decision trees to generate co-location rules.

3. Proposed Solution

As mentioned before, this paper proposes a knowledge discovery process to find co-location relations between spatial features. This process serves to find correct relations around reference features without using a reference feature-centric model to generate transactional data.

This work is based on the work of Kim et al. (2014), proposing a transactional framework that uses an event-centric model to find co-location patterns using maximal cliques as a way to generate complete and correct transactions.

Given a neighboring graph, a clique is a complete sub-graph, which means that all its nodes are neighbors with each other. A maximal clique, in consequence, is a clique that is not included in any other clique. Each maximal claque corresponds to a transaction where all its elements are neighbours, ensuring that the transactions are correct [Kim et al., 2014, Lemma 1].

On the other hand, the utilization of maximal cliques as transactions ensures the completeness of the transactions, because all the relations are considered in at least one maximal clique [Kim et al., 2014, Lemma 2].

For this reason, a knowledge discovery process for co-location pattern discovery is proposed, focused on reference features, that uses an event-centric model for transaction-based data generation through spatial maximal cliques and using a Process of Discovery of Behavior Rules using TDIDT algorithms [Britos, 2008].

As is shown in Figure 1, the process takes a set of spatially referenced data as input, represented in different formats such as *inter alia*, plain text, databases, tables and geographic information system maps. These data are integrated to a table comprised of the object identifier, the spatial feature and the object location.

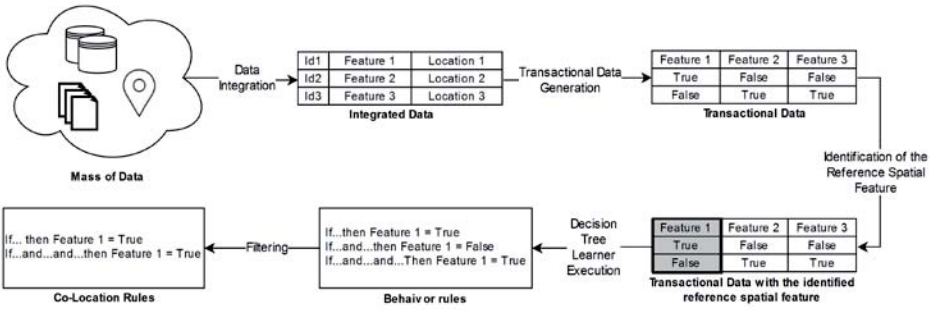


Fig. 1. Proposed co-location rules discovery process

Then, the integrated data are used to generate the transactional dataset. In this sub-process, as shown in Figure 2, all the neighbor relationships are calculated by evaluating the distance between the spatial objects. Afterwards, finding all the maximal spatial cliques inside then neighboring graph is required to generate a transaction for each, in which the spatial features of each spatial object from that clique are presented.

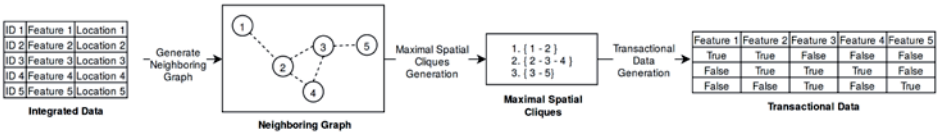


Fig. 2. Sub-process to generate transactional data

Once the transactional data is obtained, the reference spatial feature must be specified to find the co-location relations around it. That spatial feature will be used as the target attribute of a TDIDT algorithm, using the rest of the attributes as input.
A set of rules will be obtained from the generated decision tree in the last step as output. Due to the fact that the transactions have boolean values that show the presence or absence of the spatial features in the neighborhoods, it is necessary to filter the rules that show the presence of the reference spatial feature in the consequent.
This process gives the possibility of reusing the transactional data to discover co-location patterns around many spatial features without the need to calculate the neighboring graph on each opportunity. To do this, it is enough just to select another spatial feature as target and execute the decision tree learner again.
On the other hand, the rules obtained not only describe the spatial features that are usually neighbors, but also the conditions required in the neighborhoods for the presence of these relations, because each rule also shows whether the absence of any features is necessary, thus adding information to the results.

4. Proof of Concept

Below is a proof of concept of the Knowledge Discovery Process proposed in the last section. This proof aims to compare the proposed process with a Reference Feature-Centric Model Algorithm to determine if our solution can find more correct and complete co-location relations than the alternative.
For that purpose, 10 synthetic sets of 500 points are automatically generated and classified in 7 types, with random location in a small 2D space, as shown in Figure 3.

The synthetic sets are used as input for our proposed process and for the selected algorithm: Co-Location Miner with a Reference Based Model [Shekhar & Huang, 2001].
On the other hand, the symmetry property of the distance function has been used to calculate the neighboring graph in the proposed process in order to reduce the number of times this operation is performed. The algorithm CLIQUES has been used for the generation of maximal spatial cliques because of its superior efficiency over other methods [Uno, 2005; Tomita et al., 2006]. The software Tanagra [Rakotomalala, 2005] was used to run the selected Decision Tree Learning Algorithm C4.5 [Quinlan, 1993].
After the execution of both methods, the co-location relationships obtained were evaluated to corroborate their correctness in order to determine how many correct relationships were found with each method.
To show that the proposed process can find a greater number of co-location relationships, a non-parametric statistical hypothesis test was used: the Wilcoxon signed-rank test [Wilcoxon, 1945], in order to reject the null hypothesis H_0 and accept the alternative hypothesis H_A. These hypotheses are shown in Table 1.

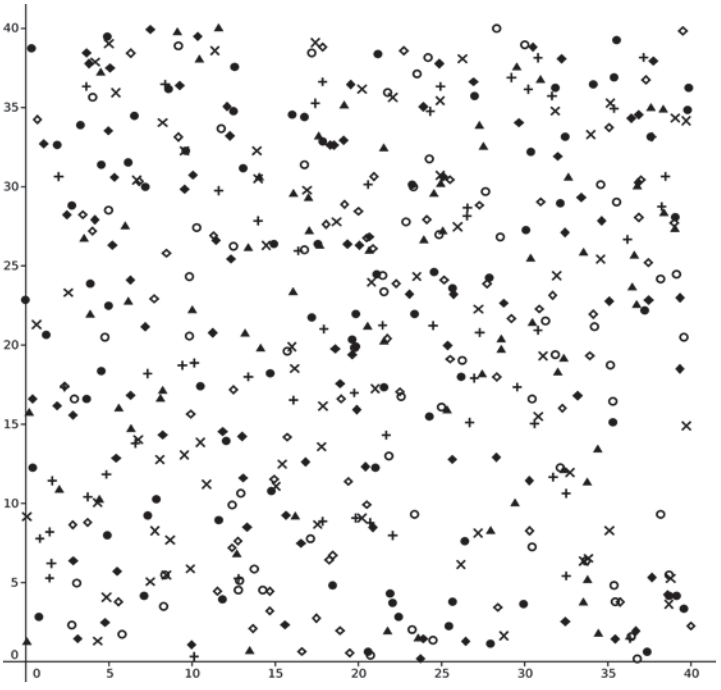


Fig. 3. First synthetic dataset distribution. Each symbol corresponds to a different spatial feature

Table 1. Null hypotesis and alternative hypotesis considered in the Wilcoxon signed-rank test

H_0:	The number of correct co-location relationships discovered using the Co-Location Miner Algorithm is greater than or equal to the number of correct co-location relationships discovered using the proposed process.
H_A:	The number of correct co-location relationships discovered using the proposed process is greater than the number of correct co-location relationships discovered using the Co-location Miner Algorithm.

The Wilcoxon Test execution is shown in Table 2 with the values obtained in each experiment. The critical value of W for 9 examples with a significance level of 1% is 1. The yielded W-Value is equals to 0. For this reason, the null hypothesis was rejected, confirming that the knowledge discovery process proposed in this paper serves to find a greater number of correct co-location relationships that the method using a reference feature-centric model.

Table 2. Wilcoxon signed-rank test execution, sorted by the absolute value of the differences.

Set	Correct relationships discovered with the proposed process	Correct relationships discovered with co-location miner algorithm.	Absolute value of the differences	Rank
Set 7	3	3	0	-
Set 8	2	1	1	2
Set 9	2	1	1	2
Set 10	5	4	1	2
Set 1	3	1	2	5
Set 2	3	1	2	5
Set 5	4	2	2	5
Set 4	4	1	3	7
Set 3	6	2	4	8.5
Set 6	7	3	4	8.5
Sum				45

5. Conclusion

This paper has described a knowledge discovery process that can be used to find correct and complete co-location patterns around reference spatial features. This process uses an event-centric model through maximal spatial cliques in order to generate transactional data from neighboring relationship between spatial data, and a decision tree learning algorithm, to obtain behavior rules that describe the neighborhoods that contain the spatial reference feature, an innovative method to achieve this goal.

The proof of concept, by means of a non-parametrical statistical test, shows that the proposed process finds a greater number of correct co-location patterns than the methods that use a reference feature-centric model to generate transactional data.

The proposed method also serves (i) to search co-location relationships between many reference spatial features without the need to calculate the neighboring relationships on each opportunity, and (ii), to add information to the results, showing what the conditions for the occurrence of the co-location relationships are.
The next planned step is to conduct validation proofs in the fields of accident prevention, civil defense and environmental determinants of diseases.

6. Acknowledgements

The research presented in this paper was partially funded by the PhD Scholarship Program to reinforce R+D+I areas (2016-2020) of the Technological National University, Research Project 80020160400001LA of National University of Lanús, and PIO CONICET-UNLa 22420160100032CO of National Research Council of Science and Technology (CONICET), Argentina.

References

1. Adilmagambetov, A., Zaiane, O. R., & Osornio-Vargas, A. (2013). Discovering co-location patterns in datasets with extended spatial objects. In Data Warehousing and Knowledge Discovery (pp. 84-96). Springer Berlin Heidelberg.
2. Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (Vol. 1215, pp. 487-499).
3. Britos, P. V. (2008). Procesos de explotación de información basados en sistemas inteligentes. Tesis de Doctorado en Ciencias Informáticas. Facultad de Informática. Universidad Nacional de La Plata.
4. Celik, M., Kang, J. M., & Shekhar, S. (2007, October). Zonal co-location pattern discovery with dynamic parameters. In Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on (pp. 433-438). IEEE.
5. Chicago Police Department (2016). Reported Incidents occurred in the City of Chicago from 2001 to present [On-Line]. Chicago, USA. [Consultado el 3 de Febrero de 2016]. Disponible en: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2
6. Eick, C. F., Parmar, R., Ding, W., Stepinski, T. F., & Nicot, J. P. (2008, November). Finding regional co-location patterns for sets of continuous variables in spatial datasets. In Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems (p. 30). ACM.
7. Huang, Y., Xiong, H., Shekhar, S., & Pei, J. (2003, March). Mining confident co-location rules without a support threshold. In Proceedings of the 2003 ACM symposium on Applied computing (pp. 497-501). ACM.
8. Huang, Y., Pei, J., & Xiong, H. (2006). Mining co-location patterns with rare events from spatial data sets. Geoinformatica, 10(3), 239-260.

9. Kim, S. K., Kim, Y., & Kim, U. (2011). Maximal cliques generating algorithm for spatial co-location pattern mining. In Secure and Trust Computing, Data Management and Applications (pp. 241-250). Springer Berlin Heidelberg.
10. Kim, S. K., Lee, J. H., Ryu, K. H., & Kim, U. (2014). A framework of spatial co-location pattern mining for ubiquitous GIS. Multimedia tools and applications, 71(1), 199-218.
11. Quinlan, J. R. (1993). C4. 5: programs for machine learning.
12. Rakotomalala, R. (2005). TANAGRA: a free software for research and academic purposes. In Proceedings of EGC (Vol. 2, pp. 697-702).
13. Shekhar, S., & Huang, Y. (2001). Discovering spatial co-location patterns: A summary of results. In Advances in Spatial and Temporal Databases (pp. 236-256). Springer Berlin Heidelberg.
14. Shekhar, S., Evans, M. R., Kang, J. M., & Mohan, P. (2011). Identifying patterns in spatial information: A survey of methods. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(3), 193-214.
15. Tomita, E., Tanaka, A., & Takahashi, H. (2006). The worst-case time complexity for generating all maximal cliques and computational experiments. Theoretical Computer Science, 363(1), 28-42.
16. Uno, T. (2005). MACE_GO: MAximal Clique Enumerator (CLIQUES Implementation) [C Code]. Versión 2.0.. Disponible desde: http://research.nii.ac.jp/~uno/code/ macego10.zip
17. Venkatesan, M., Thangavelu, A., & Prabhavathy, P. (2011). Event Centric Modeling Approach in Colocation Pattern Snalysis from Spatial Data. arXiv preprint arXiv:1109.1144.
18. Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics bulletin, 1(6), 80-83.
19. Xiong, H., Shekhar, S., Huang, Y., Kumar, V., Ma, X., & Yoo, J. S. (2004, April). A Framework for Discovering Co-Location Patterns in Data Sets with Extended Spatial Objects. In SDM (pp. 78-89).
20. Yoo, J. S., Shekhar, S., Smith, J., & Kumquat, J. P. (2004, November). A partial join approach for mining co-location patterns. In Proceedings of the 12th annual ACM international workshop on Geographic information systems (pp. 241-249). ACM.
21. Yoo, J. S., Shekhar, S., & Celik, M. (2005, November). A join-less approach for co-location pattern mining: A summary of results. In Data Mining, Fifth IEEE International Conference on (pp. 4-pp). IEEE.
22. Yoo, J. S., & Shekhar, S. (2006). A joinless approach for mining spatial colocation patterns. Knowledge and Data Engineering, IEEE Transactions on, 18(10), 1323-1337.

LSA64: An Argentinian Sign Language Dataset

FRANCO RONCHETTI[*1], FACUNDO QUIROGA[*1], CESAR ESTREBOU[1], LAURA LANZARINI[1], ALEJANDRO ROSETE[2].

[1] Instituto de Investigación en Informática III-LIDI, Facultad de Informática, Universidad Nacional de La Plata
{fronchetti,fquiroga,cesarest,laural}@lidi.unlp.edu.ar
[2] Instituto Superior Politecnico Jose Antonio Echeverría
{rosete}@ceis.cujae.edu.cu

Abstract. Automatic sign language recognition is a research area that encompasses human-computer interaction, computer vision and machine learning. Robust automatic recognition of sign language could assist in the translation process and the integration of hearing-impaired people, as well as the teaching of sign language to the hearing population.Sign languages differ significantly in different countries and even regions, and their syntax and semantics are different as well from those of written languages. While the techniques for automatic sign language recognition are mostly the same for different languages, training a recognition system for a new language requires having an entire dataset for that language. This paper presents a dataset of 64 signs from the Argentinian Sign Language (LSA). The dataset, called LSA64, contains 3200 videos of 64 different LSA signs recorded by 10 subjects, and is a first step towards building a comprehensive research-level dataset of Argentinian signs, specifically tailored to sign language recognition or other machine learning tasks. The subjects that performed the signs wore colored gloves to ease the hand tracking and segmentation steps, allowing experiments on the dataset to focus specifically on the recognition of signs. We also present a pre-processed version of the dataset, from which we computed statistics of movement, position and handshape of the signs.

Keywords: sign language recognition, handshape recognition, lexicon, corpus, automatic recognition.

1. Introduction

Sign language (SL) recognition is a complex multidisciplinary problem. It bears many similarities to speech recognition, but presents some additional difficulties [8]:

1. There is little formal standardization in most sign languages, even within a region.

[*] Contributed equally

2. Sign language specification languages themselves are not well standardized, and there is no consensus on which type of specification is more appropriate.
3. Signs are intrinsically multimodal: they are formed by a combination of hand shapes, positions, movements, body pose, face expression, and lip movements. In contrast, speech recognition usually requires only sound input.
4. Creating datasets for training sign language recognition systems requires being able to capture and model all of these signals.
5. There are relatively few sign language users with respect to the general population, and therefore finding expert subjects to record signs is more difficult.

For these reasons, it is generally considered that we are still far from robust sign language recognition systems.

There are numerous publications dealing with the automatic recognition of sign languages, and [1,8] present reviews of the state of the art in sign language recognition. The full task of recognizing a sign language involves a multi-step process. In the context of video-based recognition, and considering only manual information, this process can be simplified as:

1. Tracking and segmenting the hands of the interpreter in every frame of the video.
2. Recognizing the shapes of the hands, the movements they made and their positions.
3. Recognizing the sign as a syntactic entity (a visual word).
4. Assigning semantics to a sequence of signs (a visual sentence).
5. Translating the semantics of the signs to the written language.

These tasks are mostly independent from each other, and involve different techniques. Tracking, segmentation and modeling of the hand are mostly signal processing and 3D modeling tasks, while assigning semantics to a sequence of signs and translating from sign to written languages are more related to natural language processing.

1.1 Datasets

There are many datasets for sign language recognition. We distinguish three types, depending on the problem they target: handshape recognition, sign recognition or sentence recognition. Each type of dataset presents a greater challenge than the previous one, and allows experiments with more steps of the recognition pipeline.

Table 1 presents the most prominent video-based, research-level datasets for recognition. Since the dataset described in this paper focuses on sign-level recognition, we only list sign and sentence-level datasets[1].

[1] A more detailed reference about sign language datasets can be found at http:// facundoq.github.io/unlp/sign_language_datasets/index.html

Table 1. Recognition-oriented, video based, sign language datasets used in recent papers.

Name	*Classes*	*Subjects*	*Samples*	*Language level*	*Availability*
DGS Kinect 40 [1]	*40*	*15*	*3000*	*Word*	*Contact Author*
DGS RWTH-Weather [8]	*1200*	*9*	*45760*	*Sentence*	*Public Website*
DGS SIGNUM [8]	*450*	*25*	*33210*	*Sentence*	*Contact Author*
GSL 20 [1]	*20*	*6*	*840*	*Word*	*Contact Author*
Boston ASL LVD [3]	*3300+*	*6*	*9800*	*Word*	*Public Website*
PSL Kinect 30 [2]	*30*	*1*	*300*	*Word*	*Public Website*
PSL ToF 84 [2]	*84*	*1*	*1680*	*Word*	*Public Website*

In general, the datasets that are video-based must rely on skin color tracking and segmentation, and are therefore are not robust for background variations or interpreter clothes, as well as hand-hand or hand-face occlusions [6]. Usually, to perform features extraction these datasets require the addition of morphological information as a subsequent step to the color filtering to identify the position and shape of the hand, which can be extracted using depth cameras or other sensors, but these limit the application of the methods with respect to using normal video cameras, given the availability of each type of devices.

The largest sign language dataset available (in terms of number of classes), the American Sign Language Lexicon Video Dataset (ASLLVD) [3], contains more than 3300 signs from the American Sign Language, but near-perfect hand tracking and segmentation on this dataset is difficult [3], making it hard to use it to evaluate a sign recognizer that focuses on the syntactic and semantic recognition. The situation is similar for the SIGNUM and RWTH-Weather datasets. Moreover, the dataset ASLLVD has only 6 subjects, and an average of 3 samples per class.

1.2 Argentinian Sign Language Dataset (LSA64)

Sign languages are different in each region of the world, and each has their own lexicon and group of signs. Thus, sign language recognition is a problem that needs to be tackled differently in each region, since new movements or handshapes or combinations thereof require new training data, and possibly involve new challenges that were not considered before [8,1].

To the best of our knowledge, there are no available datasets for the Argentinian Sign Language (LSA). There are only a few dictionaries that focus on teaching the language. Since they were recorded with this aim in mind, they have only one sample for each sign, low image quality, and poor annotations, thus making them unsuitable for training automatic recognition systems. There is need for a research-level dataset that represents the group of signs used in LSA.

This paper presents a sign dataset called LSA64. The dataset consists of 64 signs from the LSA, and was recorded with normal RGB cameras. It is publicly available[2], and we also provide a preprocessed version of the dataset to facilitate experiments and reproducibility.
The subjects wore colored gloves for the recording (single colored gloves, with different colors for each hand). This methodology allows researchers to easily bypass the tracking and segmentation steps, and focus on the subsequent steps of the recognition [9].
While the dataset has less classes than ASLLVD, RWTH-PHOENIX-Weather or SIGNUM, it has more samples and subjects than many other datasets (Table 1), and it is publicly available with a preprocessed version.
The document is organized as follows: Section 2 describes the LSA64 dataset and the recording conditions. Section 3 presents statistics and information of the signs recorded, to aid in the understanding of the dataset. Section 4 details an experiment carried out to establish a baseline on this dataset, and Section 5 presents the general conclusions.

2. Dataset

The sign database for the Argentinian Sign Language, created with the goal of producing a dictionary for LSA and training an automatic sign recognizer, includes 3200 videos where 10 non-expert subjects executed 5 repetitions of 64 different types of signs. Signs were selected among the most common in the LSA lexicon, and include both verbs and nouns. Some examples can be seen in Figure 1.

Fig. 1. Snapshots of six different signs of the LSA64 database. There are overlaps in positions and handshapes. The images on the left are from the first set of recordings.[2]

[2] The dataset and relevant information can be found at http:// facundoq.github.io/unlp/lsa64/.

2.1 Recording

The database was recorded in two sets. In the first one, 23 one-handed signs were recorded. The second added 41 signs, 22 two-handed and 19 one-handed. The final dataset then contains 42 one-handed signs and 22 two-handed ones.

The first recording was done in an outdoors environment, with natural lightning, while the second took place indoors, with artificial lightning (Figure 1). Subject 10 from the first recordings was unavailable for the second set of recordings, and was replaced by another subject. This change in no way diminishes the utility of the dataset, since the set of classes recorded in the first session is disjoint from the ones recorded in the second session.

In both sets of recordings, subjects wore black clothes and executed the signs standing or sitting, with a white wall as a background. To simplify the problem of hand segmentation within an image, subjects wore fluorescent-colored gloves. These substantially simplify the problem of recognizing the position of the hand and performing its segmentation, and remove all issues associated to skin color variations, while fully retaining the difficulty of recognizing the handshape. Additionally, each sign was executed imposing few constraints on the subjects to increase diversity and realism in the database. The camera employed was a Sony HDR-CX240. The tripod was placed 2m away from the wall at a height of 1.5m.

In the following subsections we show statistics and information of the signs to better understand the nature and challenges of the dataset. These statistics show that signs in this dataset possess significant overlap in terms of types of movements, initial and final positions and handshapes, producing non-trivial experiment settings to test new sign language recognition models. All the information has been computed from the pre-processed version of the dataset described in Section 3.

2.2 Handshapes

In Figures 2 and 3 we can observe the different handshapes of the right and left hand respectively for each class of sign. There is plenty of repetition between handshapes, although their 2D projection may be different depending on the rotation of the hand.

2.3 Positions

Figure 4 presents the mean initial and final positions for each hand, along with the covariance. While a few signs can be identified by their positions, they overlap significantly in most cases.

2.4 Trajectories

Figure 5 shows sample trajectories of each sign, as performed by subject 2. There is much overlap in movements for both one-handed (for example, signs 1, 5, 7, 13 and 19) and two-handed signs (for example, signs 31, 32 and 61).

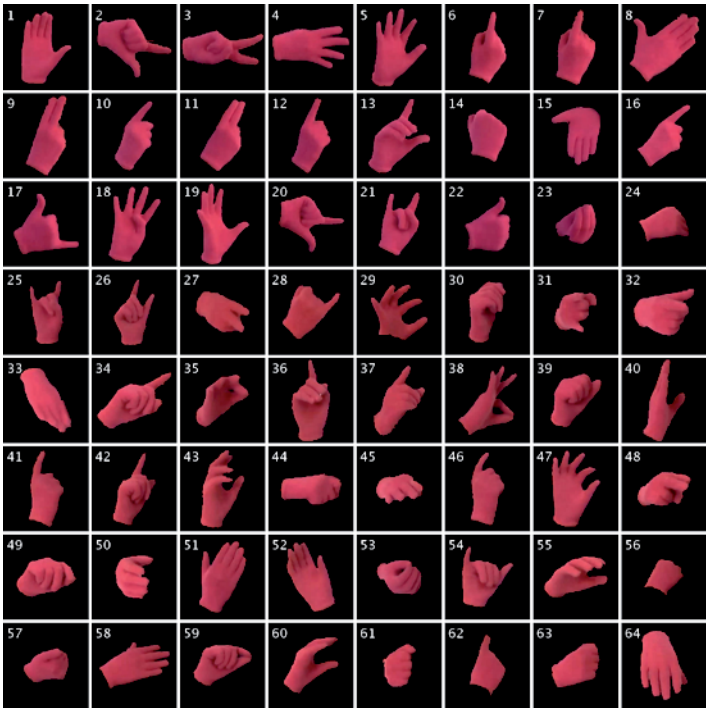


Fig. 2. Images of segmented hands as captured in the LSA64 database. Each image shows the initial handshape of the right hand for each sign in the dataset.

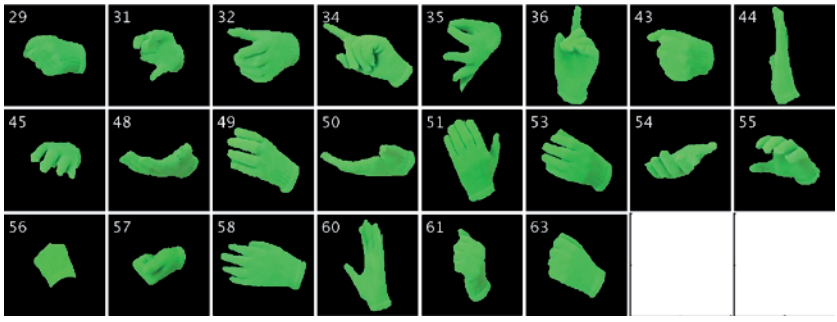


Fig. 3. Images of segmented hands as captured in the LSA64 database. Each image shows the initial handshape of the left hand for the two-handed signs of the dataset.

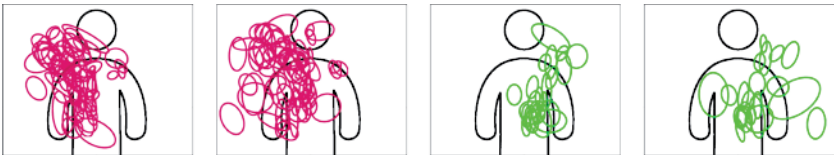


Fig. 4. Means for the initial and final positions of the right hand for each sign (a and b), and also for the left hand (c and d). Circles around means represent the covariance of the samples.

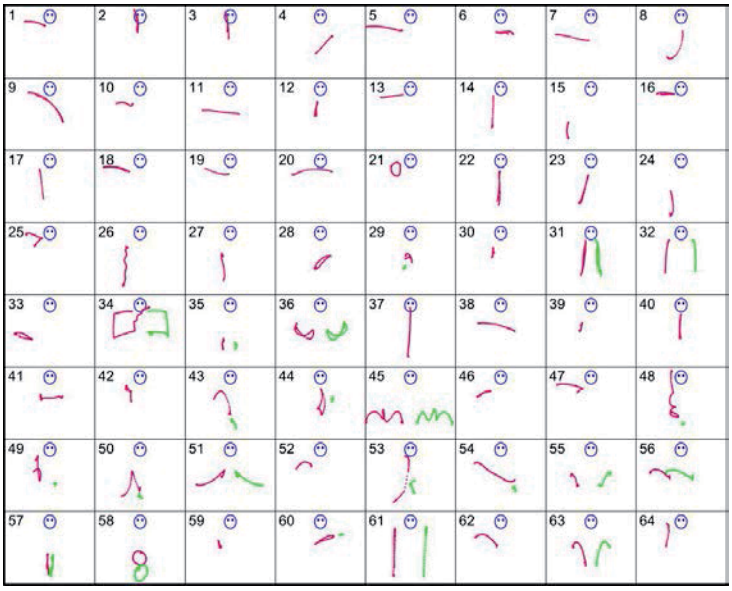


Fig. 5. Sample trayectories for each sign in LSA64. The left-hand trajectory is shown in light green, the right-hand one in red, and the head position as a blue circle.

2.5 Amount of movement

Figure 6 shows the amount of movement for each hand, measured as the maximum distance between two points in the trajectory of the hand. The movement in the left hand is significantly smaller than that of the right hand in many signs, consistent with the fact that the right hand is the dominant one for all the signers.

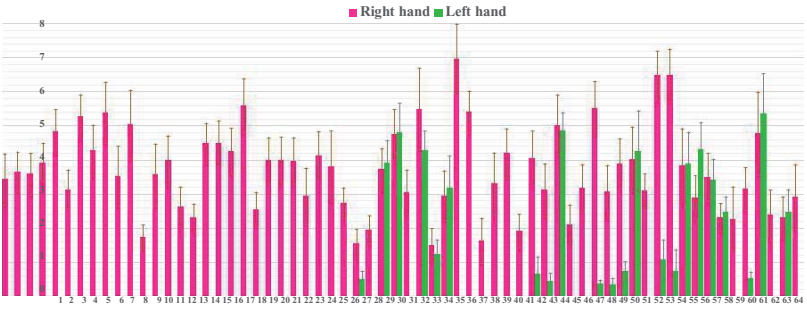


Fig. 6. Amount of movement for each class of sign. Red bars show the amount of movement of the right hand, and green bars the movement of the left hand (in two-handed signs).

3. Preprocessed version

We provide a pre-processed version of the dataset to alleviate the overhead of performing experiments with the data.

From the dataset we extracted the hand and head positions for each frame, along with images of each hand, segmented and with a black background, as show in Figure 7.

The tracking and segmentation of the hands uses the techniques described in [5]. Additionally, the head of the subject is tracked via the Viola-Jones's face detector [7]. The 2D position of each hand is translated so that the head is at the origin. The positions are then normalized by dividing by the arm's length of the subject, measured in centimeters/pixels. In this way, the transformed positions represent distances from the head, in units of centimeters.

The result of this process is a sequence of frame information, where for each frame we calculate the position of both hands, and we extract an image of each hand with the background segmented.

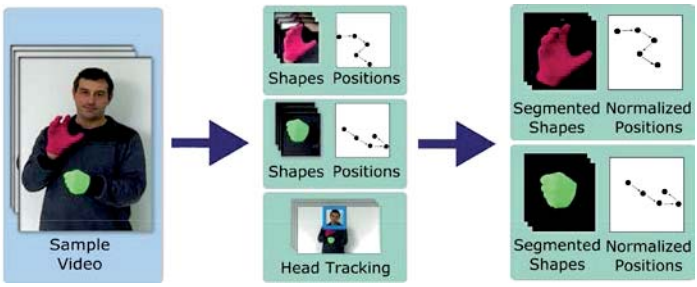


Fig. 7. Feature extraction steps for the preprocessed version of LSA64. From the sample video, we track the position of both hands and head in all frames. The shapes of each hand are segmented and the positions of them are re-centered with respect to those of the head.

4. Baseline Experiments

In this section, we briefly describe the model and results obtained in signer-dependent and independent experiments with the dataset, to establish a baseline performance. The model we used is described fully in [4].
The model we used to get a baseline performance on the dataset classifies the information for each hand separately and then multiplies the probabilities output by the subclassifiers, per class. The model for each hand includes three subclassifiers, each processing position, movement or handshape information.
The movement subclassifier contains one left-to-right Hidden Markov Model (HMM) per class, with skip transitions. All the models have 4 states. The output probabilities are modeled with a Gaussian Mixture Model (GMM) in each state. The models are trained with EM with the trajectories computed in the pre-processed version of the dataset.
The handshape subclassifier also employs HMM-GMMs, but uses as input the output of the static handshape classifier described in [5]. The static classifier is run on the segmented hand for each frame of the video, and the sequence of probabilities is fed into the handshape HMM-GMM to obtain the probability of each class for the whole sequence of frames.
The position subclassifier models the initial and final positions of the signs of each class with a set of gaussian distributions. There are two gaussian per class, one for the initial and another one for the final position.
For each class, the model outputs the product of the probabilities given by the position, movement and handshape information of both hands. For one-handed gestures, the information of the left hand is ignored, and so the probabilities output by the left hand model are not multiplied.
We performed a subject-dependent classification experiment with the model, using stratified repeated random sub-sampling validation as the cross-validation scheme, with 30 runs and an 80-20 training-test split. For each run, we measured the classification accuracy of the model on the test dataset. The mean accuracy obtained was 95.95% (standard deviation = 0:954).

5. Conclusion

We have presented a dataset of signs from the Argentinian Sign Language. To the best of our knowledge, there are currently no research-oriented datasets of this language created or available.
The subjects used colored gloves in the recording of the signs to significantly ease the tracking and segmentation steps. Nonetheless, we also provide a pre-processed version of the dataset to facilitate experimentation and reproducibility.
We have also presented a set of statistics and extra information to characterize the dataset and allow researchers to easily understand its nature. The signs in this dataset possess significant overlap in terms of types of

movements, initial and final positions and handshapes, producing non-trivial experiment settings to test new sign language recognition models.
We intend to expand the dataset with both new signs and a set of annotated LSA sentences to provide a complete basic working vocabulary for Argentinian Sign Language.

References

1. Cooper, H., Holt, B., Bowden, R.: Sign language recognition. In: Moeslund, T.B., Hilton, A., Kruger, V., Sigal, L. (eds.) Visual Analysis of Humans: Looking at People, chap. 27, pp. 539 { 562. Springer (Oct 2011)
2. Kapuscinski, T., Oszust, M., Wysocki, M., Warchol, D.: Recognition of hand gestures observed by depth cameras. International Journal of Advanced Robotic Systems 12 (2015)
3. Neidle, C., Thangali, A., Sclaro , S.: Challenges in development of the american sign language lexicon video dataset (asllvd) corpus. In: Proc. 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. Citeseer (2012)
4. Ronchetti, F., Quiroga, F., Estrebou, C., Lanzarini, L., Rosete-Suarez, A.: Sign languague recognition without frame-sequencing constraints: A proof of concept on the argentinian sign language. IBERAMIA: Iberoamerican Society of Arti cial Intelligence (2016)
5. Ronchetti, F., Quiroga, F., Lanzarini, L., Estrebou, C.: Handshape recognition for argentinian sign language using probsom. Journal of Computer Science and Technology 16(1), 1{5 (2016)
6. Roussos, A., Theodorakis, S., Pitsikalis, V., Maragos, P.: Hand tracking and affine shape-appearance handshape sub-units in continuous sign language recognition. In: Trends and Topics in Computer Vision - ECCV 2010 Workshops, Heraklion, Crete, Greece, Revised Selected Papers, PartI. pp. 258{272 (2010)
7. Viola, P., Jones, M.J.: Robust real-time face detection. International journal of computer vision 57(2), 137{154 (2004)
8. Von Agris, U., Zieren, J., Canzler, U., Bauer, B., Kraiss, K.F.: Recent developments in visual sign language recognition. Universal Access in the Information Society 6(4), 323{362 (2008).
9. Wang, R.Y., Popovic, J.: Real-time hand-tracking with a color glove. ACM transactions on graphics (TOG) 28(3), 63 (2009).

A Proposal for Outlier and Noise Detection in Public Officials' Affidavits

RODRIGO LÓPEZ-PABLOS[1,2], HORACIO D. KUNA[3]

[1]Escuela de Posgrado y Formación Continua,
Universidad Nacional de La Matanza, Argentina
[2]Escuela de Posgrado, Facultad Regional Buenos Aires,
Universidad Tecnológica Nacional, Argentina
[3]Departmento de Cs. de la Computación, Facultad de Ciencias Exactas,
Químicas y Naturales, Universidad Nacional de Misiones, Posadas, Argentina
{rodrigo.lopezpablos,hkuna}@gmail.com

Abstract. Outlier and noise detection processes are highly useful in the quality assessment of any kind of database. Such processes may have novel civic and public applications in the detection of anomalies in public data. The purpose of this work is to explore the possibilities of experimentation with, validation and application of hybrid outlier and noise detection procedures in public officials' affidavit systems currently available in Argentina.

Keywords: Anomalies and noise, Public data, Public officials, Affidavits, Databases, Outliers.

1. Introduction

Data mining and information exploitation processes have been scarcely used for solving civic issues related to the public sector, and outlier and noise detection processes are not the exception. Public data can be subject to anomalies and noise, like any kind of data in any given database (DB). However, the implications of the discovery of corrupt behavior in the public data of public officials, as well as in the quality of such data, might have significant and profound effects on society's welfare since corruption affects the social fabric and the quality of life of its populations. In this context, data mining might be a particularly useful tool for those citizens and civil organizations fighting against corruption since it can aid in the discovery of corrupt social fabric, thereby highlighting the usefulness of data mining as a tool in social sciences, monitoring and scientific research [1].

This paper is organized as follows: Subsection 1.1 presents the hypothetical questions posed in this work; Section 2 discusses the state of the art in data mining techniques aimed at detecting corrupt behavior; Section 3 analyzes open affidavit systems; Section 4 deals with the proposed outlier detection procedures; Section 5 addresses the experimentation with such procedures; and Section 6 presents the conclusions of this work.

1.1. Research Questions

This research proposal poses the following questions:

- Can the experimentation with the application and use of outlier detection techniques and processes in public databases of official affidavits be considered a useful tool to find signs of corrupt behavior in public administration and in the fight against corruption?
- Is it possible to assess the quality of such databases through noise and outlier analysis?
- Which outlier and noise detection procedures can be applied considering the nature of such systems?

2. State of the Art

Data mining and information exploitation techniques and processes have been scarcely used or even completely disregarded as an aiding civic tool in the fight against corruption in public offices. However, there is relevant research in the literature concerning mining techniques aimed at detecting financial and accounting fraud or corruption in the private sector.
Several authors [2, 3] have provided a comprehensive classification of data mining use against private corruption. Such corruption cases might take the form of financial and accounting fraud, including banking fraud, credit card fraud, money laundering and mortgage fraud, cases of complex private fraud, and complex financial fraud involving the forging of corporate information, financial speculation, etc. A classification of the techniques used in the field is presented below.

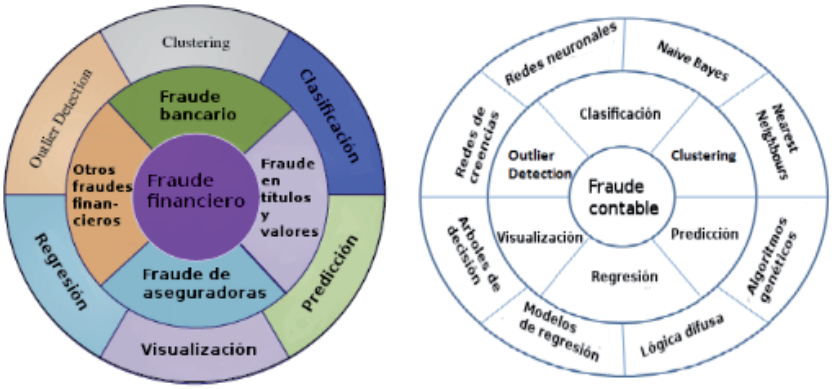


Fig. **1.** On the left, there is a classification of information exploitation techniques for financial fraud detection [2]. Data mining techniques and algorithms used for the detection of accounting or fiscal fraud are shown on the right [3].

Figure 1 shows that the use of these tools has been under-exploited in the public sector both for civic use and against corruption since there are no previous relevant works in the literature revealing any interest in the problem of public corruption.
In addition, without detriment to other data mining techniques and information exploitation processes, techniques for the detection of outliers or anomaly values have not been implemented to help solve fraud scenarios yet, let alone in the area of corruption in the public sector.

3. Open Affidavit Systems as an Alternative Solution in the Fight Against Public Corruption

In Argentina, affidavit systems are information systems regulated by affidavit regimes or systems that have three basic functions as a consequence of their implementation [4]:

[i] Controlling the patrimonial evolution of public officials in order to prevent illicit gain and other corruption-related crimes.
[ii] Detecting and preventing conflicts of interest and incompatibilities with public office.
[iii] As a mechanism for transparency and prevention of public corruption.
In terms of prevention, all public official affidavit regimes constitute a tool that enables: the control of appropriate compliance with public office duties by public officials, the prevention of deviation from their ethical duties and the correction of any detected non-compliance.

3.1 Open Affidavit Application Software and Systems

From the exploratory collection of public data in Argentina, the Open Financial Affidavit interactive system [5] jointly developed by *Diario La Nación and the NGOs Directorio Legislativo, Poder Ciudadano and Asociación Civil por la Libertad y la Justicia,* in force since 2013, was selected for the experiment. Out of 1550 financial affidavits on the interactive site, 539 affidavits corresponded to 99 public officials of the executive branch of government, 843 affidavits corresponded to 313 officials of the legislative branch, and 168 corresponded to 87 judicial branch officials, according to the updated version as of April 14, 2015. The data preparation strategy to perform the experiment consisted in only taking into account the public officials' real estate.
The open affidavit database has the following attribute structure for each public official's affidavit:

```
dj.funcionario =  ddjj_id, ano, tipo_ddjj, poder, persona_id,
                  nombre, ingreso, cargo, jurisdiccion, cant_acciones,
                  descripcion_del_bien, destino, localidad,
                  nombre_bien_s, origen, pais, porcentaje, provincia,
                  tipo_bien_s, titular_dominio, vinculo, superficiem2,
                  val_decl, valor_patrim)
```

The database created for the experiment has a total of 6627 tuples with 24 attributes since, from the original database with 39 initial attributes, 13 were discarded when partially or completely empty fields or nominal inconsistences in data imputation were found. In addition, some attributes became redundant after the homogenization and standardization of the three (3) generated attributes: `dj.patrimoniales (Gen)=(superficiem2, valor_patrim, val_decl)`.

For the purpose of the experiment, all real estate appraised in foreign currency was converted to Argentine pesos and inflation-adjusted `(valor_patrim)`, the real estate area was homogenized to square meters `(superficiem2)` and the real estate value declared by the public official `(val_decl)` was classified according to its relative fiscal valuation into fiscal, sub-fiscal, market value, or no value declared.

4. Proposed Hybrid Outlier and Noise Detection Procedures

Anomaly fields are defined as a datum which is so different from the other data belonging to the same data set [6], *i. e.* a database containing such fields, that it can be assumed to have been created by a different mechanism. It is precisely the discovery of such mechanisms that the analysis of each database pursues.

Hybrid detection methods –which combine various algorithms from different learning approaches–, have recently come to be regarded as processes, and the combination of different procedures allows outlier detection with a confidence level of over 60% [7]. Hybrid outlier detection methods offer the advantage of combining different techniques and algorithms in order to achieve the same goal. These methods include: LOF (Local Outlier Factor) and metadata, or LOF and K-Means, for numerical DBs as well as induction algorithms such as C4.5, PRISM, Information Theory, and Bayesian Networks; and clustering algorithms such as LOF, DBSCAN and K-Means for alphanumeric DBs with both supervised and unsupervised procedures.

The following table shows the use of hybrid methods with different approaches, depending on the learning approach of the algorithms involved, that is, supervised or unsupervised learning aimed at outlier and noise detection. Hybrid methods are the best alternative to obtain the highest amount of information, reduction of search space and process optimization [8, 9, 10]. Recent research has shown that the combination of different types of algorithms as well as the combination of procedures optimizes outlier

discovery [7]. As well as such paradigm, the following two alphanumeric procedures are proposed to aid civil auditors:

Table 2. Hybrid procedures proposed according to environment, algorithm and approach [7]

Hybrid procedures	Environment	Algorithms and techniques	Approaches
I	Alphanumeric DBs with a target attribute	C4.5, Information theory; LOF	unsupervised; supervised
II	Alphanumeric DBs with no target attribute	LOF; DBSCAN; C4.5; RB; PRISM; K-Means	unsupervised; supervised

As it can be observed in hybrid procedures I and II, procedure I detects outlier fields in alphanumeric databases containing a class attribute [11, 12, 7] while procedure II also detects fields in alphanumeric bases, but with no target attribute [13, 7].
Since affidavits are normally comprised of alphanumeric data, the potentiality of applying hybrid procedures I and II (table 2), corresponding to the hybrid procedures for the detection of anomaly data recently developed by [7], becomes evident. Such procedures are considered suitable due to the following reasons:

[i] They were recently developed and represent the state of the art regarding the detection of anomaly fields and noise.
[ii] They are optimal procedures for the detection of noise in alphanumeric databases, which are the type of database in which public data are generally stored.
[iii] They are easily applicable and executable with currently available data mining software.

The suitability of these hybrid procedures makes them a possible solution in civil auditing for the improvement of public and civic data of a given population. A possible solution and hypothetical application of civil auditing on open affidavit databases based on the above mentioned alphanumeric procedures are presented below:

Fig. 4. Proposal for the application of hybrid procedures for the detection of anomaly fields in alphanumeric affidavit databases [Our own design].

The proposed procedure (Figure 4) involves the application of hybrid alphanumeric anomaly detection procedures to detect outliers in the prepared affidavit databases. The possibility of detecting false positives proposes a circular feedback in both alphanumeric procedures proposed since the rest of the procedure is the same regardless of the presence or absence of outliers. Thus, the aim is to assess the quality of the public data involved in a preliminary analysis and, in a more profound *ex post facto* analysis based on detected fields and attributes, to assess the signs of implicit corrupt behavior as output of the public information processed.

5. Experimentation and Discussion of the Proposed Methodology

The experiment stage is followed by the validation of the system proposed in Figure 4 in public affidavits. Since the validity of tuples and attributes of real property affidavit databases cannot be determined, suspicion of anomalies is associated to certain characteristic economic parameters of the real property which are not trivial to its composition as well as to the quality of the collected data.

5.1 Validation of the Hybrid Procedures Used in Affidavit Databases

From the experimental implementation of procedure I in the unsupervised learning stage, considering the declared value as the class attribute of the C4.5 induction algorithm, significant attributes are obtained, thus acquiring the greatest amount of information (table 3). Six input-output bins were designed on such attributes thus simulating an information system. Then, the mining flows with the LOF algorithm were executed *ex post facto*, where "• • is the number of tuples which are suspected of containing anomalies.

Table 3. Input-output bins with detected outliers.

Input Bins – (Output)	Detected outliers	Suspicious anomaly bins Mean or Mode (Mode)
superficiem2(val_decl)	968 (•)	38733.15 (not declared)
ano(val_decl)	122 (•)	2001 (Market)
nombre_bien_s(val_decl)	209 (•)	Horizontal Prop. (Fiscal)
porcentaje(val_decl)	252 (•)	29.18528 (no data)
val_patrim(val_decl)	1130 (•)	146528.3 (Subfiscal)
vinculo(val_decl)	30 (•)	Co-inhabitant (Subfiscal)

The suspicious anomaly bins were associated to the following mean values and modes: [i] non-declared real property of an average area of 38733m^2, [ii] 2001 affidavits of real estate declared at its market value, [iii] horizontal property at fiscal value, [iv] average shareholding percentage of 29.19% with no monetary valuation, [v] average real estate value of ARS$146,528 at sub fiscal value, and [vi] sub fiscally valued properties of the co-inhabitant. According to the information theory [14], the real estate value and the area seem to create more noise and entropy in the affidavit system, quantitatively speaking.

In the execution of procedure II without a class attribute, following the determination rules of outlier [7] in the first phase, LOF-DBSCAN and the combination of classification algorithms C4.5-RB-PRISM were applied. As a result, 2531 tuples suspected of containing anomalies were detected; *i. e.*, 38.19% of 6627 tuples. An outlier database is designed for the following phase which is transformed *ex post facto* to apply K-Means in two clusterings, obtaining the following results:

Table 4. Distance from centroid for each attribute

Atribute(id) (Cluster_0)	**Distance value (Cluster_0)**	**Mean value** (Cluster_1)
ddjj_id(1)	**1.841**	1.079
ano(2)	**1.518**	1.079
superficiem2(22)	**2.033**	1.079

Upon executing the K-Means algorithm, it was observed that the furthest centroid contained the real estate area, year and affidavit identification attributes; where the former – superficiem2 – is the furthest attribute as well as the most suspicious of containing anomaly fields.

5.2 Discussion of Results of the Experimentation

The anomaly noise produced by the area, valuation, name of the property, shareholding percentage, year and relationship attributes (table 3) foresees an information system that shows the irregularities committed by public officials when appraising and declaring their properties to the citizenship. On the other hand, the area attribute shows an important source of dispersion that could reveal signs of extremely large properties together with other properties which are too small to be considered as such.
From the first induction analysis with a class attribute –procedure I–, through C4.5 algorithm, the following rules were observed:

[i] Properties less than 6500m^2 tend to be declared at sub fiscal value.
[ii] Properties larger than 6500m^2 – between mid-2005 and 2012 – tend to be not declared, just as garages, lands, plots and horizontal properties without specification greater than 6500m^2 and before 2005.
[iii] When the official owns a house and a share percentage higher than 46.3% [100%; 37.9%) or lower than 37.9 (37.9%; 0%] they will prefer not to declare its valuation, but will tend to declare it sub fiscally when they hold a share of between 46% and 38% (46.3%; 37.9%).
[iv] If the public official had an apartment, between mid-2003 and mid-2005, with an area less than or equal to 26m^2, it tends to be declared at its fiscal value, while it is not declared if it is prior to 2003. However, apartments larger than 26m^2 are simply not declared.
[v] Public officials' stores with an appraised value greater than ARS$76,493 [• , $76,493) or less than $10,615 ($10,615, 0] tend to be declared at their market value; except for the stores whose value is between [ARS$76,493 and ARS$10,615]. In such case, they will only be declared at their market value provided that the real property is registered under the name of the official's spouse and not under the official's name, in which case they will tend not to declare it.

6. Conclusions and Future Work

In this work, outlier and noise detection hybrid processes were developed and combined to be used in alphanumeric databases of affidavits of real property. It constitutes an original work dealing with the use of data mining techniques and processes as a tool against public corruption, disclosing the potentiality of anomaly detection processes when evaluating public databases quality, on the one hand, and the discovery of information suspected of containing corrupt behavior, on the other.
The attributes that contributed the highest amount of entropy to the affidavit system, thus jeopardizing the quality of the database, were: property area, property value, shareholding percentage, property name, and year. Due to

their low relative density, the property area is the most suspicious of containing anomaly data. In addition to revealing inconsistencies in the database, the possible detected anomalies could also show signs of corrupt behavior in relation to the fiscal value of the properties declared by the public officials, since the characteristics of the property could affect its patrimonial value, according to whether it is declared at its fiscal or sub fiscal value, or simply by avoiding its value declaration. In this regard, both the rules of behavior and the input-output bins could serve as a strategy of civic and accounting investigation in the fight against public corruption, and tax evasion and avoidance by public officials; all outrageous behaviors when considering the ideal exemplariness of the official in respect to the represented community.

Future research work could include the design of algorithmic variations in the proposed alphanumeric processes without discarding the application of variants to the alphanumeric processes proposed, as well as to the processes of information mining external to the detection of anomalies and noise on the same affidavit databases.

7. References

[1].Ransom J.: Replicating Data Mining Techniques for Development: A Case of Study of Corruption, Lund University, Master Thesis, Master of Science in International Development and Management, http://lup.lub.lu.se/record/3798253/file/3910587.pdf, (2013).

[2].Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X.: The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559-569 (2011).

[3].Sowjanya, S., Jyotsna G.: Application of Data Mining Techniques for Financial Accounting Frad Detection Scheme. International Journal of Advanced Research in Computer Science and Software Engineering, Noviembre, 3(11), 717-724 (2013).

[4].Gómez N., Bello M. A.: Ética, transparencia y lucha contra la corrupcion en la administracion publica, Manual para el ejercicio de la funcion publica, 1ra ed.: Oficina Anticorrupción, Ministerio de Justicia y Derechos Humanos de la Nación, Mayo, CABA, http://www.anticorrupcion.gov.ar/documentos/Libro%20SICEP%202da%20parte.pdf, (2009).

[5].DD.JJ. Abiertas: LNData. Actualizado al 13/1/2014, CABA, http://interactivos.lanacion.com.ar/declaraciones-juradas/, (2015).

[6].Hawkins, D. M.: Identification of outliers, Chapman and Hall., 11, London (1980).

[7].Kuna H.: Procedimientos de explotación de la información para la identificación de datos faltantes con ruido e incosistentes, Tesis doctoral, Universidad de Málaga, Marzo (2014).

[8].Kuna, H., García Martínez, R., Villatoro, F.: Identificación de Causales de Abandono de Estudios Universitarios. Uso de Procesos de Explotación de Información. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, 5, 39--44 (2009).

[9].Kuna, H., García-Martínez, R. Villatoro, F.: Pattern Discovery in University Students Desertion Based on Data Mining. In Advances and Applications in Statistical Sciences Journal, 2(2): 275–286 (2010).
[10].Kuna, H., Pautsch, G., Rey, M., Cuba, C., Rambo, A., Caballero, S., Steinhilber, A., García-Martínez, R., Villatoro, F.: Avances en procedimientos de la explotación de información con algoritmos basados en la densidad para la identificación de outliers en bases de datos. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Artículo 3745 (2011).
[11].Kuna, H. , Pautsch, G., Rey, M., Cuba, C., Rambo, A., Caballero, S., García-Martínez, R., Villatoro, F.: Comparación de la efectividad de procedimientos de la explotación de información para la identificación de outliers en bases de datos. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación, 296--300 (2012b).
[12].Kuna, H., Pautsch, G., Rambo, A., Rey, M., Cortes, J., Rolón, S.: Procedimiento de explotación de información para la identificación de campos anómalos en base de datos alfanuméricas. Revista Latinoamericana de Ingeniería de Software, 1(3): 102--106 (2013b).
[13].Kuna, H., García-Martínez, R., Villatoro, F.: Outlier detection in audit logs for application systems. Information Systems (2014).
[14].Ferreyra M.: Powerhouse: Data Mining usando Teoría de la información, (2007).

XI

Architecture, Nets and Operating Systems Workshop

Generalized state equation for non-autonomous Petri nets with different types of arcs

Orlando Micolini[1], Marcelo Cebollada y Verdaguer[1], Maximiliano Eschoyez[1], Luis Orlando Ventre[1], Marcelo Ismael Schild[1].

Laboratorio de Arquitectura de Computadoras (LAC) FCEFyN
Universidad Nacional de Córdoba
{orlando.micolini, marcelo.cebollada.y.verdaguer, maximiliano.eschoyez,
luis.ventre, marcelo.schild }@unc.edu.ar

Abstract. This work proposes to generalize the state equation of a Petri Net, with the objective of representing different types of arcs and time semantics in non-autonomous Petri nets. This generalized equation facilitates the hardware implementation of an IP-Core to execute the network.
The solution that this extended state equation expresses raises an algorithm that preserves the original model, facilitates parallel execution and allows addressing problems bigger in size and complexity. In addition, a case of application is exposed, highlighting the advantages of including events, different types of arcs and timed semantics.

Keywords: non-autonomous Petri net, state equation, Petri Processor, IP-Core.

1. Introduction

The implementation of threads in computer systems allows to exploit resources of multicore architectures [1]. These threads cooperate and execute concurrently. Applications designed this way have an additional intrinsic complexity when compared to sequential programs. This complexity becomes clear in the design, error detection, testing, validation and maintenance [2]. Hereby, it is necessary to introduce control mechanisms, such as semaphores, which penalize execution times. Because of this, it is convenient to generate a formal solution that guarantees and facilitates the system's development and implementation.
Recent investigations have shown that models obtained through Petri nets (Pn) directly facilitate the system's implementation, using software, processors or IP-Cores that execute Pn [3, 4]. Non-autonomous Pn processors (PP) are presented in [5, 6] as IP-Cores implemented in a Spartan-6 FPGA.

The simulation and execution of complex systems require a Pn modeling with grater semantics. In order to extend the Pn semantic capability, events, guards, different type of arcs (inhibitors, readers, etc.) and time semantics [7] are included. This expands the problems domain, as well as their size and complexity.
Even though in [8] an extension to the Pn state equation is presented, the authors did not consider events, guards or time. They also introduce variables in the incidence matrix, which hinders its implementation as a combinational circuit.
In this work, the Pn state equation is generalized using the firing inhibition concept to obtain different semantic applications and implement them in a combinational system. Thus, the FPGA resources are more efficiently used, maintaining all of the non-autonomous and timed Pn properties.

2. Notation and Fundamentals

a. State equation

Pn are used to graphically represent a system's dynamic behavior. Therefore, the graphic characteristics of Pn can be exploited to inspect a system's dynamics. This approach is suitable for small systems. However, the graphic methodology is not efficient when systems are large and complex. In this article, a state equation is developed to represent models of these systems and implement them as an IP-Core.
It's important to introduce the Pn's state equation in order to extend it with different types of arcs, events, guards and time. Through this equation, it is possible to obtain the next state of the system. This way is simpler than the graphic approach when analyzing the system's evolution.
The Pn state equation for a net with n places and m transitions, arcs with a weight greater or equal to one and an initial mark of M_0, is:

$$M_{j+1} = M_j + I * \sigma. \quad \textbf{(1)}$$

Where: I is the incidence matrix, with a size of $n \times m$. σ is the firing vector, with a size of $m \times 1$. M_0 is the initial marking vector, M_j is the current marking vector, and M_{j+1} is the next state's marking vector, all of them have a size of $n \times 1$.
The i_{ij} elements in the I matrix are obtained as:

$$i_{ij} = w(t_i, p_j) - w(p_j, t_i). \quad \textbf{(2)}$$

Where w is the weight of each arc (which takes integer signed values):

- Arcs from place -i to transition-j, are $-w(p_j, t_i)$.
- Arcs from transition-j to place-i, are $w(t_i, p_j)$.

When an enabled transition is fired, the next state can be calculated using (1).

The state equation mathematically represents the system's dynamic behavior and the incidence matrix characterizes the system's behavior. Therefore, the incidence matrix represents the system itself.

b. Enabled Transition

A transition is enabled if all of its input places have a marking greater or equal to the weight of the arc that runs from each place to the transition. This is expressed as:

$$M(p_i) \geq w(p_j, t_i), \forall p_j \in I(t_i).$$

Enabled transitions are expressed as a binary vector E, which size is $m \times 1$. Each of the vector's components indicates an enabled transition with a value of one, and a disabled transition of a value of zero.

$$E = \left(sign(S^0), sign(S^1), \ldots, sign(S^{n-1})\right).$$

Were the relation $sign(S^i)$ for each S^i vector is:

- S^i is equal to $M_j + I^i$, a vector with the marking that would be obtained by firing transition i once.
- $I^0, I^1, \ldots, I^{n-1}$, are the columns of the I matrix
- $sign(S^i)$, is a binary value. It is worth zero if S^i has any negative components, or one otherwise.

c. Firing a transition

A transition can be fired only if it is enabled. When it's fired, the new marking M_{j+1} is calculated, along with the new vector of enabled transitions E.

d. Incidence matrix implementation

From the firing semantics of a Pn, the incidence matrix can be interpreted as the conjunctive evaluation between columns (transitions) and the restrictions that are imposed by the rows (places). So, in an incidence matrix with dimensions $m \times n$, n combinations of m logic variables are evaluated, as expressed in the following equation:

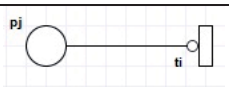
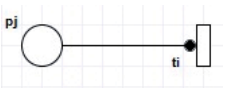
$$e_i = \left(\bigwedge_{h=0}^{m-1} M(p_h) \geq i_{hi}\right), \forall i = 0, \ldots, n-1.$$

Where i_{hi} are the elements if the I matrix and $M(p_h)$ is the marking for place h. This results in the components of vector E.

3. Extending the State equation

a. Reset, Reading and Inhibitor Arcs

Inhibitor, lector and reset arcs connect places with transitions, enhancing Pn's expression capability. It should be noted that inhibitor and lector arcs allow to model priorities.

Arc	Representation	Effect over semantics
inhibitor		If $h(p_j, t_i)$ is an inhibitor arc and p_j has a token, then t_i is not enabled.
lector		If $l(p_j, t_i)$ is a lector (or reading) arc, and there's not any token in p_j, then t_i is not enabled. Firing t_i does not consume any tokens present in p_j.
reset		If $r(p_j, t_i)$ is a reset arc and t_i is fired, then all tokens in p_j are consumed. Firing t_i sets the mark in p_j to zero.

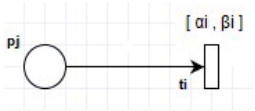
b. Guards

A guard is a logic variable associated to a transition that enhances its expression capability. Graphically, a guard is represented with a tag next to a transition. If and t_i is a transition and $g(t_i)$ is a guard associated to t_i, t_i can only be fired if it's enabled and if the guard's value is true. Guards enable Pn to communicate with the environment, making them non-autonomous.

c. Events

An event is stored in a queue associated to a transition [3], enhancing the Pn's expression capability. Graphically, the queue is represented with a tag next to a transition. The queue c_i is a counter, which increments each time a new event comes in, and decrements when the associated transition is fired. For the transition to be enabled, it is required for the queue to have at least one event. Events communicate the Pn with the environment, making them non-autonomous.

d. Time



There are several time semantics, as proposed in [9-11]. In this work, we have taken Pn with time, introduced in [12]. These Pn impose time limitations over transition shots, associating a time lapse with each transition.

That is to say, that each active transition has an implicit chronometer associated, which measures the time elapsed since the last time that the transition was enabled.

An enabled transition can be fired if the timed value associated with it is within the predefined time lapse. For example, the time tag $[\alpha_i, \beta_i]$ associated to t_i has α_i as starting time and β_i as finishing time.

e. Shot selection policy

The shot selection policy consists in the selection of the next transition to be fired among all of the ones that are currently enabled. For Pn, including non-autonomous Pn and Pn with time, if this selection is random, the resulting system is non deterministic. There are different solutions that provide deterministic behavior, such as including priorities, probabilities, inhibitor arcs, lector arcs, etc.

In order to avoid conflicts among the transitions, we use the single server policy. This means that we calculate the new state (marking) considering only one shot of one transition. If multiple shots are required, a sequence of shots is performed.

4. Extended state equation

a. Considerations for generalizing the state equation

In every enhancement of the exposed semantics, selecting which transitions are enabled was always highlighted. Once we define which is the next transition to be fired, the shot is performed with the original state equation, except for the reset arc, which removes every single token from the associated place.

In order to mathematically represent the existence of the previously enumerated arcs, matrixes that indicate the connections between places and transitions are needed. These matrixes are similar to the I matrix. The matrixes's elements are binary, since the arc's weight is always one.

So, for a transition to be enabled, the following conditions must be met:

- If it has an inhibitor arc, the place must not have any tokens.
- If it has a lector arc, the place must have at least one token.
- If it has a guard, the guard value must be true.
- If it has events, the event queue must have at least one event.

- If it has a tag with a time lapse, the counter's value must be contained within the specified time lapse.

b. New state equation

From the previous considerations it can be seen that, in order to fire a transition, a logic conjunction among all of the enumerated conditions and, the E vector of enabled transitions, is required.

The B vector of transitions disabled by inhibitor arcs, is a binary vector of dimensions $m \times 1$, which shows which a zero which transitions are disabled by inhibitor arcs and with a one those that are not. It is obtained as:

$$B = H * Q.$$

- Where H is a matrix of dimensions $m \times n$ and Q a binary vector of dimensions $n \times 1$.
- H is the incidence matrix that relates the places that are connected to transitions through inhibitor arcs. The elements h_{ij} from the matrix have a value of one if there's an arc that runs from place p_i to transition t_j, and a value of zero if there's not.
- The q_i components from the Q vector are obtained as:

$$q_i = zero\big(M(p_i)\big).$$

- The relation $zero(M(p_i))$, is zero if the marking on place p_i is different than zero, and one otherwise.

The L vector of transitions disabled by lector arcs, is a binary vector of dimensions $m \times 1$, that indicates with a zero which transitions are disabled by a lector arc and with a one the transitions that are not. Itis obtained as:

$$L = R * W.$$

- Where R is a matrix of dimensions $m \times n$ and W a binary vector of dimensions $n \times 1$.
- R is the incidence matrix that relates the places that are connected to transitions through lector arcs. The elements r_{ij} from the matrix have a value of one if there's a lector arc that runs from place p_i to transition t_j and zero otherwise.
- The w_i components from the W are obtained as:

$$w_i = one\big(M(p_i)\big).$$

- The relation $one(M(p_i))$, is one if the marking on place p_i is different than zero, and zero otherwise.

The G vector of transitions disabled by guards, is a binary vector of dimensions $m \times 1$, that indicates with a zero which transitions are disabled by a guard and with a one the transitions that are not. It's obtained directly from the guards.

The V vector of transitions disabled by events, is a binary vector of dimensions $m \times 1$, that indicates with a zero which transitions are disabled because there are not any events requesting to fire the associated transition. The v_i components of the V vector are obtained as:

$$v_i = one(eventBuffer_i).$$

The relation $one(eventBuffer_i)$, is one if there is at least one event in the event buffer associated with transition t_i, and zero otherwise.

The Z vector of transitions disabled by time, is a binary vector of dimensions $m \times 1$, that indicates with a zero which transitions are disabled because the starting time has not been reached or because the time lapse has passed since the transition was first enabled. Otherwise its value is one, and is obtained as:

$$Z = Tim(q(E, B, L, G, clk), intervals).$$

- E is the "enabled vector", B is the "disabled by inhibitor arcs" vector, L is the "disabled by lector arcs" vector and G is the "disabled by guards" vector.
- The relation $Tim(q, inervals)$ is a binary vector of sions $m \times 1$. The value of the i component is one if the value of q_i, which is a counter, is within the time lapse indicated by the component $intervals_i = [\alpha_i, \beta_i]$. Otherwise it's zero.
- For the q_i counter to start, the equation "e_i and b_i and l_i and g_i" must return a value of one. Otherwise the counter is set to zero. This counter is incremented one time-unit per clock cycle of the base time unit (clk).
- The $intervals$ matrix of dimensions $m \times 2$, contains the lower and upper limit of the time frame in which the transition's firing is enabled.

The A vector of transitions reset, is an integer vector of dimensions $m \times 1$, which contains the current marking value of the place that will be set to zero, while other components have a value of one.
The Re Matrix mathematically expresses reset arcs. The arc that runs from p_i to t_j is indicated as a value of one in the re_{ij} component of the matrix, otherwise re_{ij} is zero.

$$A = Marking(Re * M_j).$$

The product $Re * M_j$ results in a vector which components are zero if there's no reset arc, or the proper marking if there actually is a reset arc.
The relation $Marking()$ results in a value of one if the component is zero, otherwise results in the value of the component.
The A vector is multiplied, element by element, with σ the firing vector. This operation is indicated with #. The σ vector has a one in the transition that we want to fire, therefore we obtain the quantity of shots necessary to take all of the tokens from the place to be reset, otherwise we obtain just one shot.

c. Extended enabled vector and state equation

Ex, the extended enabled vector, is obtained through the logic conjunction of all of the previous vectors.

$$Ex = E\ and\ B\ and\ L\ and\ V\ and\ G\ and\ Z.$$

To introduce the reset arc, it is necessary to multiply element by element (#) the vector that results in this conjunction with the vector A. This way, the extended state equation results in:

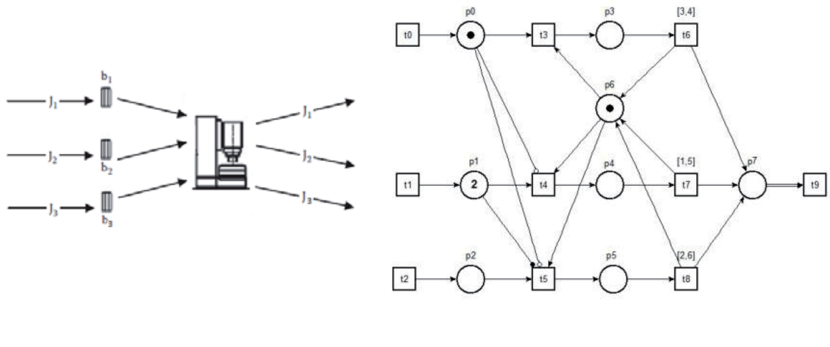
$$M_{j+1} = M_j + I * ((\sigma\ and\ Ex)\#A.$$

The relevance of this state equation relies in its simple implementation as a combinational circuit in a FPGA.

5. Case of Application

a. Production system modeled with an extended Pn

Figure 1 shows the use of inhibitor and lector arcs to model priorities and sequences in a production system. Let's consider the fabrication process represented in Figure 1(a) where a machine processes three types of workloads J1, J2 and J3. J1 has a higher priority than J2 and J3, and there must be at least one J2 standing by in order to process J3.



(a) **(b)**

Figure 1: (a) Production System and (b) Pn that models the system.

The Pn in Figure 1(b) shows the system's model. In this model t0, t1 and t2 represent the arrival of requests for J1, J2 and J3, the tokens in p0, p1 and p2 represent the requests themselves. A token in any of the places p3, p4 or p5 means that the work is being performed. Transitions t6, t7 and t8 have a time tag associated, which is required for each work and which firing corresponds to the work's completion. Place p6 represents the machine and place p7 counts the amount of works performed. Firing t9, sets place p7 down to zero. The following are the matrixes and vectors that correspond to the model presented in Figure 1.

Matrix **I**

-1	0	0	1	0	0	0	0	0	0
0	-1	0	0	0	0	0	0	0	0
0	0	-1	0	1	1	0	0	0	0
0	0	0	-1	0	0	1	0	0	0
0	0	0	0	-1	0	0	1	0	0
0	0	0	0	0	-1	0	0	1	0
0	0	0	1	1	1	-1	-1	-1	0
0	0	0	0	0	0	-1	-1	-1	0

Matrix of inhibitor arcs ***B***

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Matrix of lector arcs **L**

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Matrix of reset arcs ***R***

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1

Matrix of time lapses	Vector of transi-tions disabled by time	Marking Vector	Vector of enabled transitions	Vector of transitions disabled by inhibitor arcs	Vector of transitions disabled by lector arcs	Extended Vector of enabled transitions
$intervals$	Z	M_j	E	B	L	Ex
0 0	0	1	1	1	1	1
0 0	0	2	1	1	1	1
0 0	0	0	1	1	1	1
0 0	0	0	1	1	1	1
0 0	0	0	1	0	1	0
0 0	0	0	0	0	1	0
3 4	0	1	0	1	1	0
1 5	0	0	0	1	1	0
2 6	0		0	1	1	0
0 0	0		0	1	1	0

The implementation of this equation requires 6 matrixes. It is possible to apply net division techniques as the ones presented in [13], in order to obtain matrixes of smaller dimensions, which allows to save resources.

6. Conclusions

This article presented the state equation and the enabled transitions vector of Pn, generalized for multiple kinds of arcs, events, guards, and time. This equation and vector are of conjunctive logic, which facilitates their implementation with combinational circuits in an IP-Core. As a precaution, no variables were introduced in the matrixes, since this would hinder the implementation and validation of the state equation in a combinational circuit.
This new state equation facilitated the software and hardware development, which implements and executes non-autonomous and timed Pn. This software and hardware were used to obtain the system's states as well as the program's both concurrent and parallel execution logic.
The generalized state equation was obtained through the conjunction of vectors. In order to achieve this, a specific vector was obtained for each type of arc, time tags, events and guards. This provided modularity in software-hardware depending on the Pn that is to be implemented.
As future work, a new PP is being designed with a pipelined architecture based on this equation, which will allow to execute systems that are greater in complexity.

References

1. David R. Martinez, R.A.B., M. Michael Vai, *High Performance Embedded Computing Handbook A Systems Perspective*2008, Massachusetts Institute of

Technology, Lincoln Laboratory, Lexington, Massachusetts, U.S.A.: CRC Press.
2. Domeika, M., *Software Development for Embedded Multi-core Systems*2008, 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA Linacre House, Jordan Hill, Oxford OX2 8DP, UK.
3. Micolini, O., *ARQUITECTURA ASIMÉTRICA MULTI CORE CON PROCESADOR DE PETRI*, in *Informatica*2015, UNLaP: La Plata, Argentina.
4. Moutinho, F. and L. Gomes, *Distributed Embedded Controller Development with Petri Nets: Application to Globally-Asynchronous Locally-Synchronous Systems*. Vol. 150. 2015: Springer.
5. Micolini, O., J. Nonino, and C.R. Pisetta. *IP Core Para Redes de Petri con Tiempo*. in *CASIC 2013*. 2013.
6. M. Pereyra, N.G., M. Alasia and O. Micolini, *Heterogeneous Multi-Core System, synchronized by a Petri Processor on FPGA.* IEEE LATIN AMERICA TRANSACTIONS, 2013. **11**: p. 218-223.
7. Diaz, M., *Petri Nets Fundamental Models, Verification and Applications*2009, NJ USA: John Wiley & Sons, Inc.
8. BaşkocagİI, C. and S. Kurtulan, *Generalized state equation for Petri nets.* WSEAS TRANSACTIONS on SYSTEMS, 2011. **10**(9): p. 295-305.
9. Sifakis, J., *Performance evaluation of systems using nets.* Springer-Verlag Berlin Heidelberg, 1979. **84**: p. 307-319.
10. Roussopoulos, J.E.C.y.N., *Timing requirements for time-driven systems using augmented Petri nets.* EEE transactions on Software Engineering, 1983. **9**(5): p. 603-616.
11. Jensen, K. and L.M. Kristensen, *Coloured Petri Nets Modelling and Validation of Concurrent Systems.* Springer 2009.
12. Merlin, P.M., *A Study of the Recoverability of Computing Systems*, 1974, University Microfilms, : University of California.
13. Micolini, O., M. Cebollada, and L.O. Ventre. *Localidad estructural, criterio de división para la ejecución de redes de Petri no autónomas en IP-Core*. in *XXI Congreso Argentino de Ciencias de la Computación (Junín, 2015)*. 2015.

Design of a CAN Simulation Device for Communications in Sensor Networks

FERNANDO G. TINETTI[1,4], FERNANDO ROMERO[1], MARTÍN PI PUIG[1], SANTIAGO MEDINA[1], ARY BATISTA[2], DIEGO ENCINAS[1], ARMANDO DE GIUSTI[1,3]

[1]Instituto de Investigación en Informática LIDI (III-LIDI),
Facultad de Informática, Universidad Nacional de La Plata,
50 y 120 2do piso, La Plata, Argentina.
[2]Universidad Nacional de Quilmes, Roque Sáenz Peña 352, Bernal, Argentina.
[3]CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas
[4]CIC – Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires
{fernando, fromero, mpipuig, smedina, dencinas, degiusti}@lidi.info.unlp.edu.ar,
arypbatista@gmail.com

Abstract. Real-Time Distributed Systems must run algorithms under deadline time constraints defined by the applications. Hardware verification and validation stages usually include several changes and running new experiments. One way to decrease the complexity and probability of errors in hardware development is simulation. A design, implementation, and validation of a simulation model of a Controller Area Network (CAN) communications system is proposed and tested in this paper. The purpose of the model is to predict the transmission performance in different scenarios.

Keywords: Modelling and Simulation, Communications in Real Time Systems, Real Time Distributed Systems.

1. Introduction

The communication infrastructure in a real-time distributed system must meet non-functional requirements [15] very different from those required by other systems. These requirements include accuracy and reliability. The accuracy requirement usually involves short latency and minimum jitter, this being one of the main characteristics of real-time communication systems. In sensor networks, the latency is the elapsed time between the sensor reading in each of the nodes and the output of the corresponding actuator (data processing time is included). Jitter is measured as the difference between worst case and best case latencies, and a high value of the jitter may produce negative effects to bind the receiving time of a message.
Reliability involves features that communication systems are required to have to guarantee that the data sent is properly received (and this is usually

achieved with redundancy). Also, a real-time communication protocol must be flexible enough to support changes without having to change software for each minimum change. The structural requirements of the network are determined not only by technical considerations (functional requirements) but also by economic considerations. One of the first steps in the design phase of these systems is to perform an analysis of their operation. At present, this is achieved by means of simulation techniques which allow to study several aspects, such as:

1) Interaction among different components.
2) Communications protocols.
3) Operation modes.
4) Subsystem power consumption.

The purpose of the functional simulation is not to reproduce the exact behavior of the components as it would lose the advantage of abstraction of the physical characteristics [1]. In this paper, we propose a simulation model of a CAN communication device [2], using the Proteus simulation environment [3] on which it is not currently available. The proposed model validation will be carried out by means of a physical system with which the data obtained with the tests in the simulator will be compared.

This paper is organized as follows. Section 2 includes a general description of the CAN communication protocol. The simulation methodology is explained in Section 3, which provides the basis for the model development. Section 4 describes the simulation model implementation. Simulation model verification is described in Section 5. Finally, Section 6 presents the conclusions and further work we expect to carry out in the future.

2. The CAN Protocol

The CAN protocol has been formally defined by the ISO 11898 standard [4]. This protocol was designed by Robert Bosch GmbH in Germany and the initial purpose was to create a fast and reliable communication network within the automotive field. The standard defines a bus deterministic communication, with priorities and error detection. The physical bus is generally made up of a pair of cables, and stations or nodes. The specification of the cabling transmission medium is not indicated in the original standard, and it can be defined depending on the area of application. The encoding used is Non Return to Zero (NRZ) with stuffing on a differential (balanced) signal channel, which ensures compact messages with a minimum number of transitions and a high resistance to noise. Fig. 1 schematically shows the typical connection of the nodes to the network and the elements that compose it.

CAN is known as a serial communication protocol with high reliability, robustness and performance. With a properly designed and implemented application network layer CAN is also suitable to distributed real-time systems control. CAN most relevant characteristics can be enumerated as:

- Message priorities.
- Multi-master system.
- Flexible configuration.
- Medium transmission speed (up to 1 Mbit/s).
- Failure detection and signaling.

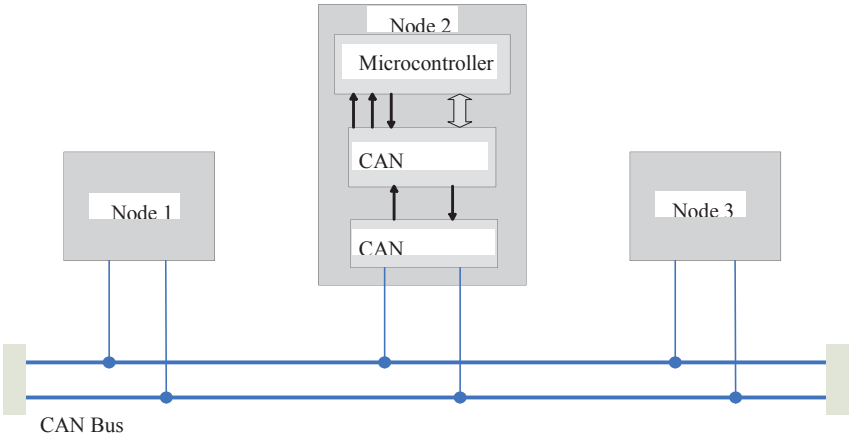


Fig. 1: CAN Connection Network.

The transaction mechanism used by the CAN protocol for data transferences is like the producer/consumer model, in which each node sends messages (in multicast or broadcast mode) without requiring receiver acknowledgement. Bus access is multi-master, and is not assigned by a central node. All nodes can try to access at the same time. The arbitration procedure involves the node identifiers simultaneously using the bus. There are four communication frames in the CAN protocol: data frames, remote frames, error frames, and overload frames. Fig. 2 schematically shows the fields of a data frame.

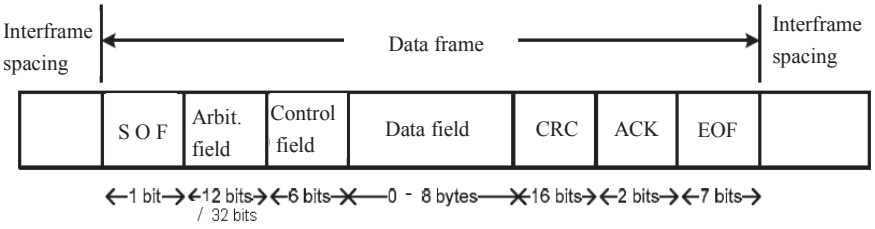


Fig. 2: CAN Data Frame.

3. Methodology

Several specific CAN technical definitions as well as some simulation design environment details are explained in this section. CAN technical definitions are needed for implementing the simulation, and the environment will be useful for the experiment measurements taken for model simulation performance analysis.

3.1 Message Transmission

The simulation model takes into account the CAN message transmission scheduling as well as simulation specific constraints. The CAN planning model used is the one proposed in [5] and previous works. One important planning concept is that each message is queued by a task, process, or software interrupt that runs on the microcontroller/host processor. This task is invoked by an event or status query with a limited amount of time to queue the message. This time varies between 0 and Jm, where Jm is called the maximum message jitter.

Fixed priority planning models such as the CAN protocol, have jitters despite having only periodic tasks [6]. There are several works that carry out a protocol scheduling analysis feasibility [7] [8] [9] and introduce jitter as a necessary parameter to carry out the planning of messages that are periodically sent.

Random jitter causes a delay of the transmission time. In [10] a stochastic analysis of the response times is performed and a uniform distribution function for the jitter is proposed. However, considering that jitter is the result of multiple different subsystems, which in turn have their own jitters, it is possible to use the central limit theorem. The central limit theorem states that the probability distribution function of the sum of a sufficiently large number of random variables with different distributions, can approximate a normal or Gaussian [11] distribution. This is one of the reasons to propose and use a normal or Gaussian function of probability distribution of jitter in this work.

3.2 Read and Write Message Windows

Sending and receiving messages is implemented through the concept of writing and reading message windows. Message windows are designed to have a fixed time interval and represent the message constraints when they arrive at a time or time interval when they cannot be transmitted despite having a high priority. Therefore, when a message arrives during a reading window time interval, it must wait until the start of writing window to compete with other possible messages at the time of arbitration.

3.3 Arbitration

Each node writes the bit corresponding to its identifier during the writing window period. In the next reading window, the node verifies the bit in the bus: if this bit differs, it means that a message issued by a node with a higher priority exists. Then the nodes with lower priority only receive the message that is being transmitted by the channel. That is, a node gains access to the bus in the arbitration process if it can completely write its identifier. Once a node obtains the control of the bus it sends its message.

3.4 Global Clock

Network nodes synchronization is given by a signal provided by the simulation environment. This signal is emitted at regular intervals over a synchronization channel and determines the start and end of the write and read windows. Therefore, a global clock is assumed possible by the Hard Synchronization and Resynchronization [12] methods which are considered sufficient to synchronize the clock signals of each node in the real system.

3.5 Tools

A set of tools was used for simulation modeling and implementation. The most important ones are listed below:

- Proteus v7.7.
- CCS Compiler v5.0.

Also, within the Proteus simulation suite, it was decided to use PIC16F877A microcontroller as a basic implementation component for CAN nodes and controllers.
For physical system implementation, the selected microcontroller was the AT89S52, SJ1000 was chosen as the CAN controller and PCA82C250 as the CAN transceiver component.

4. The Simulation Model

In this section, we describe the implementation as well as the programming model developed for simulation. Most design decisions and tools have been listed above, and will be further explained only for better understanding of the model.

4.1 Implementation

The system basically contains two main modules: the CAN Node and the CAN Bus. As it happens physically, the modelled node is composed of a microcontroller and a CAN controller. In this simulation model, the CAN transceiver has not been taken into account since this component adds a fixed

delay of the order of nanoseconds, which is a small value compared to the time necessary for the transmission of a bit (1 μsec). Fig. 3 schematically shows the implemented behavior for sending-receiving a CAN frame.

4.2 Programming Model

Fig. 4 shows an activity diagram of the main simulation system components at runtime. The microcontroller behaves as containing the system sensor as well as the CAN controller. This component generates the message data and id (identifier). Message data and id are sent to the CAN controller via the RS-232 protocol. The CAN controller and the bus are among the main components of the simulation system. By means of different functions, the fields of the data frame are completed, the data to be transmitted is entered, and the jitter corresponding to the transmission buffer is added. As mentioned in section 3.1, it is proposed to represent the jitter with a normal distribution function. The generation of random numbers with normal or Gaussian probability density function was implemented applying the Box-Muller transform [13].

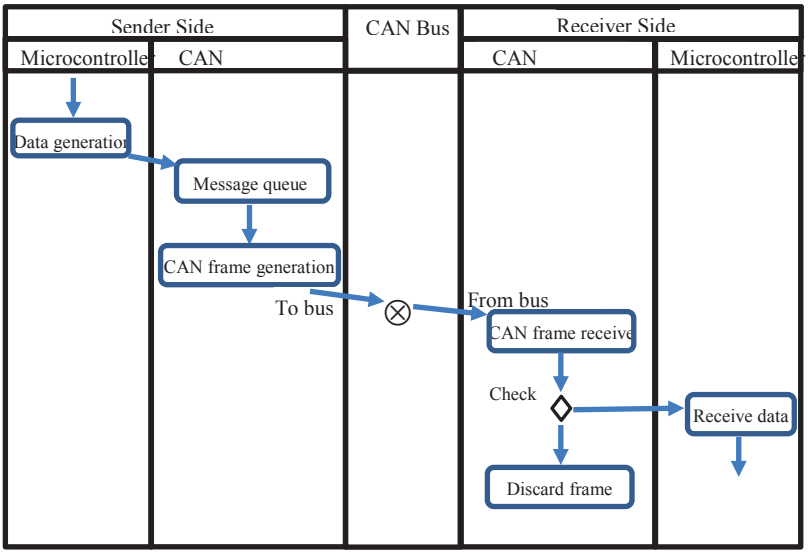


Fig. 3: Send-Receive a Data Frame.

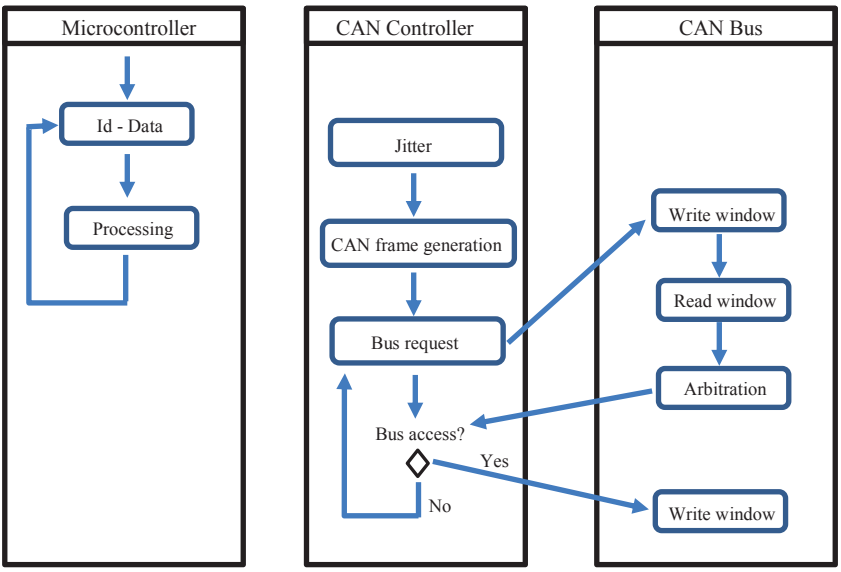


Fig. 4: Simulation System Activity Diagram.

The data bus (CAN Bus) was implemented using OR logic gates to allow each node to increment the value of the current window. An AND logic gate type is used for resetting the bus every time the global clock resets the synchronization channel.

5. Results

Once we have implemented the simulation model, and having a real CAN bus we can carry out several experiments and collect simulation and real experiments data for comparison. Since the real system has several restrictions, we initially show the results on the real system and we later use the simulation model to collect the corresponding data.

5.1 Experiments on the Real System

The sending period was set at 4 msec, and the transmission rate was 3.96 msec. This is a consequence of the microcontroller's architecture (size of registers, specifically), the transmission process of the CAN controller, and the frequency of the crystals in each node.

In general, traffic in traditional data networks is analyzed taking into account network throughput, latency, and jitter [14]. Although in a network for strict real-time systems these parameters are necessary to quantify the quality of service (QoS), because of the specific characteristics of the networks in the CAN protocol the throughput is almost meaningless. In fact, the throughput

is fixed and 1 Mbps, since it is the maximum transfer speed that the protocol supports. That is why the evaluated parameter is the message Inter Arrivals Time (IAT).

The number of messages handled by the CAN bus was 37339 in about 150 sec (about 4.02 msec per message in average). Fig. 5 shows the number of received messages in specific IAT (shown in the x-axis).

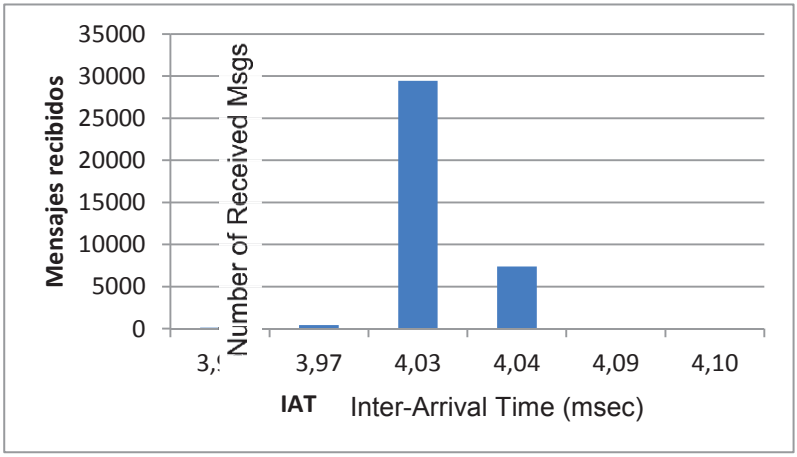


Fig. 5: Inter Arrival Time Measurements in a Real CAN System.

Fig. 5 shows that the highest number of messages is received at an average IAT of 4.03 msec and the rest of the messages do not deviate more than 70 μsec from that (modal) value.

5.2 Experiments with the Simulation Model

The simulation model adds several execution overheads, such as: a) processing time to implement the CAN Protocol CRC, b) arbitration (reading and writing windows) and, c) the Gaussian distribution jitter. The scenario that was simulated was based on the physical system, a node that transmits an 8-byte data message to a receiving node at a periodic rate of 3.96 msec. Again, the test consisted of sending 37339 messages. Figure 6 shows the simulation model IAT.

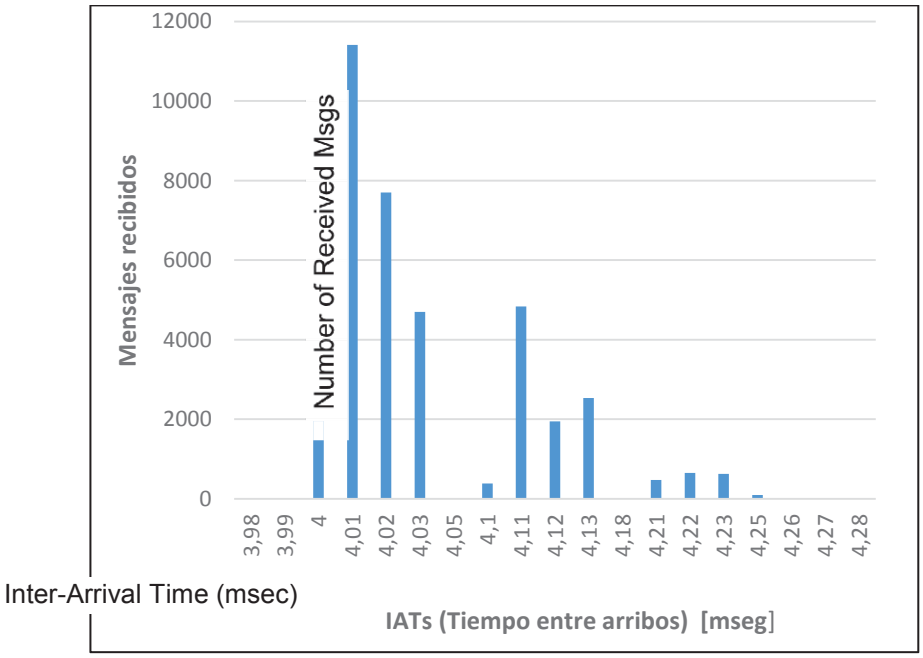


Fig. 6: Inter Arrival Time Measurements in a Simulated CAN System.

Fig. 6 shows that the main difference with the physical system: the average IAT value is about 4.05 msec with a maximum deviation of 230 μsec. On the other hand, the simulated environment increases the number of different IATs, allowing for a greater resolution. Even with different resolutions, a similar behavior can be observed in the two models.

6. Conclusions and Further Work

From the observation of the results of the simulation model it can be said that it complies with the CAN protocol specification and is thus validated. In addition, comparing the results of the simulation with the results obtained with the physical system, it can be stated that the two systems produce similar outputs given the same input data. One of the next steps to proceed with the present work would be to develop the necessary implementation for the simulation model in order to support other types of CAN frames, to improve the bus scaling, and to analyze other probability distributions for the jitter.

References

[1] EICKHOFF J. (2009). Simulating Spacecraft Systems. Springer.
[2] Robert Bosch GmbH. CAN Specification 2.0. 1991

[3] Proteus. https://www.labcenter.com. 2016
[4] ISO 11898: Road Vehicles –Interchange of digital information– Controller Area Network (CAN) for high speed communication. 1993.
[5] R. Davis, «Controller Area Network (CAN) schedulability analysis: Refuted, revisited and revised.,» Real-Time Systems. Springer, vol. 35, nº 3, pp. 239-272, 2007.
[6] A. Burns, Real-time systems and programming languages, Addison Wesley, 2009.
[7] R. Davis, «Controller Area Network (CAN) Schedulability analysis with FIFO queues», 23rd Euromicro conference on Real-Time Systems, pp. 45-56, 2011.
[8] P. Yomsi, «Controller Area Network (CAN): Response time analysis with offsets,» 9th IEEE International workshop on Factory Communication Systems, pp. 43-52, 2012.
[9] N. Navet, «Controller Area Network (CAN) Schedulability analysis for messages with arbitrary deadlines in FIFO and work-conserving queues,» 9th IEEE International workshop on Factory Communication Systems, pp. 33-42, 2012.
[10] M. Di Natale, Undertanding and Using the Controller Area Network Communication Protocol. Theory and Practice, Springer, 2012.
[11] J. Devore, Probabilidad y estadística para ingeniería y ciencias. Sexta Edición, Thomson Learning, 2005.
[12] Philips semiconductors, Application note. Determination of Bit Timing Parameters for the CAN Controller SJA 1000. AN97046, 1997.
[13] Box, G. E. P.; Muller, Mervin E. A Note on the Generation of Random Normal Deviates. Ann. Math. Statist. 29. 1958
[14] J. Kurose, Redes de Computadores. 2° Edición, Pearson-Addison Wesley, 2004.
[15] Hermann Kopetz.Real-Time Systems.Design Principles for Distributed Embedded Applications. Second Edition. Springer. 2011.ISSN 1867-321X e-ISSN 1867-3228 ISBN 978-1-4419-8236-0 e-ISBN 978-1-4419-8237-7

Using White Spaces: A solution for frequency spectrum overloading

ANTONIO CASTRO LECHTALER[1, 2], ANTONIO FOTI[3, 4], ALEJANDRO ARROYO ARZUBI[1], JORGE GARCÍA GUIBOUT[5], FERNANDA CARMONA[6], RUBÉN FUSARIO[1, 2], ALEJANDRO OLIVEROS[3].

{[1] Escuela Superior Técnica, Universidad de la Defensa, C1426AAA, Ciudad de Buenos Aires; [2] CISTIC - Facultad de Ciencias Económicas, Universidad de Buenos Aires, C1120AAQ, Ciudad Autónoma de Buenos Aires; [3] Universidad Nacional del Oeste, B1718, San Antonio de Padua, Provincia de Buenos Aires; [4] Universidad Nacional de 3 de Febrero, B1674, Sáenz Peña, Provincia de Buenos Aires; [5] Instituto Tecnológico Universitario, Universidad Nacional de Cuyo, M5500, Mendoza, Provincia de Mendoza; [6] Universidad Nacional de Chilecito, F5360, Chilecito, Provincia de la Rioja} República Argentina.

{antonio.castrolechtaler, antonio.foti, aarroyoarzubi, fbcarmona64, rfusario}@gmail.com; aoliveros@untref.edu.ar; jgarcia@itu.uncu.edu.ar

Abstract. The need to provide universal communication services to rural and sparsely populated areas has led to the search of new technologies to prevent further overloading of the frequency spectrum through the use of white spaces. These technologies are based on techniques known as ***Cognitive Radio*** and ***Software Defined Radio***. The frequencies used by digital TV systems through radio broadcasting seem to offer an opportunity to expand this kind of equipment that combines telecommunication techniques and computer applications by using specific software. Additionally, this IEEE 802.22 equipment's adopt widely known and accepted elements from standard 802.3 for their data link layer.

Key words: White Spaces, 802.22, Frequency Spectrum, Rural Communications, *Whitespace Alliance*.

1. Introduction

Funded by FONCyT – ANPCyT, the Private Community Networks Project [1] was created to research, test, and discuss different technologies to provide communication services to sparsely populated and remote communities. These areas have no commercial interest.

Through the work in [2], [3], [4], [5], [6], it was found that there are locations in Argentina with low population density (1000 inhabitants or even less in certain cases) that do not have 24-hour energy supply. Therefore, these areas are not covered by either landlines or mobile phone services. Hence, accessibility to data networks and internet connection are not available.

Many of these locations are found near the old railroad network, no longer in service. The layout makes it difficult for service providers to invest in communication links and take care of the problem. Our idea is to find alternative solutions to solve this problem at a reasonable cost.

Due to practical and economic considerations, we have analyzed and tested various wireless digital solutions. We were led to consider these alternatives based of on speed and feasibility considerations. Another criterion was based on the need to use frequencies that do not require prior approval from authorities or government bureaus. Additionally, we are trying not to depend on a local broadcasting operator.

Our work included prior experiences in which coverage of significantly remote locations was analyzed. These locations require solutions that are different from those required for big cities, which are meant for short distances, as was originally the case in technology 802.11, [7]

Simultaneously, and due to the high overload of the frequency spectrum, a way to use it more efficiently is currently being researched at national and international levels. To this purpose, new state of the art technologies have been developed, as is the case of those equipment's existing in the market that comply with IEEE Recommendation 802.22 [8]

Modern societies depend on and use increasingly more the radioelectric spectrum. The omnipresence of wireless services and communication devices, such as mobile phones, police communications, Wi-Fi and the recent broadcasting of Open HD Digital TV, among others, illustrate this dependency. The frequency spectrum has become one of the most requested and scarce resources of modern times.

During the last four years, global demand for mobile data traffic has increased at rates that in many cases exceed 100%. The expected growth rate is even greater [9]. It is currently estimated at 134 Exabyte[1] per year, which means a year-over-year increase of 66% for the period 2012-2017 [10].

The intensive future use of the spectrum up to 10 GHz or even higher frequencies has led to a review of the regulatory policies. Due to the overload of these frequencies, the study of the so-called white spaces has become more intensive.

CEPT, the European Conference of Postal and Telecommunications Administrations, has defined the white space as: "***A portion of the spectrum that is available to be used for a radio communication application at a given moment, in a specific geographic area, simultaneously with another one, characterized by not interfering with other services having higher national priority to access those same frequencies*** "[11].

Great research efforts are being undertaken by NGOs, different countries and telecommunication companies in order to take advantage of this part of the spectrum that is currently unused. To this purpose, the Whitespace Alliance has been created with the mission to "***Promote the development and***

[1] EB (Exabyte) = 1.000 PB (Petabyte) = 1.000.000 TB (Terabyte).

widespread use of standards based on products and services as a medium to provide broadband capacity through existing whitespaces in the frequency spectrum".

In view of the strong investment and deployment, that Argentina has recently made for the creation and installation of the Open Digital Television system (TDA, in Spanish) the existing infrastructure offers an unparalleled opportunity to solve the problem of rural or remote locations through the use of white spaces.

2. Current situation

Globally, two out of three people have no access to internet and more than half of the world population lives in rural areas with no broadband access[2]. On the one hand, it is expensive to wire up rural or remote areas with copper or optic fiber, especially in those places with low population density. Although satellite solutions are possible, their installation, service and maintenance costs are not reasonable considering the potential users of such services. Therefore, radioelectric solutions seem to be the most viable.

In many countries, providers of traditional wireless service have focused on urban areas where the high population density guarantees a quick return for their investments. However, even if investments are made at a loss, there is the additional problem of determining and allocating a portion of the frequency spectrum for their exploitation [12]. It is not a minor problem because the spectrum is always a scarce commodity.

In addition, existing technologies have not been successful in providing coverage of significant reach using radioelectric means, most where there is not a direct line of vision. The experience at Corral de Lorca [5] [6], using Recommendation 802.11, albeit interesting, proved that green barriers created by wooded areas of a certain height together with the distances involved conspired against the possibility of obtaining good signal levels to provide a continuous service.

Precisely, one of the reasons for the creation of the White Spaces Alliance was to promote and find solutions to transform this digital gap into an opportunity through the utilization of unused or underused frequency spectrum. It also aimed to adopt new bandwidth technologies to provide connectivity at a reasonable cost, facilitating the use of white spaces and assisting in the implementation of the systems interoperability.

On the other hand, in many countries such as Argentina, the switchover from analogic to digital TV could represent the opportunity to bridge the aforementioned gap. According to the aforementioned recommendation, by digitalizing each analogic TV channel it would be possible to obtain up to 5 signals of standard definition digital TV.

[2] http://www.internetworldstats.com/stats.htm.

The spectrum surplus, often referred to as ***"digital dividend"***, could be used to provide access to broadband, provided that no interferences are caused to those users who already have bands allocated to them by the regulatory body.

In addition, the channels used by TV stations in VHF / UHF bands have broadcasting characteristics that are highly favorable for long distance reach. Regulatory bodies of various countries are currently establishing the standards that will permit the unlicensed use of the spectrum generated by white spaces, provided that such use does not interfere with TV receivers.

The equipment's that are able to comply with such requirements use cognitive radio techniques and, when using the white spaces permitted by TV channels, they obtain a reach that is ten times bigger than WI-FI solutions in bandwidths above 1 GHz.

3. New technologies for using white spaces

3.1 Introduction

The need for equipment that can be used in white spaces has motivated different telecommunication companies and research groups in quest for viable solutions.

In many countries, rural areas are of particular economic interest because they supply crops and grains in different stages of production. They also represent a major source of basic exports. They represent a significant percentage of their GDP.

Recommendations in 802.XX include standards that regulate the operation of wireless communications. Some of these standards have been analyzed and evaluated during the aforementioned experience at Corral de Lorca.

Subsequent to standard 802.11, new technologies began to appear and expand as a result of the work of various research groups.

On July 1st, 2011 the standard IEEE 802.22 - "IEEE 802.22: Cognitive Wireless Regional Area Network - Medium Access Control (MAC) and Physical Layer (PHY). Specifications: Policies and Procedures for Operation in the TV Band[3]" was approved with the support of IEEE [8]'s LAN/MAN[4] Committee.

The standard includes an option to establish full duplex wireless links for distances between towers ranging from 30 to 70 km by using frequencies that are not restricted by government regulations.

3 "IEEE 802.22 - Wireless Cognitive Regional Area Network for Access Control to MAC Medium) and Physical Layer PHY). Specifications, Policies and Procedures for the operation of TV Bandwidths".

4 LAN: Local Area Network; MAN: Metropolitan Area Network.

The goal of the standard in series 802.XX[5] is to determine the criteria to deploy its multiple interoperable products, providing access to fixed bandwidth to various geographical areas, including specific sparsely populated regions in rural areas, and to avoid interference with those TV services that already operate with radiobroadcasting bands.

Currently, this is known as Wireless Regional Area Network (WRAN) and is meant to operate mainly as a means to access broadband services for private data networks located in rural areas.

3.2 Standard 802.22 General Characteristics

In addition to offering a solution to the problem that concerns us, this recommendation permits to solve two other problems that seriously affect the use of the frequency spectrum. One is the so-called white spaces and the other one is the interference between adjacent channels, both becoming increasingly more common due to the intensive use of wireless communications for all kinds of services.

From the viewpoint of information systems, these characteristics are very interesting due to the fact that they combine a communication problem – such as the adequate use of the spectrum – with the development of special software to enable the use of radio cognitive techniques.

This technique's reduce the interference that might be caused by other existing operators using the same frequencies and also enable geo-localization. This is accomplished by accessing a database of established services and, in order to detect the presence of other services through the detection spectrum. It is used together with another standard known as WRAN or by its IEEE.802.22.1 denomination.

WRAN systems use channels ranging from 54 to 862 MHz in VHF and UHF bands. The use of cognitive radio technologies enables the use of spaces located between two open TV channels, thus avoiding the interference of these services with TV stations. Both operate in the same bands.

The objective is to use the frequencies allocated for the transmission of national open TV through their integration to a system based on these standards. It would enable their use for rural communications and other similar cases.

These solutions shall require the redesign of norms and regulations related to radio-communications and mobile services, compression standards, and wireless service substitution through cable or satellite technologies. Undoubtedly, the solution to this problem shall involve dynamic access through cognitive radio technology, based on the 802.22 standard.

Currently, Cognitive Radio Technology (CRT) is considered to be one of the strongest possibilities to meet the increasing spectrum shortage. Its objective

5 Wireless Networks.

is to take advantage of underused frequencies, such as temporary gaps in primary signals and the different kinds of white spaces.

Although it is already being used for various applications, once it is further developed it will be capable of providing technologies for rural broadband, public safety and emergency response, urban frequencies, etc. This technology will also have a significant impact on dynamic detection and spectrum management.

3.3 Radio Defined Software

The explosive growth of the ways and means that people use to communicate using wireless devices has resulted in their being designed taking into account two key features: ***user friendliness***[6] and a convenient cost/service relationship. Both have become essential success factors.

A new technology known as "***Software Defined for Radio - SDR***[7]" provides flexibility and cost-effectiveness to final users and to service providers and product developers [14]. A special forum created under the name "Wireless Innovation Forum[8]", made up of researchers, telecommunication equipment manufacturers and different service operators is involved in the promotion of these new technologies.

This forum has defined the concept of Software Defined Radio as "***radio-communication equipment's in which a portion or all the physical layer functions are executed by software programs***".

The radio is a device that transmits or receives wireless signals by using a portion of the radio spectrum. Due to their characteristics, traditional radio devices based exclusively on hardware (i.e.: mixers, filters, amplifiers, modulators/demodulators and detectors) are limited by the fact that they can only be modified through physical intervention.

On the other hand, equipment using Software Defined Radio technology could execute many functions with specific software, either through a computer or through an embedded system.

It is not a new concept, but the capabilities generated as a result of the rapid evolution of digital electronics has made it possible for many processes that used to be feasible only in theory to be executed today with the sole use of these applications[15].

With this technology, the software has proved to be efficient at a relatively low cost. Additionally, improvements can be made through software updates. In many cases, the software manages all the functions required to operate the equipment, including the processing of the physical layer.

6 Could also be described as " ***intuitive management***".

7 Software Defined Radio.

8 http://www.wirelessinnovation.org/

3.4 Cognitive Radio[9]

At the end of the 90's, Joseph Mitola[10] and Gerald Maguire, researchers at the Royal Technology Institute of Stockholm, developed a technology they named ***Cognitive Radio.*** This was an improvement of a prior work they had presented regarding ***Software Defined Radio*** [15] [16].

Although Software Defined Radio has a great potential, it requires a significant amount of processing that could limit its flexibility, in addition to requiring an adequate network response.

Radio Cognitive technology consists in the introduction of embedded communications software using the language "***Radio Knowledge Representation Language – RKRL***".

This can be considered an intelligent and efficient system for radio communications and the functioning of protocols. It provides mechanisms based on intelligent technology that permit the optimization of the frequency spectrum.

As mentioned before, the use and allocation of frequencies in an overloaded spectrum is not perfect and white spaces are created, specially in bands used by open digital TV operators.

Among others, these were the reasons that led to the development of the radio cognitive technology for wireless communications with the purpose of detecting – and then using – those portions of the radiofrequency spectrum that were being used inefficiently, allowing to reuse them without causing interference to existing allocated services.

Through the allocation of variable frequency, this procedure enables other services to take advantage of Recommendation 802.22 to occupy white spaces

Cognitive Radio intelligent software analyzes the spectrum regularly searching for white spaces, detects the use allocated to each one, and then determines whether they can be reused. If so, the system changes the transmitter parameters based on the environmental interaction.

It has the necessary capacity and technology to capture or detect information from other radio equipment's that are operating in the same frequency. Through systems of dynamic programming, it can reconfigure the transmission frequency allowing to transmit and receive in various frequencies, as well as to use different transmission access technologies supported by its hardware design.

---

9 Mitola defines cognitive as "***the combination of intuitive, declarative and procedural knowledge within a system that is learning from its own experience***".

10 Joseph Mitola III was awarded a doctoral degree from this Institute for his thesis: ***Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio.***

3.5 Additional Requirements[11]

In order to operate in the same frequencies as other services, and at the same time to protect the transmissions of the main operators allocated to those frequencies, the standard includes a set of requirements that include: spectrum detection, geo-positioning service, access to database with information on the spectrum status, recording and tracking of all channels operating at a given time in a specific geographical area [9].

The standard has the capability of detecting and taking advantage of operating channels that could cause interference, such as: television broadcasting, wireless microphone transmissions, transmissions from safety devices such as wireless lighthouses, or even medical telemetry equipment that require protection from the local regulatory body.

4. 802.xx Recommendations

As stated above, 802.22 is within the set of 802.xx standards. IEEE's 802 LAN/MAN Standardization Committee has developed a broad and diverse family of wireless data communication standards.

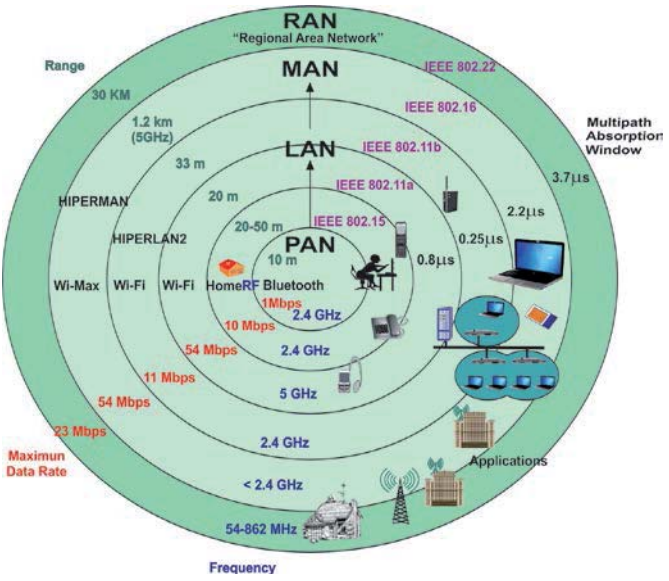


Figure 1. Different wireless standards developed by the IEEE 802 Committee

Different requirements for wireless communications have been addressed since the first 802.3.

11 Mitola defines cognitive as "***the combination of declrative and procedural knowledge within a system that is learning from its own experience***".

Figure 1 shows the most relevant standards and the relative position of standard 802.22.

It can be observed that different standards cover various radial points that can be small - up to 10 m - or large, such as the one being analyzed that covers up to 100 km.

5. Field Tests

The National Communications Regulatory Body – ENACOM – has recently approved the "Universal Service General Regulation" through Resolution 2642/2016 dated May 17, 2016. Said Regulation, under the title "Title III, Article 19, Programs, Section c. Connectivity in rural and unfavorable geographic areas for the development of ITC services" contemplates the development of applications using available modern radio-electric technologies, with the purpose of enabling the distribution of digital signals to convey voice, data and internet service to rural areas by taking partial advantage of the Argentine System of Land Digital TV (SATV-T) infrastructure.

Having become aware of our work, said regulatory body has invited this work group to perform field tests in order to verify whether these technologies could provide Universal Communications Service to rural and sparsely inhabited locations, as contemplated by the above mentioned regulation.

To this end, we are currently studying the products in the market, their costs, and availability, in order to begin with these field tests. We are already quite advanced in this matter and upon conclusion of this research; the available equipment's shall be purchased to perform the corresponding tests.

Similar tests are currently being conducted in Canada, United States, Singapore, Uruguay, South Africa, Kenya, India and the Philippines, among other countries.

6. Conclusions

Considering its characteristics and benefits, Standard IEEE 802.22 could be adequate to organize a system of rural communications. Meant for distances up to 100 km, this standard might cover largely the requirements of remote locations.

It is supported by a set of successful prior standards from which it takes numerous ideas, specially regarding the operation of the link layer since it adopts widely known and accepted elements from standard 802.3.

Therefore, the use of Standard IEEE 802.22 will enable to solve the communication problem of small and remote locations by taking advantage of the modern infrastructure offered by digital terrestrial television (DT)

As stated before, among possible techniques to optimize the use of the DDT bandwidth, it is worth mentioning those directly related to information systems known as Software Defined Radio (SDR). This technique consists in a portion or all the functions of radio-electric communication equipment's to be managed by computer software. This kind of technology opens an important path for I.T. specialists due to the significance that it will have for the development of communication equipment's.

.Cognitive Radio, the new system that appeared at the end of the 90's, is a variation of the prior system and consists in a truly intelligent system that enables, among other applications, to manage radio-electric communications with an optimum use of the frequency spectrum of digital broadcasting TV systems.

It is worth recalling that there are DDT stations that enable digital antennas to send digital signals to users' receivers. These stations transform digital signals into images and sounds that can be seen in any kind of screen. This is extremely important for the implementation of a rural communications system.

At the same time, the existence of a radio-broadcasting television system already in place should avoid having to use an additional portion of the frequency spectrum that, as stated above in this paper, is increasingly scarce and overloaded.

7. Future Work

We expect to continue studying the technical features of this recommendation in order to determine its limitations, if any.

It will be necessary to study the equipment being offered in the market for this standard and to analyze its capabilities. This will give an estimate of the required costs to cover different zones within the national territory.

Just as it was done at Corral de Lorca, it will be necessary to perform a field test to determine the real output of the equipment and the difficulty involved to deploy it.

Being mounted on shelters, the infrastructures of the DTT base stations will undoubtedly be useful to deploy part of the equipment's required by 802.22

References

[1] Antonio Castro Lechtaler (Director). PICTO 11-18621. Private Community Networks Project FONCyT, ANPCyT. Working Paper.

[2] J. Garcia Guibout, C. García Garino, A. Castro Lechtaler, R. Fusario and Guillermo Sevilla. Physical and Link Layer in Power Line Communications Technologies. Proceedings of 13th of Argentine

Congress on Computer Science. ISBN 978 - 950 – 656 – 109 – 3. pp. 56 a 67. Corrientes. October 2007.

[3] J. García Guibout, C. García Garino, A. Castro Lechtaler, R. Fusario and Guillermo Sevilla. Power Line Communications in the Electric Network. Proceedings of 13th of Argentine Congress on Computer Science ISBN 978 - 950 – 656 – 109 – 3. pp. 68 a 79. Corrientes. October 2007.

[4] J. García Guibout, C. García Garino, A. Castro Lechtaler and R. Fusario. Transmission voice over 802.11. Proceedings of 14th of Argentine Congress on Computer Science. ISBN 978 - 987 - 24611 - 0 - 2. pp. 307 a 318. Chilecito. October 2008.

[5] A. Castro Lechtaler, A. Foti, R. Fusario, C. García Garino and J. García Guibout. Communication Access to Small and Remote Communities: The Corral de Lorca Project. Proceedings of 15th of Argentine Congress on Computer Science. ISBN 978 - 897 - 24068 - 4 - 1. pp.1.117 a 1.126. Jujuy. October 2009.

[6] A. Castro Lechtaler, A. Foti, C. García Garino, J. García Guibout, R. Fusario and A. Arroyo Arzubi. Proyecto Corral de Lorca: A connectivity solution for small, isolated and distant locations. Proceedings from the IXth Latinamerican Conference on Cyber-Computer Systems: CISCI 2010. - Volume III - ISBN - 13: 978 – 1 – 934272 – 96 - 1. PP. 121a 127. Orlando, USA. June 2010.

[7] http://www.cplus.org/rmw/index.html (Radio mobile software).

[8] IEEE 802.22 - Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Policies and Procedures for Operation in the TV Bands.

[9] Carlos Cordeiro, Kiran Challapali, and Dagnachew Birru, Sai Shankar N. IEEE 802.22: An Introduction to the First Wireless Standard based on Cognitive Radios Journal of Communications, Vol. 1, N° 1, april 2006.

[10] http://www.cisco.com/c/es_es/about/press-2013/2013-02-05-trafico-global-de-datos-moviles-se-multiplicara-por-trece-en-2017.html.

[11] CEPT Report 24. A preliminary assessment of the feasibility of fitting new/future applications/services into non-harmonized spectrum of the digital dividend (namely the so-called "***white spaces***" between allotments. Report C from CEPT to the European Commission in response to the Mandate on: Technical considerations regarding harmonization options for the Digital Dividend. 1 July 2008.

[12] http://www.cse.wustl.edu/~jain/cse574-14/ftp/j_09wsp.pdf

[13] http://www.wirelessinnovation.org/introduction_to_sdr

[14] Dillinger, M; Madani, K; Alonistioti, N. Software Defined Radio: Architectures, Systems and Functions. Ed. Wiley & Sons, 2003.

[15] J. Mitola, G. Maguire. Cognitive radio: making software radios more personal. IEEE Personal Communications Magazine, vol. 6, N° 4, pp. 13–18, Aug. 1999.

[16] J. Mitola. Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio. Dissertation Submitted in Partial Fulfillment of the Degree of Doctor of Technology. Royal Institute of Technology - KTH Teleinformatics. ISSN 1403 – 5386. Sweden. May 8. 2000.

VIII

Innovation in Software Systems Workshop

InfoUNLP3D: An interactive experience for freshman students

FEDERICO CRISTINA [1], SEBASTIÁN DAPOTO[1], PABLO THOMAS[1], PATRICIA PESADO[1,2]

[1] Instituto de Investigación en Informática LIDI
Universidad Nacional de La Plata – Argentina
[2] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires - Argentina

{fcristina, sdapoto, pthomas, ppesado}@lidi.info.unlp.edu.ar

Abstract. The use of technology in educational environments is becoming a common standard. Current high-end mobile devices provide the necessary hardware to allow the execution of complex and resource-demanding 3D applications, at least in their simplest forms under low-end devices. Under this context, a multiplatform 3D mobile application is proposed, presenting a complete virtual scenario of the Faculty of Informatics (Universidad Nacional de La Plata, Argentina), with the main purpose of being an interactive guide for students. The application provides a navigational model of the building, several playing modes in order to get to know the facilities, and updated information about classrooms and courses.

Keywords: virtual buildings, 3D navigation, mobile devices

1. Introduction

Nowadays, the use of technology for helping purposes in everyday situations is something increasingly common. In this context, mobile devices such as cell phones, smart phones or tablets increase the possibilities of creating new technology to assist people under certain circumstances. This technology has the necessary potential to be also used in learning activities [1, 2, 3].
Additionally, the computational capacity of these devices have increased considerably in the past few years, allowing the execution of 3D applications with high levels of hardware requirements with ease. By doing this, it is now possible to create tridimensional experiences in which the user can even achieve a complete immersion within a virtual environment.
Thus, these two concepts (learning software and virtual 3D experiences) can be related in order to create a vast number of applications that aid users in their lives through easy-to-use and visually attractive solutions. One of these applications is the one proposed in the present paper. Its main goal is to

visually help students with their first steps at the faculty to locate classrooms and place themselves within the building.
The rest of the paper is organized as follows. Section 2 presents the motivations for the proposed work. Sections 3 to 5 describe the main stages in the creation of the interactive experience. Section 6 presents the results achieved after its publication and use by the students. Section 7 presents a list of possible related work.

2. Motivation

For freshman students, the initial days within the faculty can be a challenging and stressful experience. For the most part, a new and more independent way of life begins; all this changes might overwhelm them [4, 5].
In view of this situation, and despite the fact that the building provides all corresponding signaling and information offices, this is sometimes not enough for freshman students, who, in some cases, present interaction difficulties in their first days at the university.
Nowadays, most of the students own mobile devices. This is even more noticeable in computer science careers. They are familiarized with this kind of technology and easily master a large variety of mobile applications, which are a fundamental part of their daily life.
Therefore, a mobile application that assists students in their first stages at the university turns out to be very useful, helping them to become familiar with the building, its main areas and offices.
The proposed application consists of a mobile interactive 3D virtual model of the Faculty of Informatics, Universidad Nacional de La Plata, Argentina. Users can navigate the building or take virtual tours all around its facilities, in which several signs with the most relevants points of interest are displayed in an *augmented information style*. Additionally, information about classrooms and courses is shown while navigating the virtual scenario.

3. From blueprints to 3D building model

The first stage in the development of the interactive solution was the creation of a 3D model of the whole building. This task implied constructing the mesh based on the original faculty blueprints.
The faculty is a 3 stories building, containing dozens of classrooms, administrative offices, bathrooms, elevators, a library, a cafeteria, a green area and a parking lot; all of which had to be reproduced in the mesh.
In addition to blueprints, web mapping services like Google Maps [6] and Bing Maps [7] were used as reference in order to consider additional visual information not included in the blueprints such as roof color, trees placed inside the green area, building orientation, etc.

During the design, a special consideration was given to possible performance issues if the application would be run on slow mobile devices. Key aspects like the number of objects in the scene, number of polygons per object, or the textures used for the materials were taken into account with the goal of a correct balance between quality and performance.
Figure 1 shows an early stage of the modeling process. The blueprint is used as a reference in order to create the walls, floor, etc. The corresponding constructed 3D model is shown at figure 2.

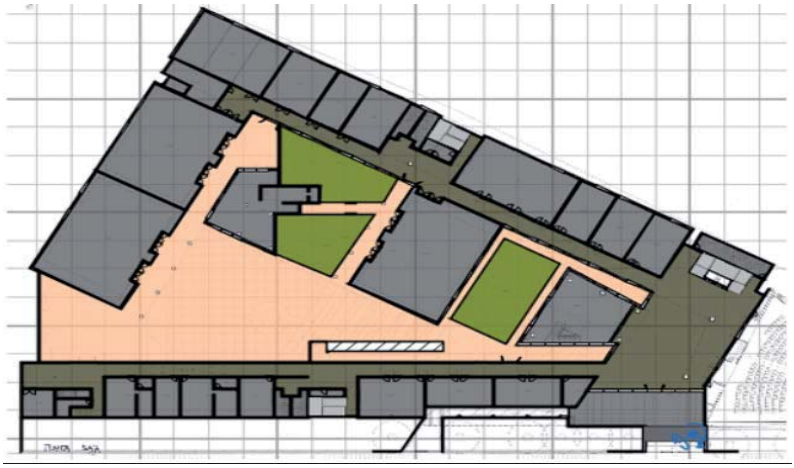


Fig. 1. Blueprint used as reference.

Fig. 2. The constructed 3D model.

4. Interactive experience

The main goal of the experience is to allow students to become acquainted with the differents areas of the faculty, and how to get to them.

This is achieved in several ways or playing modes available in the application:

Ï *General tour*: this option automatically navigates the whole building; starting from the main entrance, the tour presents a complete walkthrough though the hallways for each story, ending in a general bird's-eye view of the complex.

Ï *How to get to…*: in order to easily locate a bathroom, the cafeteria, etc. several preloaded navigations were included so that the student simply selects the location of interest, and the tour will explain how to get there from the main entrance.

Ï *Free roam*: the student will be able to navigate any area of the faculty in a first person perspective, virtually *walking* through the building hallways, stairs, etc.

These modes - in more extent the free roam one - requires an application with full user interaction support.

To achieve this goal, the application was created using Unity [8], a 3D development framework for creating interactive solutions, games, etc. Being a 3D solution, Unity supports the use of 3D meshes [9,10, 11, 12]. In particular, the model previously created in Sweet Home 3D [13] was exported and imported in Unity.

Once imported, a series of activities were necessary to adapt the 3D object for its correct use in Unity. The most important ones were the scale correction of the 3D model for proper handling and a mesh collider application on its members in order to run the simulation as expected in the free roam mode.

Together with the model, visual text hints were created as quick references for each place of interest, such as the library, bathrooms, etc. These hints are always displayed on foreground so that students can recognize these places along the use of the application, wherever they are placed.

The free roam mode required a special user interface, which allows the user to control the virtual character to move throughout the building. The mobile versions of the application present two virtual controls in order to move and change its orientation, while the desktop versions control the character by using keyboard and mouse.

Figure 3 shows the model already imported and adapted for Unity. Individual animations were created for each tour, setting camera position and rotation for each keyframe along time, as shown at the bottom of the image.

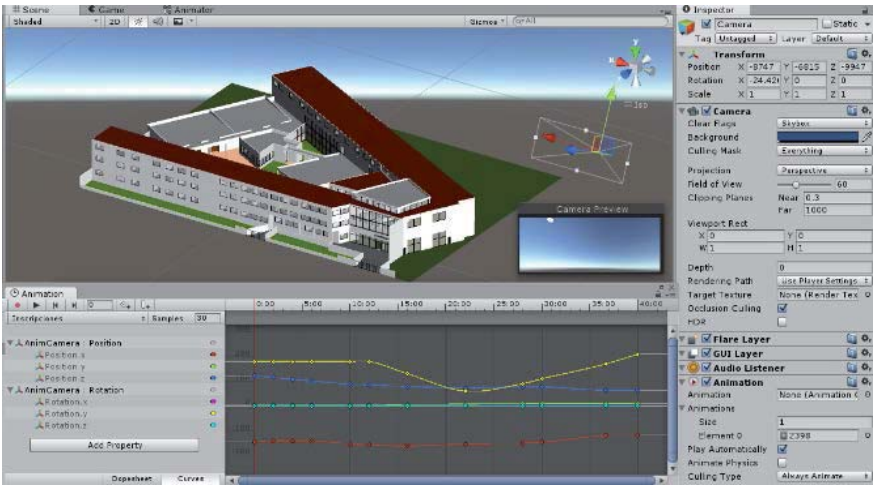


Fig. 3. The 3D model in Unity.

5. Integration with the classrooms information API

The *free roam* mode allows a useful additional feature. Every time the user is near a classroom, the daily schedule of courses for that classroom is displayed with information such as course name, professor in charge, type of course, hour of the day, duration, etc. The user can also browse the schedule through different days of the week, or even search courses for another quarter.

To achieve this functionality, the application interacts with an external system, also developed at the faculty: *Teachers Management System* [14], in particular with the *Classrooms Information API*. This API allows to query for a particular classroom information through a REST call [15].

The complete schedule for every classroom is prefetched and processed at the application startup time, so that all the information is already loaded by the time the user starts using the application. This eradicates any possibility of display lag when the user approaches a classroom.

Figure 4 shows an example of this feature. Every time the user moves towards a different classroom, the information will change accordingly.

Fig. 4. Detailed information about the courses schedule is displayed while the user walks near the classroom.

6. Results

Once the development stage was completed, the resulting application was released for distribution. The binaries are available in the *institutional* section of the faculty's official website [16].

The application was released under different builds for several platforms:

- An Android version for mobile devices such as cell phones or tablets.
- A Windows version for traditional desktop/notebook devices
- A Web version for any other type of device/operating system.

The main difference between the native versions (Android/Window) in contrast to the Web one is that the achieved performance is better in the former ones. New builds are planned to be created, such as distributions for iOS and Linux in order to include the whole range of users using native versions.

Students were notified about the application through the use of social networks like the faculty page at Facebook (around 3000 followers). A positive feedback was received by users, which *shared* and *liked* the publication [17].

The application's *General Tour* mode was even used at the inauguration of a new area of the faculty building [18, 19].

Figure 5 depicts the final application running on an LG device with Android OS. The different *playing modes* can be selected from the menu on the left, or any of the predefined destinations in the *How to get to…* mode. Virtual

controls are displayed along the screen for moving and rotating the camera in the *free roam* mode.

Fig. 5. The final application running on an LG device with Android OS.

7. Conclusions and future work

This paper presented a novel application that aids students to have a better knowledge of the faculty where they are starting their studies, through a virtual experience which shows the highlights of the building.
Several playing modes are available in order to enrich the experience, including a free-roam mode that allows a completely free navigation of the installations.
The application is available for several platforms, spanning a wide range of desktop and mobiles devices. There is also a web-based version in order to extend its use to any type of device.
The application was well received by students and general users, who gave positive feedback about it and even suggested new features, which now are the main objectives planned for future work.
First, a mid-term goal is to expand the interactive experience increasing the augmented information displayed in the application, which will be useful not only for students but also for teachers. For instance, when a user is positioned in front of a room, the system will provide related information such as:

- Room capacity
- Room equipment (computer, projector, etc.)
- Number of students

Second, a long-term goal is to extend the virtual scenario with new models from other faculties, so that the whole university will be included in the experience. In this way, the application will not be targeted only for informatics, but for a wider range of students from any other area.

References

1. Cristina, F.; Dapoto, S.; Thomas, P.; Pesado, P. "A simplified multiplatform communication framework for mobile applications". IEEE International Conference on Computer Engineering & Systems (ICCES). December 2014. ISBN 978-1-4799-6593-9.
2. Cristina, F.; Dapoto, S.; Thomas, P.; Pesado, P. "Prototipo móvil 3D para el aprendizaje de algoritmos básicos". XXI Congreso Argentino de Ciencias de la Computación CACIC. October 2015.
3. Goldin C.; Katz L. "The Race between Education and Technology". 2010. ISBN-13: 978-0674035300
4. Caballero Z.; Gómez M.; Borgobello A.; Ciarla D. "Una experiencia de taller para alumnos pre-ingresantes". 2004. Facultad de Psicología. UNR.
5. Romero H. "Características de los ingresantes a la universidad". V Encuentro Nacional y II Latinoamericano La Universidad como objeto de investigación. 2007. ISBN 978-950-658-187-9.
6. Google Maps. http://google.com/maps
7. Bing Maps. http://bing.com/maps
8. Unity online manual. http://docs.unity3d.com/Manual/index.html
9. Nystrom R. "Game Programming Patterns". 2014. ISBN-13: 987-0990582908
10. Hocking J. "Unity in Action: Multiplatform Game Development in C# with Unity 5. 1st Edition". 2015. ISBN-13: 978-1617292323
11. Smith M. "Unity 5.x Cookbook". 2015. ISBN-13: 978-1784391362
12. Linowes J. "Unity Virtual Reality Projects". 2015. ISBN-13: 978-1783988556
13. Sweet Home 3D user's guide. http://www.sweethome3d.com/userGuide.jsp
14. Sistema de Gestión Docente. Facultad de Informática. Universidad Nacional de La Plata. http://gestiondocente.info.unlp.edu.ar/
15. Leonard Richardson, Sam Ruby. "RESTful Web Services". O'Reilly. 2007. ISBN: 978-0-596-52926-0
16. Virtual tour access links at UNLP official site. November 2015. http://www.info.unlp.edu.ar/index.php/institucional/recorrido-virtual
17. UNLP official Facebook Page. Virtual Tour news publication. November 2015. https://www.facebook.com/InfoUNLP/posts/946207192138632
18. "Avanzan las obras de ampliación del edificio de Informática". YouTube video. December 2015. https://www.youtube.com/watch?v=9DSdWxuswJk
19. "Noticias UNLP - Informática empezará 2016 estrenando instalaciones". YouTube video. December 2015. https://www.youtube.com/watch?v=m7esG8Z63sk

Knowledge Based Augmented Card System for Medical Assistance Over Mobile Devices

CRISTIAN MONTALVO, FACUNDO PETROLO, DIEGO SANZ, NAHUEL MANGIARUA, NICOLÁS VERDICCHIO, SANTIAGO IGARZA, JORGE IERACHE

[1] Grupo de Investigación en Realidad Aumentada Aplicada, Departamento de Ingeniería e Investigaciones Tecnológicas UNLaM. Universidad Nacional de La Matanza, Av. Florencio Varela 1903 (B1754JEC) San Justo, Provincia de Buenos Aires, República Argentina.
jierache@unlam.edu.ar

Abstract. This paper describes a proposed integration between Augmented Reality and Knowledge Based systems in the context of a personal id card for emergency health care. We present the initial results of augmenting physical ids such as cards and tags, with health relevant information and a preliminary categorization of the user's condition given by a Knowledge Based system.

Keywords: medical information, augmented reality, medical emergency.

1. Introduction

The Knowledge Based Augmented Card System for Medical Assistance makes use of Augmented Reality, Knowledge Based systems, mobile devices and physical id elements with the purpose of improving emergency assistance of a patient.

This system can be used in different contexts: 1) in a medical facility, by the personnel to have access to a patient medical history; 2) in an emergency vehicle, by the paramedics for quick visualization of relevant information such as allergies; 3) in the streets, by a citizen, for access to basic id information and relevant contacts.

Currently, there are several physical mediums for patient identification which can be unequivocally linked to medical history systems and processed by AR and KB systems. In particular, Knowledge Based systems demonstrated to be effective for dealing with problem solving and efficient for handling repetitive, routinary tasks. In consideration with the classic discussion of AI replacing human labor, we consider KB systems to be a complement to human experts when dealing with complex problems [1].

Augmented Reality is usually associated with Virtual Reality but they are different in principle. While VR seeks to introduce the user in a completely new, virtual environment, AR complements, augments the physical reality, allowing the user to keep its senses connected to it while perceiving some additional virtual elements [2]. AR does not replace reality, it overlaps virtual information [3]. A combination of newly available information sources in

addition to the sensors and processing power included in modern handheld devices have become the angular stone for AR systems [4].

This characteristic feature of AR is exploited by our system to quickly display personal and medical information relevant for decision making during an emergency, augmentating the physical ids of the affected individual.

For this purpose our system is composed of several components: a) a physical id such as a card or tag; b) a Marker, a printed image adhered to the physical id, displaying a code which triggers the AR content; c) a handheld device such as a tablet or smartphone which captures the Marker with a camera and displays augmented information in the screen; d) categorization, knowledge based system which classifies the patient into one of three possible states. Figure 1 shows a conceptual diagram of the system.

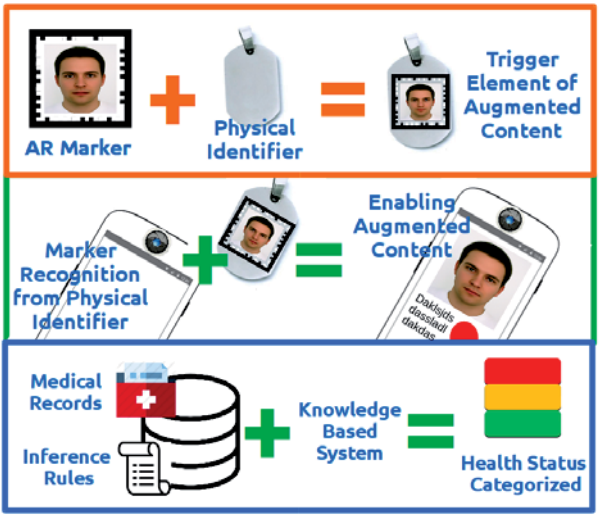


Fig. 1. Conformación del Sistema.

2. Knowledge Based Augmented Card System for Medical Assistance

As shown in Figure 2 and briefly described in [5], the system was implemented in three modules, the Web, Mobile and Main modules. The Main module acts as the system server, storing and handling communication for the mobile and web modules as well as hosting the knowledge based sub-system. It exposes a REST API [6] and was developed with Spring IO [7] and runs over Apache Tomcat [8]. Persistence is handled by JDBC [9] over MySQL [10]. The Web module handles the medical data entry, generating the required Marker for the physical ids. It was developed using HTML and public libraries such as JQUERY [11] and Bootstrap [12] over a Java [13] backend. The Mobile module manipulates static and augmented reality

visualization elements, developed for Android devices using Unity3D[14] and Vuforia[15].

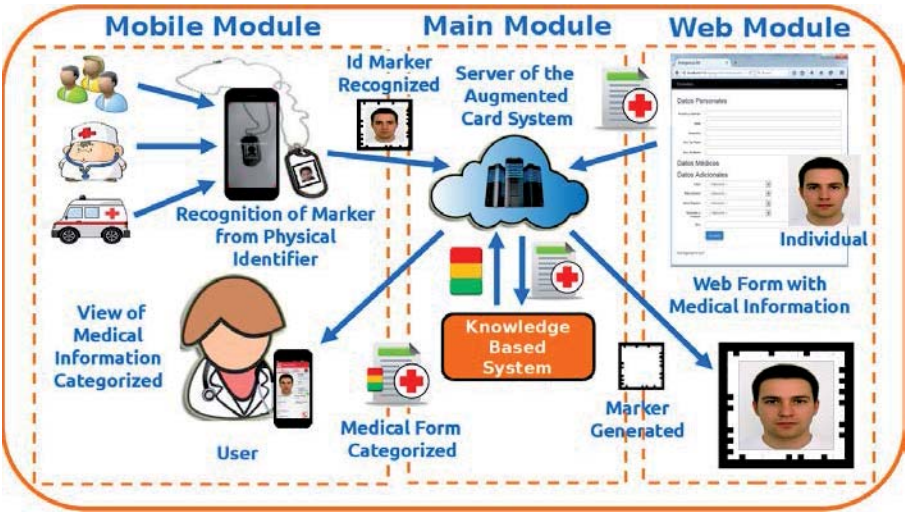


Fig. 2. Diagrama conceptual del Sistema.

In the following subsections we detail the implementation of each particular module.

a. **Main Module - Server Application and Knowledge Based System**

This module encompaces a RESTful API providing addition, deletion, modification as well as query services for the other two modules. It integrates a KB system for user health state categorization feeding from the medical history. Currently, most paramedics must work in the blind when dealing with an emergency. From this lack of information we identified the necessity to incorporate an intelligent system capable of alerting or categorizing the patient, complementing a quick presentation of the most relevant information for the assisting individuals. In Figure 3 we can see how the KB system uses the medical history and inference rules to generate an output state associated with a color. A knowledge map generated from the conceptualization of experts in the medical field is presented in Figure 4, identifying the most relevant attributes such as current medication, allergies, disabilities and past surgeries which in turn are pondered by a relevancy factor.

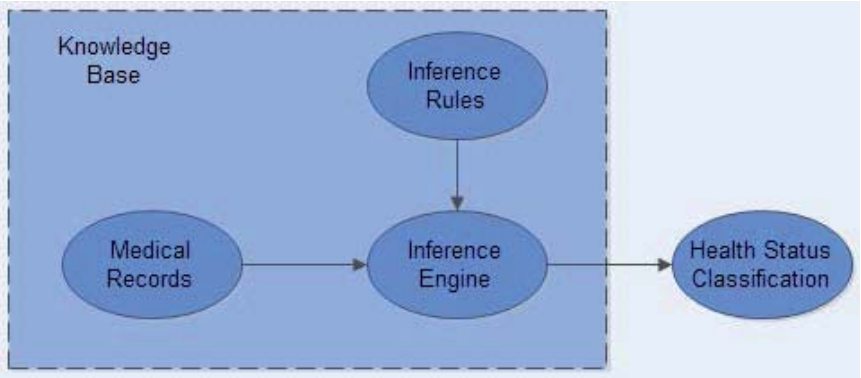


Fig. 3. Knowledge Based System

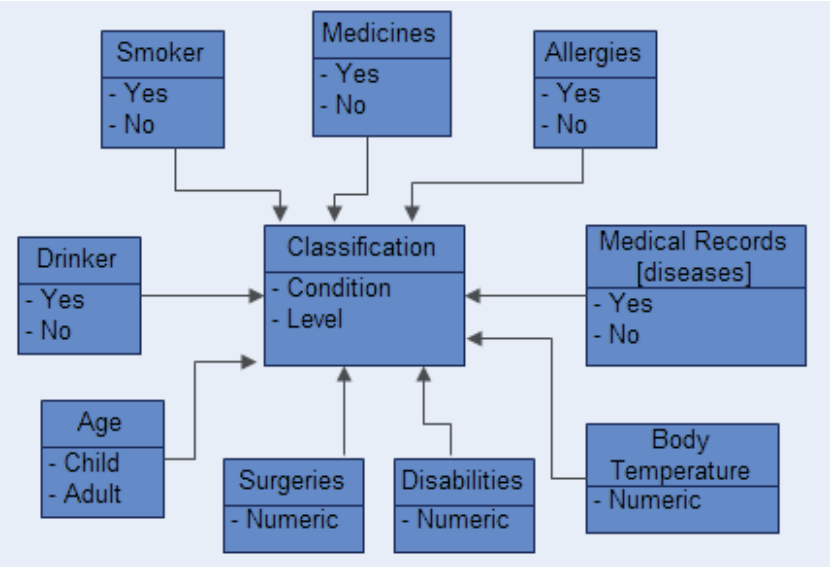


Fig. 4. Knowledge Map

b. **Web Module - Data Loading and Management**

This module provides an administrative user or medical staff a simple service in the form of a web page to load and manage the regular users medical history and personal information through web forms. Once the forms are complete, all information is submitted to the server (main module) and a Marker associated with the individual is returned to be downloaded and sent to the regular user to be attached to the physical id.

In Figure 5 we present a deployment diagram illustrating the components involved between the server and mobile modules. The mobile module has the

Marker, Catalog and Content factories, the Marker Recognition component and the Network Manager providing the connectivity. On the server (main+web) side several Controllers interact with a Persistence layer and the KB sub-system to provide the web module functionality and the REST API.

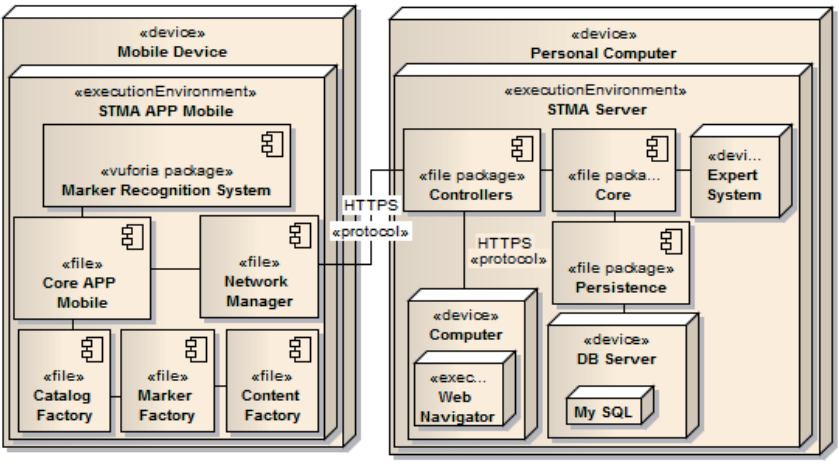


Fig. 5. Deployment diagram of the system.

c. Mobile Module - Mobile Application for the Exploitation of the Augmented Content

This module is responsible of the marker recognition, which has the physical identifier of the individual being assisted, and to display the augmented content of the application in the real world. Its workflow consist in the following: 1) If the user is previously registered at the system and he has a marker on his/her physical identification element (card) it is recognized by the mobile application. 2) Once the marker is recognized, a content download request is sent to the system server. 3) After completing the download of content, the application will be responsible for displaying useful personal data, relevant medical information and health status of the individual, on the screen of the mobile device. The figure 2 shows the basic cycle of exploitation of this information.

3. Preliminary tests and results

The figure 6 shows the mobile application running, it is focusing the marker on the physical element (identification card) of an individual. We can see the options of the augmented content and displayed on the device screen.

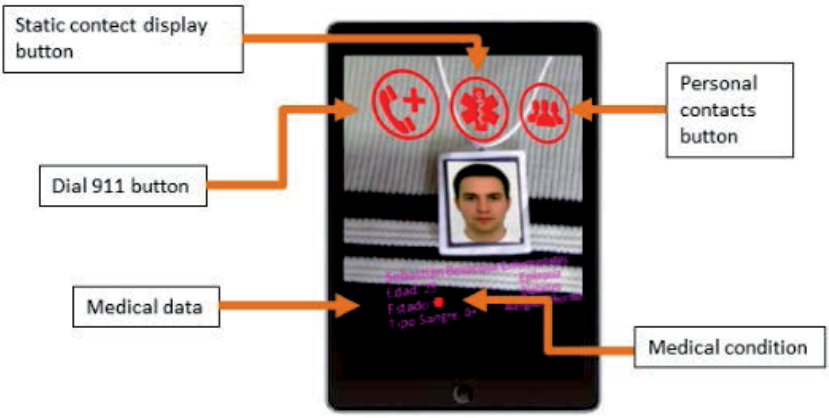


Fig. 6. Augmented Reality's view of Mobile Module.

The previous figure shows us that this application uses augmented reality. The button options that are presented on the screen of the device are: a) call the emergency service, b) call the individual's contacts, c) show relevant medical data of the individual, d) display contents in static form. The individual information is distributed in three sections, it is shown in figure 7. The first section provides information about the individual's personal data (name, age, blood type, allergies, health condition, previous illnesses) and the health status classification by the knowledge-based system, which is indicated with a circle in this case in color red (considerable background). The second section shows the individual's health history, where we can see details about diseases, surgeries, allergies, operations, etc. Finally, the third section contains data of a complementary nature, for example: If a person is an organ donor, if he is a smoker, if he has a phobia, and so on.

Fig. 7. View of the three informative sections of Mobile Module.

In December of the year 2015, a group of students of the medical career of the National University of La Matanza used an initial version of the mobile application to recognize markers and to visualize relevant medical information of a group of individuals preloaded in the web module. Then, they completed a survey to evaluate the level of usability, response time, innovation and the intuitive interface provided by the application. Satisfactory results were obtained, which were published in [16] and can be visualized in Figure 8. A second instance of tests of this initial version of mobile application was carried out in WICC 2016 [17], presenting the same survey to the participants of the event, Figure 9 shows the results of the event.

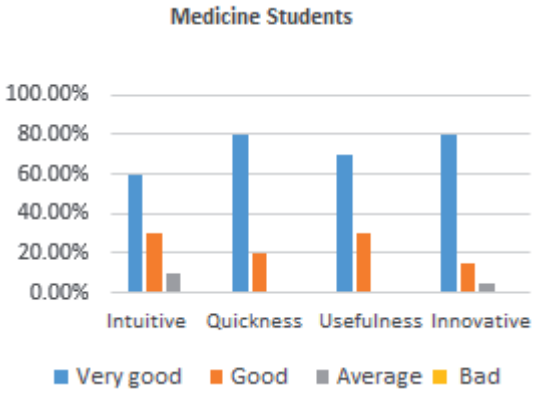


Fig. 8. Results of survey from medical career students.

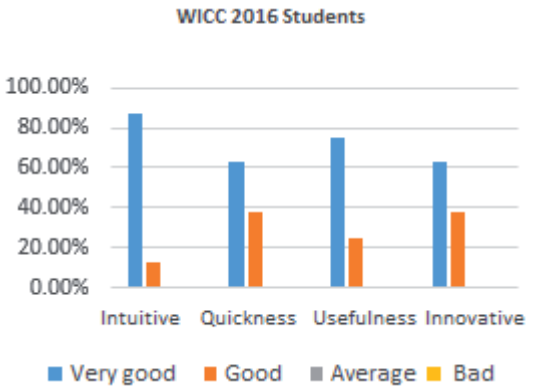


Fig. 9. Results of WICC 2016 survey.

4. Conclusion

The system contributes to the care of a particular patient in an emergency situations and provides categorization according to their health history through the use of a knowledge-based system. It has potential to improve the time of attention to health care situations in terms of efficiency and quality, providing relevant information through the use of augmented reality technology. In addition, it can help medical staff to get a first impression of the patient's history. Future developments are designed to expand system functionality with comprehensive capabilities that reflect individual parameters (pressure, pulsations, sweat, etc.), integrated through your mobile device.

Acknowledgement. This work is funded by the PROINCE program of the department of engineering and technological research of the National University of La Matanza (UNLaM), based on the initial requirement posed by Prof. Dr. Daniel Eduardo Martínez.

References

1. Gisselle Rey Salazar, Alex García Araya: Sistema Experto para determinar tipo de diabetes. pp. 290-294. (2007);
2. Manresa Yee, M. Abásolo, R Más Sansó and M Vénere:. Realidad virtual y realidad aumentada. Interfaces avanzadas. (2011)
3. Jorge Mario Gaviria Hincapie1, Guillermo Alonso Castaño Perez, Byron PortillaRosero, Jose León Sierra Ospina: SLD203 Realidad Aumentada En el Tratamiento de las Enfermedades Mentales y las Adicciones. XV edition of the Convention and International Information Fair (2013).
4. Pablo J. Iuliano, Claudia A. Queiruga, Francisco J. Diaz: UNLP Aumentada: Desafíos y Retos. In: Biennial Congress of Argentina (ARGENCON), 2014 IEEE pp. 43-18. (2014)
5. J. Ierache, N. Mangiarua, N. Verdicchio, D. Sanz, C. Montalvo, F. Petrolo and S. Igarza, "Augmented. Card System Based on Knowledge for Medical Emergency Assistance". I IEEE CACIDI Congreso Argentino de Ciencias de la Informática y Desarrollos de la Investigación Dic 2016 IEEE Xplore Digital Library (en prensa) ISBN 978-1-5090-2938-9 2016
6. APIRest, http://www.restapitutorial.com/lessons/whatisrest.html.
7. Spring IO, https://spring.io/
8. Apache Tomcat, http://www.tomcat.apache.org/
9. JDBC, http://www.oracle.com/technetwork/java/javase/jdbc/index.html
10. MySQL, https://www.mysql.com/
11. JQUERY, https://jquery.com/
12. Bootstrap, http://getbootstrap.com/
13. Java, https://www.java.com/
14. Unity3D, http://unity3d.com/es
15. Vuforia, https://developer.vuforia.com/
16. Jorge Ierache, Nicolas Verdicchio, Nicolas Duarte, Cristian Montalvo, Facundo Petrolo, Diego Sanz, Jonathan Barth, Nahuel Mangiarua, Santiago Igarza, "Augmented Reality Card System for Emergency Medical Services", IWBBIO

2016 (International Work-Conference on Bioinformatics and Biomedical Engineering) Proceedings Extended abstracts 20-22 abril 2016 Granada (SPAIN), pp.487-494, ISBN 978-84-16478-75-0.

17. Nicolás Verdicchio, Diego Sanz, Jonathan Barth, Cristian Montalvo, Facundo Petrolo, Nahuel Mangiarua, Santiago Igarza, Jorge Ierache, "Líneas de investigación de realidad aumentada aplicada a la asistencia médica en el campo de la emergentología", XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina) pp.667-671, ISBN:978-950-698-377-2.

NMEA-0183 sentence processing for the analysis of satellite geometry using low cost GPS receivers

ALBERTO EDUARDO RIBA[1], JORGE DAMIÁN TEJADA[1], NELSON ACOSTA[2,3], JUAN MANUEL TOLOZA[2,3]

[1]Dept. of Basic and Applied, Universidad Nacional de Chilecito
9 de Julio 22, Chilecito, La Rioja, Argentina
{ariba, jtejada}@undec.edu.ar
[2]Universidad Nacional del Centro de la Provincia de Buenos Aires
General Pinto 399, Tandil, Buenos Aires, Argentina
{nacosta, jmtoloza}@exa.unicen.edu.ar
[3]Universidad Nacional de Tres de Febrero
Mosconi 2736 - Sáenz Peña (B1674AHF), Buenos Aires, Argentina

Abstract. NAVSTAR-GPS is currently the most widely used satellite navigation system. There is a great amount of low cost GPS receiver devices that can be used for different applications but these do not deliver the precision required. This article presents a tool that allows NMEA sentences obtained from one or various low cost GPS receivers to be analyzed in order to predict possible indicators of the relationship between satellite geometry and positional precision.

Keywords: GPS, NMEA-0183, satellite geometry, low cost, data fusion.

1. Introduction

Nowadays a large number of activities require navigation over the Earth crust and geo-positioning, some of these activities require more precision than others.

The current trend in geo-referencing lies in the use of GNSS systems (Global Navigation Satellite System) that use satellites to locate a receiver on earth based on triangulation techniques and measurements with delay of the signal [1].

There are currently two GNNS systems in fully operational conditions and with global coverage: NAVSTAR GPS developed by the US Defense Department and GLONASS which belongs to the Ministry of Defense of the Russian Federation. Two other systems are being developed and are estimated to be fully operational in the coming years: Beidou-2/COMPASS which belongs to the Chinese Government and Galileo which belongs to the European Union – European GNSS Agency.

NAVSTAR-GPS technology is the oldest and the most well-known system within an array of receivers ranging from low cost equipment for civilian use

to professional equipment that can cost thousands of dollars. The difference in the price lies in the precision of the position obtained.
Professional equipment such as the one used in cadastral institutions or the Star Fire GPS 3000 by John Deere [2] can reach centimeter-level precision but its price is very high while civilian use receivers are low cost and deliver a 10 to 15 meter level distance 95% of the times [3] depending on the manufacturer.
There exist a large number of applications that can carry out trouble-free tasks in terms of the obtained precision but some areas such as robotics, precision agriculture, air navigation, sea navigation, and rescue operations among others, need greater precision [2], [4], [5].
Currently there are different systems that allow positional precision to be augmented, for example DGPS (Differential GPS) [6], AGPS (Assisted GPS) [7], RTK (Real-Time Kinematic) [8] o e-Dif (extended Differential), but these augmentation systems are not available in certain regions; they are too expensive to acquire or to implement and, in many cases, its use implies an onerous monthly payment.
For these reasons, that are cost and availability, it is important to develop technological solutions to meet these needs. Several institutions face the daily challenge of finding new techniques to improve positioning precision, many of them with successful and verifiable results.
This work is within the framework of the research line presented by the authors in WICC 2016 called "Improvement of the positional precision using low cost GPS receivers". This line of research aims to reduce the gap between these developments and end users who need to perform tasks with greater positional precision than originally delivered by the NAVSTAR-GPS system using low cost GPS receivers.
The aim of this article is to present a tool that allows the processing of multiple files with information of NMEA-0183 standard sentences obtained from low cost GPS receivers to relate satellite geometry to positional precision.
The article is structured as follows: Section 2 describes the general aspects of the NMEA-0183 standard and points out the structure of the sentences to be used in the work. Section 3 presents the related works to this topic. Section 4 shows the stage of experimentation establishing the sampling points, proposed scenarios, and the tool for the analysis of the information corresponding to the geometry of the satellite. Finally, Section 5 presents conclusions and future prospects.

2. NMEA-0183 Standard

The NMEA-0183 standard interface (Standard for Interfacing Marine Electronic Devices) developed by the National Marine Electronics Association, defines the electric signaling requirements, the data transmission

protocol, the specific sentence format for a serial transmission at a speed of 4800 baud rate.
This standard supports one-way transmission from a single transmitter to one or more receivers. The transmitted data is encoded in 7-bit ASCII format and each line is a sentence that follows a well-defined format composed of comma-separated fields, which identify the type of information contained as position, speed, and depth among others.
Among the sentences that deliver the low cost GPS receivers in their outputs the following can be found:

- GPRMC (Recommended Minimum Specific): information referred to the position, latitude, longitude, date, time and magnetic variation, etc. for example:
  $GPRMC,194421,A,2909.1567,S,06730.2625,W,000.1,095.4,270716,001.3,W*68
- GPGGA (Global Positioning System Fixed Data): information referred to the amount of satellites in use, HDOP, altitude, etc. For example:
  $GPGGA,194421,2909.1567,S,06730.2625,W,1,07,0.9,1150.4,M,29.6,M,,*72
- GPGSA (GNSS DOP and Active Satellites): information referred to PRN (satellite identifier) in use, PDOP, VDOP, HDOP, etc. For example:
  $GPGSA,A,3,07,13,30,05,28,09,08,,,,,,1.6,0.9,1.3*36
- GPGSV (GNSS Satellites in View): information referred to satellites in view elevation, azimut, SNR, etc. For example:
  $GPGSV,3,1,12,07,35,139,44,13,21,232,33,30,59,167,43,05,44,238,36*71

For more detailed information related to the structure of these sentences, check the reference manual [9].
From the previously mentioned instructions the information will be processed according to:

- Date: snapshot date.
- Time: snapshot time.
- Latitude: latitude of the position in the format GGMM,MMMM (2909.1567).
- Longitude: longitude of the position in the format GGGMM,MMMM (06730.2625).
- Magnetic variation:
- Satellite in use: amount of satellites used in the solution of the position.
- Altitude: altitude of the position.
- Geoid: geoid value of the calculated position in the sample area.
- PRNs: An array of 12 values with the PRN of each satellite used in the position setting.
- PDOP: dilution of precision in the position.
- HDOP: dilution of precision in the horizontal component

- VDOP: dilution of the precision in the vertical component.
- Satellites in View: amount of satellites in view at the time of acquisition. For each satellite:
    - PRN: satellite identifier
    - Elev: elevation
    - Az: Azimut
    - SNR: Signal to noise ratio

3. Related Work

Of the set of works related to the topic, the following stand out.
Di Lecce, et al [10] use neural networks with a single GPS receiver achieving an improvement in position precision of up to 25%, this correction system can be applied in static and dynamic devices.
Toloza J.M. [11] presents a methodology and the study of different techniques and algorithms for the treatment of the information delivered by standard GPS receivers. He develops a tool that allows raw processing of the NMEA-0183 sentences of the GPS receivers used, for later treatment with the techniques and algorithms proposed, implementing a differential GPS system of relative positioning to improve the precision of the positions delivered by the GPS system originally. This tool is based on portability so that it can operate in regions where there are no augmentation services available to improve precision.
Schrader, et al. [12] suggest a tool to improve positional precision by using NMEA-0183 sentences in a system of multiple low cost GPS receivers connected to eight bit microcontrollers. They present two centralized and one decentralized hardware diagrams based on calculations related to the latitude and longitude average, without resorting to additional sensors or complex calculations to obtain the position.
Having read these works we have identified that in none of the proposals a study is made that relates the parameters of the satellite geometry (elevation and azimuth) with the positional precision obtained in the measurements.

4. Testing

Initially, for the development of the tool, different sets of samples with multiple GPS receivers were taken. These samples were collected on different days, times, and different places in order to have a heterogeneous sampling.

4.1 Scenarios

The chosen scenarios for the snapshots correspond to geodectic points belonging to the National Geographic Institute of the Argentine Republic

whose coordinates are known and precise, which will serve as a reference point to establish subsequent comparisons.
The snapshots were taken in the open air, without significant obstacles nearby, or objects that could cause alterations in the reception of signals that degrade the quality of the snapshots, which ensures a greater number of visible satellites. Samples were taken day and night taking into account that the ionosphere at night is less ionized.
During the experiments 360000 samples were collected from different data points with different amounts of GPS receivers. In those places where the snapshots were made with multiple receivers the data can be fused to perform different analyzes and calculations of the geometries and positions obtained.

4.2 Equipment that was used

In order to take the snapshots three Garmin GPS receivers connected to notebook were used. The price of these receivers does not exceed 100 dollars. The device includes a receiver and an embedded antenna. It can track several satellites (maximum 12) at the same time, providing navigation data updates with a frequency of 1 Hz. Its consumption is low and it also includes the ability to work with the WAAS (Wide Area Augmentation System) GPS differential augmentation system to increase precision to 3 meters. The problem that arises is that this augmentation system does not work in our region due to the absence of signal from the geostationary satellites. This receiver is designed to work under extreme conditions and is even water resistant and can be submerged at 1 meter for 30 minutes.
It has an internal flash memory that allows you to retain critical data such as the orbital parameters of the satellites, the last known position, date and time. It provides standard NMEA-0183 data output. Drivers for Windows and Macintosh are available. The working voltage is between 4.4 and 5.5 volts. The precision in normal mode, it is less than 15 meters 95% of the measurements, these are specifications provided by the manufacturer.

4.3 Proposed Tool

For the analysis of the data obtained from the different receivers a tool was developed that allows to take one or several files with NMEA format and to analyze the evolution of the geometry of the satellites. Each capture file is assigned a name and a shape with a color to be able to identify them in the graphic when they are consulted. Figure 1 shows the screen for selecting the files.

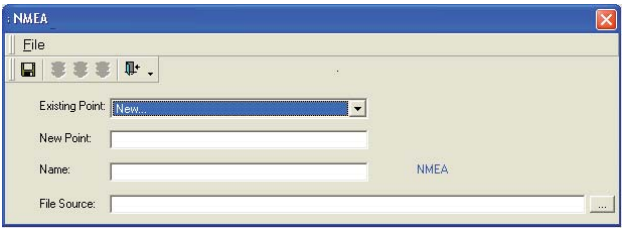


Fig 1. File selection interface.

Once imported the files with the sentences to be processed, different filters can be established such as date, start time and end time, number of satellites in use, signal noise ratio and precision dilution parameters, PDOP , VDOP and HDOP. These filters allow you to remove tuples that may contain errors.
Once the filters have been established the tool displays a grid with a list that details the time of the measurement with its latitude and longitude and by selecting a tuple a chart is shown with the sky map of the measurement as shown in Figure 2. Double click on a given row shows the rest of the information related to that measurement.

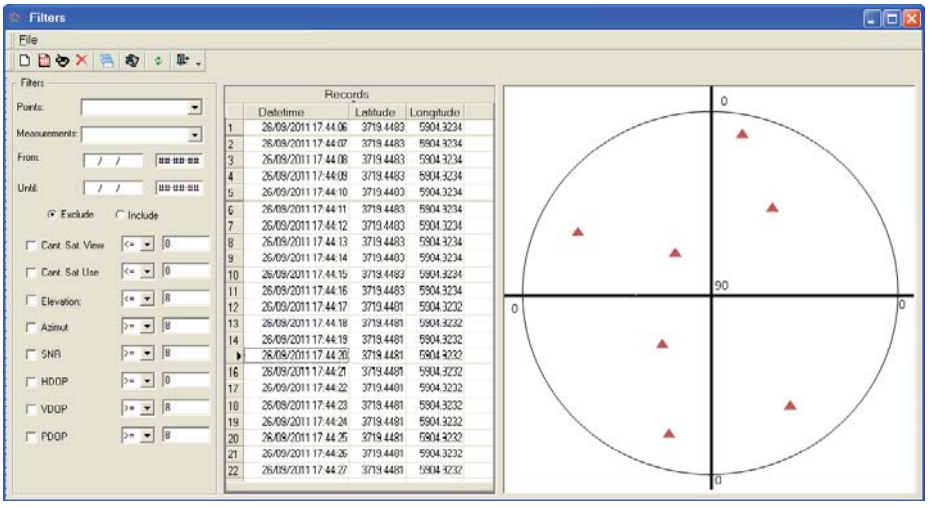


Fig. 2. Satellite geometry display interface

The estimated position is calculated through the arithmetic mean, mode and determination of the pair of points that is most repeated.
The tool presents an advanced interface for an expert user to personalize the information search according to certain criteria that cannot be included in the proposed filters.

5. Conclusion and Future Prospects

The proposed tool organizes and orders data from NMEA sentences allowing these data to be manipulated in a fast, dynamic and visual way making possible the analysis of the obtained information related to the satellite geometry.

The data processing based on the tests conducted made it possible to postulate possible indicators between the quality of satellite geometry and positional precision.

Satellite dispersion is one of these possible indicators and it is related to the elevation and azimuth parameters of each satellite tracked in the NMEA sentences obtained from the GPS receivers.

In future works, new indicators will be suggested in order to calculate a coefficient that will allow satellite dispersion to be tested to improve the positional precision delivered by GPS receivers without compromising computational cost so that is can be implemented on a microcontroller.

References

1. Gleason S., Gebre-Egziabher D.: GNSS Applications and Methods. Artech House, 508 pp. (2009).
2. Tomkiewicz, S. M., Fuller, M. R., Kie , J. G., Bates , K. K., Global positioning system and associated technologies in animal behaviour and ecological research, Philosophical Transactions of the Royal Society B: Biological Sciences 365 (1550) 2163–2176. (2010).
3. Arnold, Lisa. L., Zandbergen, Paul. A.: Positional accuracy of the Wide Area Augmentation System in consumer-grade GPS units. Computers & Geosciences Volume 37 Issue 7, Elsevier, pp. 883-892. (2011).
4. Cui , Y., Ge , S. S.: Autonomous vehicle positioning with gps in urban canyon environments, Robotics and Automation, IEEE Transactions on 19 (1) 250 15–25 (2003).
5. Elnabwy, M. T., Kaloop, M. R., Elbeltagi , E.: Talkha steel highway bridge monitoring and movement identification using rtk-gps technique, Measurement 46 (10) 4282–4292. (2013)
6. Clarke, Bill:#

11. Toloza, Juan Manuel. “Algoritmos y técnicas de tiempo real para el incremento de la precisión posicional relativa usando receptores GPS estándar”. SEDICI, Universidad Nacional de La Plata. (2012).
12. Schrader, D.K., Min , B-C., Matson , E.T.: Real-time averaging of position data from multiple GPS receivers, Measurement (2016).

VII

Signal Processing and Real-Time Systems Workshop

Functional Prototype of a Fall Detection System Based on the CIAA Platform

MATÍAS DELL'OSO[1], LAURA LANZARINI[1], PABLO RIDOLFI[2]

[1]Institute of Research in Computer Science LIDI (III LIDI), School of Computer Science, National University of La Plata
[2]Digital Processing Laboratory, Department of Electronic Engineering, Regional School of Buenos Aires, National Technology University
Argentina
{mdelloso, laural}@lidi.info.unlp.edu.ar
{pridolfi}@frba.utn.edu.ar

Summary. Falls are the main cause of injuries in people above 65 years old. In this paper, we describe the prototype of a device that can detect when an individual falls and send an alert to a monitoring center and/or family members. Even though there are currently similar solutions that are commercially available, these are not manufactured domestically. The device described in this paper will allow offering a low cost fall detector that will be extremely useful both for the elderly as well as for people with reduced mobility. Its purpose is helping reduce the adverse consequences that appear when a fall is not detected quickly, using a device that is fully developed in Argentina.

Keywords: Fall Detection, Teleservice, CIAA Project, Real Time, Embedded Systems.

1. Introduction

Thirty percent of people above 65 years of age fall once a year, and this number is constantly increasing because the population above this age threshold grows every year [1-3]. The WHO (World Health Organization) reported in 2012 that every year there are 37.3 million falls that require medical assistance, 424,000 of which result in death. This makes falls the second highest cause of death worldwide due to accidental injuries, following trauma caused by car accidents [4].

Some studies [5-8] show that this situation has similar rates of occurrence in various parts of the world. In Spain, for instance, a study carried out by the MAPFRE Foundation showed that 14.7% of adults above 65 years of age living alone suffer at least one fall each year [6]. In this study, 63.4% of the participants state that they have experienced only one fall, 11.3% indicate that they have fallen twice, 6.7% report having fallen three times, 0.8% have fallen four times, and 7.6% have fallen more than four times. As regards the places where falls occur, 45% fell outside of their homes, while 39.9% did so

in their own homes. In the United States, it is estimated that 1 every 3 elderly adults suffers some kind of fall every year. This translates as one fall every 13 seconds and one death every 20 minutes [7] [8].

In Argentina, an article published by CoKiBA (Professional Association of Kinesiologists of the Province of Buenos Aires) in 2014 [5] warns that 1 every 5 adults older than 65 suffers at least one fall a year, and that more than 80% of the episodes occur in their homes. Additionally, public hospitals in the Province of Buenos Aires report that 2 out of 5 adults over 80 years old have experienced at least one fall each year. Based on these statistics, the development of systems that help quickly detect falls by triggering an alarm for medical staff or family members, minimizing potential adverse consequences, is of the utmost importance.

Taking all this into account, in this paper we present the design and implementation of the prototype for a device that can identify if an individual has fallen and, in the event of a fall, will send an alert as applicable.

The remaining sections of this article are organized as follows: Section 2 provides an overview of the system, Section 3 includes the design details of the device, Section 4 discusses the tests carried out, and Section 5 presents the conclusions and future lines of work.

2. System Overview

This fall detector has been designed as a carry on device that is worn by individuals on their waists, clipped to their clothes or to a belt, and it measures movement acceleration periodically. It has a panic button that the user can press if they need assistance, and status indicators that indicate if an alert has been triggered or not. Two communication modules, WiFi and GSM, generate the required redundancy to ensure that the alert is sent even in case of failure (no signal, disconnection from the Internet, etc.). It also has an EEPROM memory board to store contact information (family members telephone numbers, central service server, etc.), a battery, and an accelerometer that will be used to detect falls (see Figure 1).

As regards operation, the device measures patient acceleration periodically. If a fall is detected (or if the user presses the panic button), the data send sequence is initiated.

As regards the software used for monitoring, a web interface was designed that allows monitoring the status of the devices carried by each patient. Also, if a fall is detected, a text message is sent to all telephone numbers recorded for the corresponding patient.

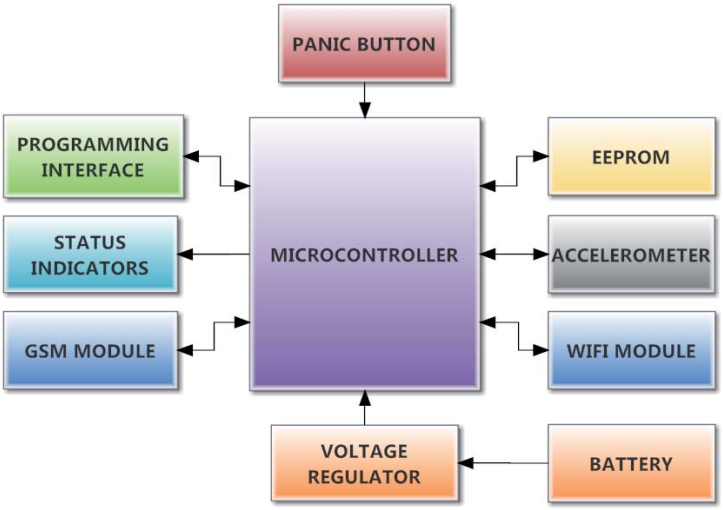


Fig. 1: Schematic of the fall detection device

3. Design and Implementation

3.1. Hardware Design

To detect a fall, the individual needs to carry a device that sends a help signal when necessary. When choosing this device, two options were considered: using a modern smartphone with all the features required for detecting falls and sending the corresponding alerts (processor, GSM and WiFi connection, accelerometer), or designing an embedded system. Finally, we decided to design an embedded system for the following reasons:

- The elderly are resistant to the use of mobile phones, which would mean that it would not have been possible to make sure that they carried the phone with them at all times.
- Mobile phones do not have a real-time operating system, so another, non-fall-detection-related task could potentially delay the activation of a critical task. Additionally, since there would be applications running in the background, battery life would be significantly affected.
- A mobile phone would not meet the necessary standards and certifications to develop an electronic product for the health sector.

However, the possibility of developing an application for mobile devices in the future is not ruled out, building on the algorithm implemented for the embedded system, to carry out performance comparisons or offer it as an add-on.

When selecting the micro-controller, we looked for a low-consumption model, but one that would at the same time have enough computation power to process complex algorithms in the future. Under these conditions, we selected the asymmetric, multi-core micro-controller LPC4337. The main advantage of using an asymmetric micro-controller is that, first, a lot of energy can be saved by using the low-power core for operations not requiring high computation power, activating the high-power core only when the program requires more complex calculations. Another reason for selecting the LPC4337 was our desire to include our work in the context of Project CIAA (Computadora Industrial Abierta Argentina, Argentine Open Industrial Computer) [9].

It should be noted that, for the design of this prototype, development board EDU-CIAA-NXP was used [10]. Also, both the accelerometer and the communication modules used are separate modules with the basic electronics required for operation. As future work, when developing the final product, the modules and the micro-controller will have to be integrated into a single printed circuit.

To connect the modules, push buttons, status indicators and EEPROM memory, an expansion board was designed for EDU-CIAA-NXP (see Figure 2) with the following: three status indicator LEDs, the EEPROM memory, three connectors to insert both communication modules and the accelerometer, a switching source to feed the EDU-CIAA with voltages above 5 V, and a linear, 3.3V regulator.

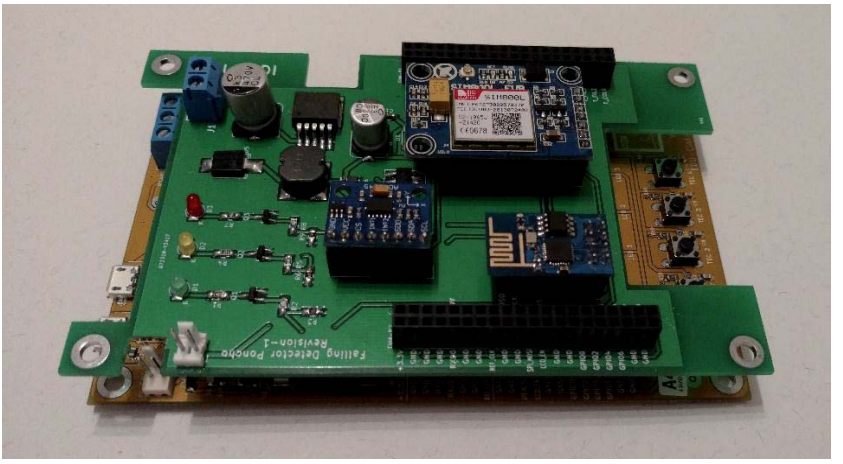


Fig. 2: Expansion board connected to EDU-CIAA-NXP

3.2. Fall Detection Algorithm

An algorithm based on thresholds that reads the accelerometer every 8 msec (125 Hz), processes the values and, if found to be above the threshold, starts

the process for sending alerts, was developed. In the following paragraphs, the algorithm used is detailed:

First, the thresholds to be used for detecting falls were established. To do that, using [11] [12] as reference, a first threshold of 6g was defined for the acceleration magnitude vector on the three axes (VMA), calculated with (1.1).

$$VMA = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad \textbf{(1.1)}$$

Where ax, ay and az are, respectively, the accelerations on axes X, Y and Z, taken as shown in Figure 3.

The second threshold was set at 2g for the sum of accelerations on plane XZ (VMP), calculated with (1.2). Plane Y was not included because, based on tests carried out, large acceleration values along this axis were detected while individuals were running or if they sat down abruptly, and these situations would have triggered false alarm signals.

$$VMP = \sqrt{a_x^2 + a_z^2} \quad \textbf{(1.2)}$$

The third threshold was set at 1.5 $\frac{m}{s}$ for the velocity (V0) of the person at the time they touch the ground, calculated using (1.3).

$$V_0 = V_1 + \int_{t1}^{t0} VMA(t)\, dt \quad \textbf{(1.3)}$$

To solve this integration operation, the acceleration measured on all axes (VMA) is backward integrated from t1 (1500 msec after initial contact) to t0 (point in time when the person touches the ground), which gives the times to be used for the integration operation. The value of 1500 msec was empirically obtained since, in all the intentional falls carried out, no significant changes were observed in the acceleration values on any of the coordinate axes after one second and a half.

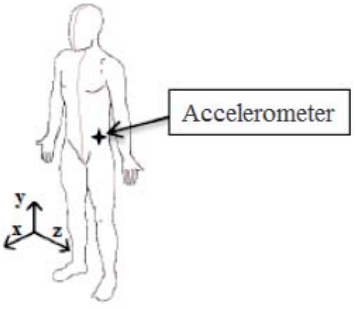


Fig. 3: Placement of the accelerometer on the body

To calculate the value of the integral, the composite Simpson numerical integration method was used [13]. The value of n (number of subintervals) was set at 94, which is half the number of samples captured during the integration period (188 samples). This value is aimed to achieve a midpoint between calculation approximation quality and computation to be performed by the microcontroller. As future work, tests with different values of n remain to be carried out.

The system will detect falls when the first threshold is exceeded or when the second and third thresholds are simultaneously exceeded, as shown in Figure 4.

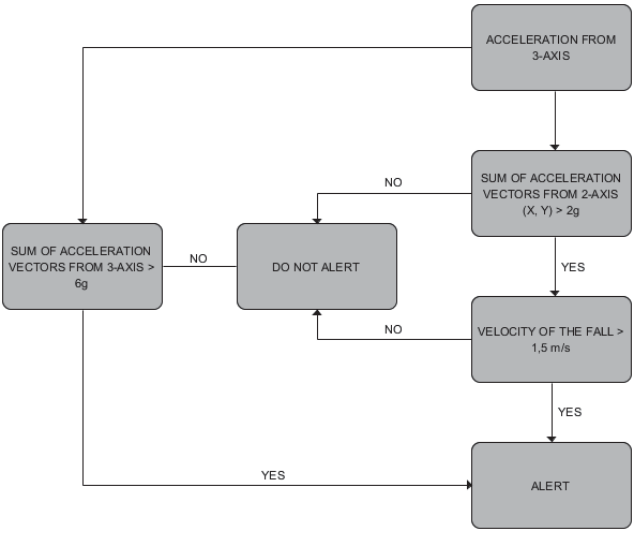


Fig. 4: Block diagram of the algorithm used to detect falls

4. Tests and Results

In the following sections, some of the tests carried out to check and validate the correct operation of the modules used and the system as a whole are presented. Both individual-component and whole-system tests were carried out using the expansion board for EDU-CIAA-NXP mentioned in Section 3.1.

4.1 Individual Testing of Communication Modules and Accelerometer

To check that the resolution selected for the accelerometer (±8g) was suitable to detect the movements of the individual wearing the device, a function was created in MATLAB that receives on its serial port the acceleration data on the three axes and creates the corresponding chart for any given period. To carry out this test, the device was placed on the waist of the individual and

several tests were run with the person performing sudden actions (jumping, running, shaking body, and so forth) for 8 seconds (1000 samples). As a result, it was observed that the acceleration on all axes is at all times lower than 8g. Figure 5 shows the acceleration chart on the three axes while the test subject was doing short jumps and shaking his body.

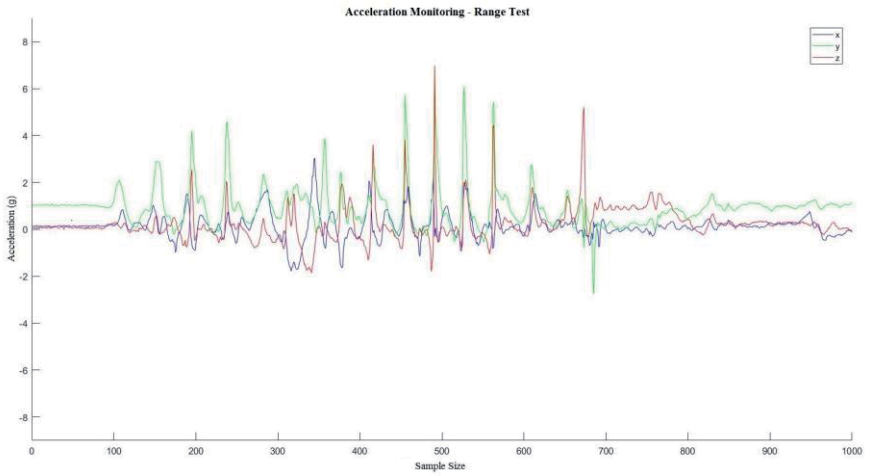


Fig. 5: Acceleration monitoring – Range test. Values obtained with the accelerometer when jumping and shaking the body

To check that communication is correctly established through the communication modules, a UART was used to monitor each command sent to the modules. First, the modules were disconnected; then, the emergency button was pushed; and finally, the modules were reconnected. Figure 6 shows the messages generated for the WiFi module. It can be seen that, even if communication fails, the device recovers and, after a given time, tries to reestablish communication.

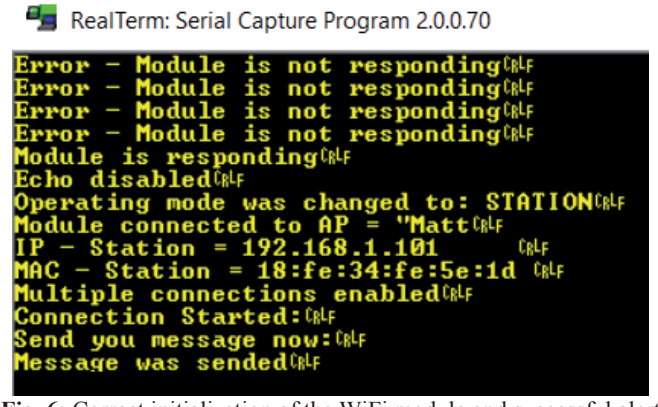


Fig. 6: Correct initialization of the WiFi module and successful alert

4.2. Testing the System as a Whole

To check if the system is working correctly, 4 different everyday life scenarios were recreated, and the behavior of the device observed in each of them. To that end, the function in MATLAB mentioned earlier was used to monitor the acceleration measured by the device in each scenario and, in the cases that involved a fall, we corroborated that the alert had been received both in the web page and the mobile phone linked to the device. In all cases, the fall detector was placed on the waist of the test subject. The different scenarios are described below:

i. **A person trips, falls to the ground, and remains lying on the floor.** This test involved falling on a mattress pad and watching the acceleration during the fall. Figure 7 shows that, when the fall occurs, the acceleration magnitude vector exceeds the predefined threshold (6g), which results in a fall being detected.

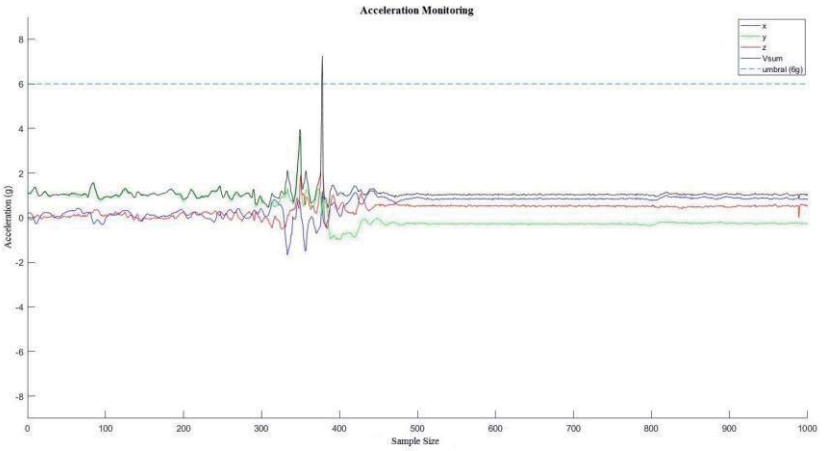


Fig. 7: Acceleration monitoring. Values obtained with the accelerometer when falling from a large height

ii. **A person goes up and down stair steps.** In this scenario, none of the thresholds was exceeded and, as a result, no falls were detected.

iii. **A person looses balance and, as he/she falls down, he/she grabs on to an object to avoid hitting the floor.** In this case, the device was not able to detect the fall. Since the test subject grabbed onto an object as he fell, his acceleration decreased and impact on the floor was not strong enough to exceed the preset thresholds. In this scenario, the threshold-based algorithm fails. It is for this reason that it is of the utmost importance to implement in the future a smart algorithm that can detect this type of falls. However, if the subject is conscious after

the fall, he/she will be able to push the alert button and trigger the help signal.

iv. **A person walks to a chair, sits down abruptly and then falls to the floor on his/her right.** In this scenario, the threshold set at 6g was not reached, but the device detected the fall because the velocity threshold is exceeded when the person touches the floor.

5. Conclusions and Future Work

The design and implementation of a prototype for a fall detection device based on the CIAA platform has been presented. Based on the tests carried out, its performance in identifying falls and sending alerts has been successful.

This is the first step in a highly promising direction, both in relation to human resource training as well as for the generation of qualified labor in Argentina and for social welfare. No comparisons with other existing solutions [14] [15] have been included because we consider that the contribution of this work is successfully developing the device proposed using national, low-cost technology, which limits device capabilities.

We are currently working on building a falls database not only to strengthen threshold selection to detect falls, but also to build a non-linear, adaptive model that allows improving the success rate of the device by recognizing any movement patterns that may be characteristic of the individual wearing the device.

As regards hardware, there is still work to be done to design the case, which must be water-resistant to allow wearing the device under the shower. Finally, we will look for ways to separate the data acquisition and processing modules from the GSM and WiFi communication modules to minimize consumption and prolong battery life.

Bibliography

[1] Deandrea, S., Lucenteforte, E., Bravi, F., Foschi, R., La Vecchia, C., Negri, E.: *Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis*. 2010; 658–668.

[2] Major Injury Hospitalizations Due to Unintentional Falls in Canada 2009–2010. Report.

[3] Peeters GM, Pluijm SM, van Schoor NM, Elders PJ, Bouter LM, Lips P. *Validation of the LASA fall risk profile for recurrent falling in older recent fallers*. J Clin Epidemiol 2010;63-1242.

[4] Organización Mundial de la Salud (2016, Apr 4). Centro de prensa [Online]. 2012. Available at: http://www.who.int/mediacentre/factsheets/fs344/es/

[5] Colegio de Kinesiólogos de la Provincia de Buenos Aires (2016, Jun 14). Centro de prensa. [Online]. Available at: http://www.cokiba.org.ar/web/?q=node/116

[6] Sáinz M. *Estudio de investigación sobre Seguridad en el domicilio de personas mayores* (2016, Jun 27). Fundación MAPFRE [Online]. 2008. Available at: http://www.mapfre.com/documentacion/publico/i18n/consulta/registro.cmd?id=128697

[7] Stevens J. A., Mack K. A., Paulozzi L. J., Ballesteros M. F. *Self-Reported Falls and Fall-Related Injuries Among Persons Aged >65 Years*. Morbidity and Mortality Weekly Report. Vol. 57. No. 9. 2008.

[8] Centers for Disease Control and Prevention (2016, Apr 2). *Important Facts About Falls* [Online]. Available at: http://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html

[9] Proyecto CIAA (2016, Jul 5). Computadora Industrial Abierta Argentina [Online]. Available at: http://www.proyecto-ciaa.com.ar/devwiki/doku.php

[10] Proyecto CIAA (2016, Jul 7). EDU-CIAA-NXP [Online]. Available at: http://www.proyecto-ciaa.com.ar/devwiki/doku.php?id=desarrollo:edu-ciaa:edu-ciaa-nxp

[11] Perry J.T., Kellog S., Vaidva S. M., Jong-Hoon Y., Hesham A. Sharif H. *Survey and evaluation of real-time fall detection approaches*. 6th International Symposium on High-Capacity Optical Networks and Enabling Technologies (HONET), 2009.

[12] Lindemann U. *Evaluation of a fall detector based on accelerometers: a pilot study*. *Medical & Biological Engineering & Computing*. Vol. 43. No 5.2005.

[13] Simpson's Rule (2016, Apr 7). Wolfram MathWorld [Online]. Available at: http://mathworld.wolfram.com/SimpsonsRule.html

[14] Bagalà, F., Becker, C., Cappello, A., Chiari, L., Aminian, K. *Evaluation of Accelerometer-Based Fall Detection Algorithms on Real-World Falls*. 2012.

[15] Rodriguez, J., Mercuri, M., Karsmakers, P., Soh, P.J., Leroux, P., Schreurs, D. *Automatic Fall Detector based on Sliding Window Principle*. WIC 2013.

Architecture and Implementation of A Low-Cost Prototype for On-Field Measuring of Goat Fibre Diameter

RAFAEL ZURITA, MIRIAM LECHNER, RODOLFO DEL CASTILLO, EDUARDO AISEN, EDUARDO GROSCLAUDE

Facultad de Informática, Facultad de Agronomía
Universidad Nacional del Comahue
Buenos Aires 1400, Neuquén, Argentina
{rafa, mtl, rdc, eduardo.aisen, oso}@fi.uncoma.edu.ar
http://faiweb.uncoma.edu.ar

Abstract. Goat raising is a customary economic activity in the north of the province of Neuquén. Lack of adequate technology makes producers' profits smaller. This article introduces the hardware and software architecture of a real prototype designed and developed for animal textile fibre classification. The system is to be used by the crianceros right out in their farm.
From a hardware point of view, one can envision the need of an embedded system. This aims at low production costs, robustness and mobility. From the software side, a linear order algorithm is used as the base for image processing. The prototype has been evaluated by taking measurements of animal fibre diameter and comparing the values obtained against a similar commercial instrument. The validation process shows a high correlation between results. Basing upon this experience and after a cost/benefit analysis, a final solution for the criancero is shown as possible.

Keywords: Embedded system, image processing, textile fibre.

1. Introduction

Small producers in the agroecological system in the north of Neuquén province are the transhumant communities (who call themselves *crianceros*) that raise goats. One of their most important income comes from textile animal fibre production [1]. However, they lack technology or professional advice to: i) breed selection to improve fibre quality, which would lead to better incomes; ii) objective fibre classification directly on the farm, optimizing marketing; iii) negotiate the right price based on objective fiber measurement, and market price; iv) choose best biotypes adapted to the nature of the place in order to stop, and eventually revert, the current desertification problem [2].

There exist several professional equipments for the measurement and characterization of fibre diameter (for wool fibre mainly): OFDA 2000, WoolView 20/20, LaserScan, air flow, and Lanámetro, among others [3, 4]. However, almost all of them are high-cost systems, and mostly suitable for laboratory use [5]. WoolView 20/20 is a tool for measuring wool fibre diameter on the farm, but the product has been discontinued [6]. This article presents the hardware and software architecture of a prototype equipment designed and developed by Universidad Nacional del Comahue (named "prototipo UNCOMA") for measuring animal fibre diameter. The system is to be used by the crianceros right out in their farm, so there were two goals to achieve as for usability: minimal amount of user practice required, and least possible preparation of fibre samples required.
The results show that the cost, performance and precision of the proposed "prototipo UNCOMA" is adequate for the objective selection at the origin of animals of high fiber production (especially cashmere), which would increase production per head and increase the annual income of the criancero. In this way, proposals for managing the animal load can be made effective, thus helping to revert the serious problem of desertification.
The remaining of this work is organised as follows. On section 2 the software and hardware architecture proposed are described, with emphasis on the measurement method by image processing. Section 3 exposes several experimental results in order to validate the process of "prototipo UNCOMA" to determine the diameters of fibres. Finally, section 4 discusses the conclusions and future works.

2. Architecture

2.1 Hardware Architecture

The hardware architecture consists of two main components:

- Mobile handpiece;
- Embedded hardware for image processing.

The mobile handpiece has a digital microscope, a LED for contrast, a pressing tool for holding the fibers, and a trigger mechanism. The 400x microscope was modified and calibrated; it requires 5v, and it is built into the handpiece as shown on Fig. 1.
Several tasks take place when the trigger is pulled. First, the acrylic plate in front of the microscope optic presses the fibres firmly, leaving them focused. Second, at the end of the trigger action, a digital switch is turned on, sending a signal to the embedded hardware over a digital I/O port. Finally, the embedded system will get the event signal, starting the execution of the image capture software.
At this point the image processing software running on the embedded hardware will request a capture of the now focused and held still fibre strands to be processed afterwards. The image is transferred from the handpiece to

the embedded hardware through a USB 2.0 interface, which interconnects both components.

The embedded hardware where the image processing software is located features a main Allwinner A10 SOC with a 700Mhz ARM CPU; 512MB of main memory (RAM); and an SD interface where a 2GB microSD memory was attached. The Allwinner SOC also has, internally, General Purpose I/O Ports (GPIO), a USB 2.0 port and, a Wireless WiFi 802.11b/g/n interface. During a reset, the SOC bootloader loads the OS from the microSD; which in turn will run the prototype application.

The resulting embedded hardware size is 10x4x3cm, and the handpiece is 20x20x5cm. Total equipment weight is 900g. The cost is determined by three main components:

1. 400x Microscope embedded on the handpiece: USD 60.
2. Embedded board with a 700Mhz+ ARM CPU and 512MB RAM: USD 30.
3. Custom made mechanical parts: USD 20.

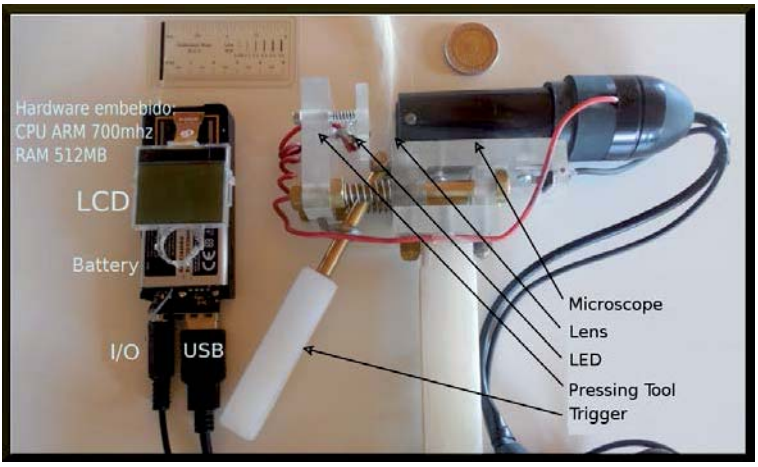


Fig. 1. prototipo UNCOMA.

2.2 Software architecture

The lowest level software layer is the GNU/Linux Debian OS which controls the embedded hardware. The main drivers from the Linux kernel which are used here, are: the universal serial bus (USB), the universal video class (UVC) and the General Purpose I/O port interface (GPIO). The digital signal sent by the handpiece is received by one of the GPIO ports. At this stage, the application, running in user space, will get an event notification driven by the signal, starting the fiber image capture and digital analysis process. Fig. 2 shows the general process.

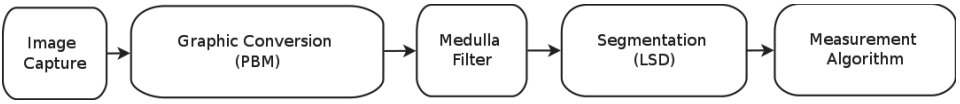


Fig. 2. Prototipo UNCOMA: whole process stages.

The Linux kernel exports to user space, through the UVC driver, tools which are used by the embedded application to capture an image. The UVC driver, which controls the microscope capture process, will obtain a complete image of the fiber on the pressing tool, and then will present it to the application in user space. As the Line Segment Detector algorithm (LSD) requires a portable gray scale (PGM) format as input, and the digital image received is in JPEG format, a conversion is required as a first step. In order to get a PGM image, the **netpbm** tool set is used. The image resolution is 640x480 pixels, where 1 pixel corresponds to 0,9743 microns, according to the optics and microscope selected for the prototype.

Each fiber strand has a medulla. Eventually, the image captured may expose the medulla, affecting the measurements. This medulla will be partially removed from the image using a filter, which computes the distance transform for pixels in a gray scale range. The range was previously determined using color values that may be expected for medulla. If the medulla detection is positive for a given pixel, the filter will replace the pixel color by one coincident with fiber (also previously established).

After that, the embedded application will use the LSD (Line Segment Detector) algorithm to obtain a set of line segments representing the original image [7]. This set will be used for fiber diameter measurement using the following algorithm:

```
# Pseudocode algorithm for fiber mean diameter calculus
# INPUT: line segment set (lsd_set)
# OUTPUT: measurements and mean diameter
    measrmt = 0;
    diameter = 0;
[1] FOR each segment IN lsd_set DO
      diameter_temp = 0;
      measrmt_t = 0;
[2]   FOR each dot (x,y) IN init_mean_final(segment) DO
        f() = perpendicular(segment, dot);
[3]     FOR each seg2 IN lsd_set DO
          IF parallel(segment, seg2) AND
            ( f() == perpendicular(seg2) ) AND
            ( f(FROM segmento TO seg2) == pixels_of_fibre() )
  THEN
[4]           diameter_temp = distance( f(FROM segment TO seg2)
);
              measrmt_t++;
          END IF
        END FOR
      END FOR
[5]   IF (measrmt_t >= 2) THEN
         diameter = diameter + diameter_temp;
```

```
        measrmt++;
        measrmt_set = add(diameter_temp);
     END FI
   END FOR
diameter = diameter / measrmnt;
RETURN measrmt_set, diameter
```

The algorithm searches for parallel line segments. These are identified as a fibre strand. If the pixel colors along a perpendicular segment running between those parallel lines are possibly fibre colors, then the length of that perpendicular segment is added to the diameter candidate set. If at least two perpendicular lines are candidates, then the parallel lines are considered fibre borders. Finally the mean diameter from the candidates is computed. With more precision:

1. For each line segment in LSD output, three points (endpoints and midpoint) are determined.
2. For each point the perpendicular line function is computed.
3. The algorithm then searches, among all line segments taken from LSD output, for a line segment being parallel to the line segment found in stage 1. Also, there must be an intersection with the perpendicular in 2. If there are not-fibre-colored pixels in the perpendicular between parallels, then the line segment is discarded.
4. If the line segment is not discarded, then the parallel line segment in 3. is considered to be the opposite border of the same fibre as line segment in 1., so the distance between parallel line segments is calculated as a candidate diameter.
5. If there are at least two points in line segment 1. with computed distances, then their values are valid measurements for the general mean diameter statistics. Otherwise, the values are considered noise and discarded.

Finally, when the processing of line segments finishes, several statistics are calculated. The statistics are: the average of the measured diameters (mean), standard deviation and variance. Based upon that information some useful values for the final user are shown on the LCD display as follows:

```
[1] mean;
[2] percentage of measurements that are under 30 microns;
[3] percentage of measurements between 17um and 30um;
[4] percentage of measurements that are above 30 microns;
```

All the statistics are saved in the embedded system, and can be retrieved anytime from an external computer through the wireless interface. Additionally, we have developed a PC graphic user interface (GUI) for the image processing analysis, so the user can verify (visually) that the measurements calculated by the equipment are correct. The GUI also allows the utilization of the handpiece from a PC.

Fig. 3 shows an example of the analysis ran via the graphical interface. On the left is the original image obtained from the microscope. On the right, there is a graphical representation of the work made by the image processing algorithm. The red segments indicate the locations of the calculated measurements. This picture is useful to make sure that the measurements were taken in the correct locations (for example, red segments outside from fibres would be incorrect).

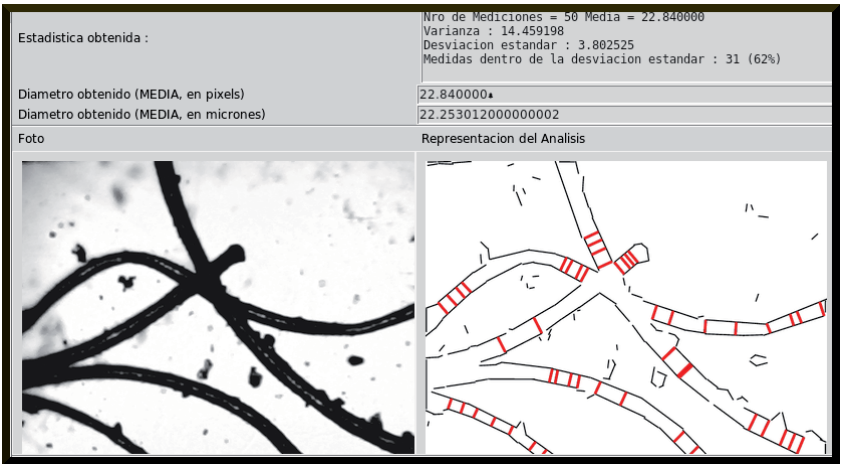


Fig. 3. Graphic User Interface for the image processing analysis.

3. Results

In order to validate *prototipo UNCOMA*, measurements of time and precision were done.

Total execution time lets us assess if the hardware and software are suitable for day to day usage. If the device takes too long to show the fibre diameter results, then the users might get confused, wondering if the device is or not functioning. Moreover, slow response may prove useless to farmers who must analyze many animals.

The most important requirement is about the precision of results. A device yielding incorrect or low-precision measurements will not be acceptable. Usually, the criancero has a trained eye for estimating measurements via visual analysis, so the instrument must produce better results than expert people.

3.1 Run-time order

Run-time cost has been observed taking 96 samples of true fiber strands. The time needed by the operator for sampling preparation has not been taken into account. The GNU **time** tool was used for this purpose. Each stage was run under GNU time and the corresponding utilization of resources was recorded. Results show that both image capture time and conversion time are constant. On the other hand, the execution time for image analysis varies.
The LSD algorithm, which is a linear time algorithm, is the most time consuming part of this image processing [7]. It has been observed that all images taken have less than 15 fibres each. Moreover, the images are all the same resolution. For our test set, the execution time has been determined to be in the range (0s, 3s] for all the 96 samples.

3.2 Precision analysis

Two measurement scenarios were analyzed in order to obtain precision results:

- Artificial fiber images, with known measures.
- A set of 96 true fiber strand images, from a same wool tuft. This test set was used to compare performance of "prototipo UNCOMA" and a commercial product, WoolView 20/20. Samples were collected independently on each equipment.

Artificial fiber images measurement. For this purpose, 74 synthetic fiber images were created using GNU image manipulation program (GIMP). They contain within 1 to 7 black lines (artificial fiber), with random orientation and direction. The thickness is also known beforehand, and documented all along the fiber.
Fig. 4 shows an example: on the left side, an artificial image with 5 fibres, each of which has a certain thickness defined along it. Diameters selected for the lines are within 10 and 40 pixels. The mean diameter value observed for each image was computed as the sum of measurement values on each line over the number of lines. Fig. 4 shows, on the right side, 35 of 74 results obtained by "prototipo UNCOMA" using the artificial images. Along each mean value bar, calculated by software, the manual calculation is shown.

True fiber images measurement. In order to analyze the proposed software, a comparison between results obtained by "prototype UNCOMA" and WoolView 20/20 was made. WoolView 20/20 is a commercial product for wool fiber measurement, with accuracy less than one micron, ready to be used directly on the animal.
Using "prototipo UNCOMA", 96 samples were obtained from a same animal tuft. This samples were saved to be published with this article. A batch script was developed to sequentially analyze each of these samples, using the prototype software. The statistics of each digital image was saved in a

separated file. Both original samples and results of diameter obtained using the prototype software, can be downloaded from [8].
A different operator took 96 samples using WoolView 20/20 equipment, and same animal tuft as the one used with "prototipo UNCOMA". This equipment shows a cumulative statistic; thus, for the purposes of comparison, intermediate details every 10 measurements were recorded.
Fig. 5 shows intermediate measurements results for both equipments. Final mean diameter obtained by the prototype proposed is 20,6263 microns (Fig. 6 shows normalized histogram for all measurements). Mean diameter obtained with WoolView 20/20 equipment is 20,9 microns. Table 1 shows final statistics for both.

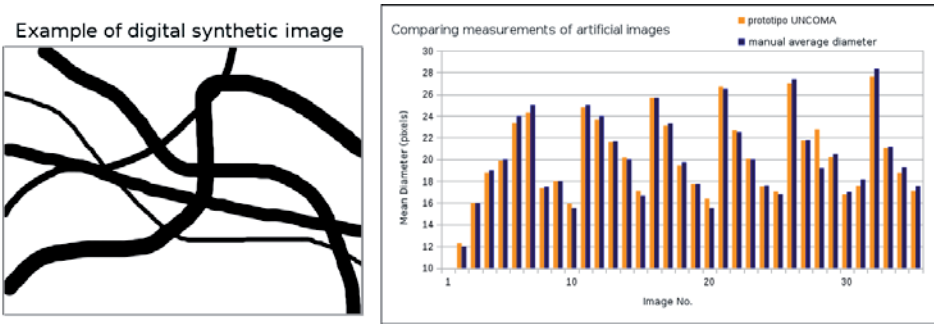


Fig. 4. Left: Artificial image. Right: Manual and by software measurements.

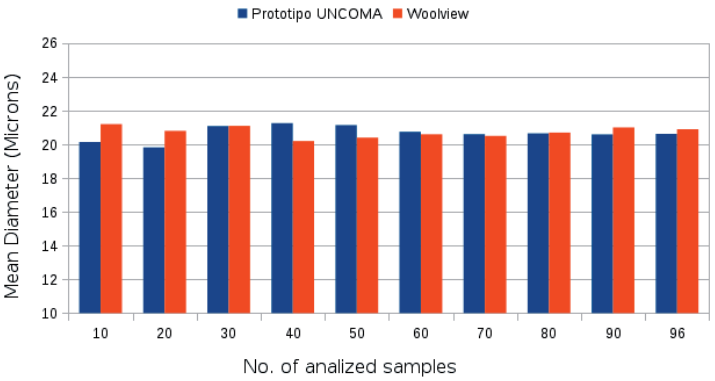


Fig. 5: Cumulative comparative from 96 measurements using two different equipments: "prototipo UNCOMA" and WoolView 20/20. Difference between the mean diameter obtained by both equipments decrease while increasing sample number using same animal tuft. After 20 measurements, difference falls down to less than a micron. After 50 measurements the difference is less than 0,5 microns.

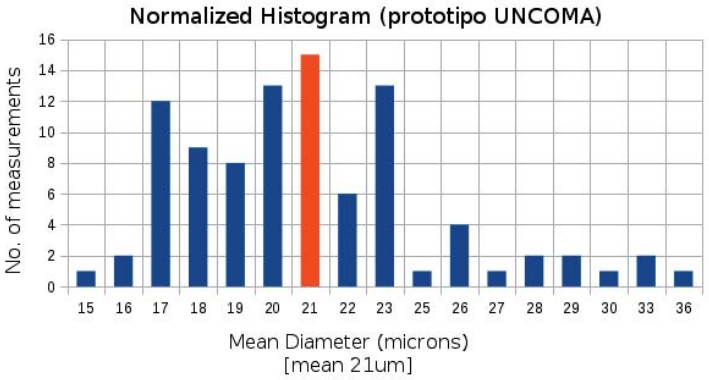


Fig. 6: Histogram for 96 measurements using "prototipo UNCOMA"

Table 1. Statistics of both equipments after 96 measurements.

	prototipo UNCOMA	WoolView 20/20
No. of analyzed images	96	96
Mean Diameter	20.6263	20.9
Comfort Factor	91.6%	91%
Standard Deviation	4.03um	5.5um
Coefficient of Variation	19.5%	27%
Samples under SD	82%	No info available

4. Future work and conclusions

This article presents the architecture of a low-cost real prototype system for the measurement and determination of goat fibre thickness. The equipment precision was validated by using a series of real and synthetic digital image files. The analysis of real samples was verified by comparisons with measurements taken with Woolview 20/20. The results from prototipo UNCOMA are less than one micron of difference in contrast to Woolview, when at least ten measurements were analyzed with both instruments. The weight and size of the handheld device are suitable for using it right on the animal, without removing the fibres from the goat fur. Preparation of the samples may take minutes, so the time taken by the analysis with prototipo UNCOMA (which amounts to a few seconds) is not significant.

There is still more to be done. The equipment might prove more accurate if validated against OFDA 2000, which is a professional instrument available only in specialized laboratories. On the other hand, we plan to test new techniques for image filtering and analysis, which our project is currently developing in software. As the current software discards samples taken with improper lens focus, the particular goal here is reducing the number of discarded samples.

References

1. Bendini, M.: Chapter 8. Transhumant Communities and Agroecosystems in Patagonia. Interactions Between Agroecosystems and Rural Communities. Cornelia Flora (2001). ISBN: 978-0-8493-0917-5
2. Bendini, M., Tsakoumagkos, P., Nogues, C.: Los crianceros trashumantes en Neuquén. Grupo de Estudios Sociales Agrarios (GESA), Universidad Nacional del Comahue (2005)
   http://investigadores.uncoma.edu.ar/cehepyc/publicaciones/Los_trashumantes_en_Neuquen.pdf
3. OFDA 4000 instrument, http://www.ofda.com/Natural_fibres/Ofda4000.html
4. Sirolan Laserscan instrument,
   http://www.itecinnovation.com/productDetails.php?id=52
5. Baxter, P.: Comparisons between OFDA, Airflow and Laserscan on raw merino wool. Technology And Standars Committee. International Wool Textile Organization. 2002.
6. WoolView 20/20 portable device,
   https://web.archive.org/web/20041102185907/http://www.woolview.com/Product1_2.htm
7. Grompone, R., Jakubowicz, J., Morel, J., Randall, G.: LSD: a Line Segment Detector, Image Processing On Line, 2 (2012), pp. 35–55,
   http://dx.doi.org/10.5201/ipol.2012.gjmr-lsd
8. Zurita, R., Lechner, M., Dataset for the validation and use of prototipo UNCOMA, http://se.fi.uncoma.edu.ar/prototipoUNCOMA/

V

Computer Security Workshop

Improving a Compact Cipher Based on Non Commutative Rings of Quaternions

JORGE ALEJANDRO KAMLOFSKY

[1] CAETI - Universidad Abierta Interamericana
Av. Montes de Oca 725 – Buenos Aires – Argentina
Jorge.Kamlofsky@uai.edu.ar

Abstract. Asymmetric cryptography is required to start encrypted communications. Most protocols are based on modular operations over integer's rings. Many are vulnerable to sub-exponential attacks or by using a quantum computer. Cryptography based on non-commutative algebra is a growing trend arising as a solid choice that strengthens these protocols. In particular, Hecht (2009) has presented a key exchange model based on the Diffie-Hellman protocol using matrices of order four with elements in Z_{256}, that provides 128-bits keys also to devices with low computing power. Quaternions are four-component's vectors. These also form non-commutative rings structures, with compact notation and lower run-times in many comparable operations. Kamlofsky et al (2015) presented a model using quaternions with elements in Z_{256}. To provide a 128-bit key is required 4 rounds of 32-bits. However, a gain of 42% was obtained. This paper presents an improvement of this cipher that reduces even more the run-times.

Keywords: Asymmetric cryptography, quaternion's cipher, non-commutative cryptography, post-quantum cryptography.

1. Introducción

1.1 Trabajos Relacionados

La Criptografía es una rama de la Matemática. Trata el problema de enviar información confidencial por un medio inseguro. Para ello, se cifra la información de manera que, aun cuando se encuentre disponible para cualquiera, no pueda ser utilizada, a menos que alguien autorizado la descifre. En una comunicación cifrada, entonces, pueden presentarse dos instancias diferentes: el intercambio seguro de claves y luego, con ello, el cifrado y descifrado del mensaje [1]. La Criptografía se divide en dos grandes ramas: de clave privada o simétrica, que cifra y descifra los mensajes y de clave pública o asimétrica, que logra el intercambio seguro de claves.
Diffie y Hellman fueron los pioneros de la criptografía asimétrica: en 1976, en [2] presentaron el revolucionario concepto de criptografía de clave pública cuya seguridad radica en el problema de la intratabilidad del logaritmo

discreto [3] (*DLP: Discrete Logarithm Problem*). Sin embargo, por facilidad de implementación práctica, el esquema criptográfico de clave pública hoy más usado es RSA [4]: su seguridad radica en el problema de la intratabilidad de la factorización de grandes números enteros *(IFP: Integer Factorization Problem)*.
En 1993 Peter Shor presentó en [5] un algoritmo que reduce la complejidad computacional del problema IFP mediante una computadora cuántica. A pesar que este dispositivo aún no se había inventado, solo la existencia del algoritmo, debilitó a esta rama de la criptografía. Hoy su existencia es un hecho: la empresa D-Wave Systems ya vendió computadoras cuánticas a Lockheed Martin, al laboratorio Los Alamos, a Google y a la NASA, entre otros [6]. Además, IBM por su lado, ofrece servicios en la nube con su computadora cuántica [7].
Desde inicios de este siglo ha crecido el interés por el desarrollo de criptosistemas asimétricos alternativos que sean resistentes a ataques de complejidad sub-exponencial y ataques vía computadora cuántica [8 – 9]. A la mayoría de estos esquemas se los denomina colectivamente como criptografía post-cuántica [10] o bien, por su naturaleza algebraica, se los denomina criptografía no conmutativa [11]. Sobre esta línea, no se conocen ataques que hayan logrado resultados concretos. Dentro de esta línea, en [12] se presentó un esquema de distribución de claves Diffie-Hellman basado en un anillo de polinomios matriciales. Al sistema se lo denominó compacto debido a que no se requieren librerías de precisión extendida, lo que hace posible su uso en procesadores de menor porte. En [13] se implementó dicho esquema en un anillo de polinomios de cuaterniones, lo cual permitió la obtención de claves de la misma longitud (128 bits) con una mejora de 42,59% en los tiempos de ejecución.
En este trabajo, se presentan resultados que muestran mejoras aún mucho mayores en los tiempos de ejecución usando cuaterniones con elementos de otros conjuntos numéricos, obteniendo claves de la misma longitud que las presentadas en [12, 13].

1.2 Motivación y Alcance

En [13] se logra el intercambio de claves de 128 bits a partir de cuaterniones con elementos de 8 bits en 42,59% menos tiempo de ejecución. Para ello, en cada instancia se obtiene un cuaternión que trae consigo 4 x 8 bits = 32 bits. El algoritmo se repite 4 veces para obtener los 128 bits. Sin embargo, en otro trabajo [14], donde se compara el desempeño de aplicaciones que utilizan matrices cuadradas y cuaterniones, se observa que la ganancia en tiempos de ejecución obtenida con cuaterniones es notoriamente superior, lo cual permitió intuir la existencia de un amplio margen para mejorar aún más los tiempos de ejecución de este cifrador.
El destino del protocolo presentado en [12] es para procesadores de pequeño porte: 16 bits. Sin embargo, se trabaja con elementos de 8 bits. De aquí se puede obtener parte de ese margen disponible solo trabajando con elementos de 16 bits.

Como los primeros ensayos obtenidos fueron favorables, se permite pensar en ampliar la idea a elementos de 32 bits, considerando que hoy también pueden llamarse a los procesadores de 32 bits como de menor porte.

1.3 Objetivo del Trabajo

La finalidad de este trabajo es mostrar que mediante el uso de cuaterniones en el conjunto numérico adecuado pueden obtenerse importantes mejoras en los tiempos de ejecución en el esquema de intercambio de claves Diffie Hellman Compacto.

1.4 Relevancia del tema

La existencia de la computadora cuántica es un hecho. Con ello, la criptografía asimétrica clásica, se encuentra muy debilitada, ya que RSA, el algoritmo más usado ha sido quebrado. Los nuevos desarrollos en criptografía post-cuántica permiten mitigar esta debilidad. Y esquemas más veloces permiten que en la práctica, éstos puedan ser implementados de manera extendida.

1.5 Estructura del Trabajo

En la Sección 2 se presenta el marco teórico. En la Sección 3 se presenta la solución propuesta con datos experimentales. En la Sección 4 se presentan las conclusiones.

2. Marco Teórico

2.1 La Importancia de la Criptografía en la Seguridad de las Comunicaciones

Nociones Básicas de Criptografía Simétrica. La Criptografía se ocupa de asegurar la integridad y confidencialidad en las comunicaciones a través de un canal inseguro. Para ello, el mensaje se transforma en el punto de emisión mediante operaciones matemáticas de manera que sea imposible de interpretar mientras viaja en el canal inseguro, o bien su costo en tiempo y/o recursos sean tan altos que su descubrimiento carezca de sentido.
Se usan algoritmos criptográficos altamente robustos que permiten además que la información se encripte bit a bit (cifradores de flujo) o en grupos de n-bits (cifradores de bloque) permitiendo que puedan cifrarse comunicaciones en tiempo real [3]. Estos cifradores usan la misma clave para el cifrado y descifrado del mensaje. A estos cifradores se los clasifica como Criptografía

Simétrica. Muchos cripto-sistemas simétricos seguros (AES, DES, Trivium) pueden iniciarse con claves de 128 bits.

Nociones Básicas de Criptografía Asimétrica. La criptografía asimétrica o de clave pública, usa elementos públicos que se comparten, y elementos privados que se mantienen en secreto. Generalmente usan propiedades y operaciones de aritmética modular en estructuras algebraicas de anillos de números enteros.
Ésta brindó soluciones al problema de presentar en forma segura claves para su uso en cifradores simétricos: mientras RSA [4] permite que se pueda enviar la clave simétrica cifrada a otro usuario usando su clave pública, con Diffie-Hellman [2] y Elgamal [15] ambas partes pueden generar la misma clave intercambiando elementos.

Amenaza a la Criptografía: El Algoritmo de Shor y la Computación Cuántica. En 1995 Peter Shor presentó un algoritmo para computación cuántica basado en la transformada rápida de Fourier (FFT) que logra resolver en tiempo polinómico el problema IFP [5]. Es decir, permite reducir drásticamente la complejidad del problema (considerado de clase NP) a niveles atacables [16].
Una computadora cuántica usa qubits en lugar de bit. Un qubit posee los estados 0, 1 y la superposición de ambos: 0 y 1 a la vez. Por ello, se puede realizar una cantidad exponencial de operaciones en paralelo en relación con la cantidad de qubits del computador cuántico.
En 2001 se implementó el algoritmo de Shor en la primer computadora cuántica. La computación cuántica prácticamente arrasa con todo lo conocido en la criptología actual: con ello desaparecen de escena todos los criptosistemas de clave pública: RSA y todas las variantes de ElGamal y Diffie-Hellman [16].

Criptografía Post-Cuántica Basada en Anillos no Conmutativos. Se utilizan estructuras de anillos de matrices cuadradas o de cuaterniones, entre otros, con elementos finitos, por lo tanto, su seguridad radica en la complejidad del tratamiento del problema DLP. Algunos esquemas como el presentado en [17] se basan en la dificultad de resolver el problema SDP (*Simple Descomposition Problem*) en un anillo no conmutativo de polinomios matriciales.
Desde el punto de vista criptográfico, solo se necesita estar seguro que no existe fórmula que permita reducir la complejidad del problema DLP (incluso con computadora cuántica). Y esto está garantizado ya que en los anillos no conmutativos no existe forma de relacionar el determinante de una matriz o bien sus eigenvalores con la potencia de la matriz [18], parte de la clave privada, independientemente de la cantidad de qubits que pudiera tener una computadora cuántica que ejecute el ataque.
En [12] se muestran más consideraciones de la seguridad de estos esquemas.

2.2 El Anillo no Conmutativo de Cuaterniones

Anillos no Conmutativos. Un anillo (A ; + ; .) es una estructura algebraica (un conjunto A con las operaciones suma y producto) donde (A ; +) forman estructura de grupo, y (A ; .) de semigrupo. Un anillo será no conmutativo si no se verifica la propiedad conmutativa entre todos los elementos de A para la operación producto.
El primer anillo de división (cuyos elementos no nulos son inversibles) no conmutativo fue el anillo de los cuaterniones. Otro ejemplo es el conjunto de matrices cuadradas de orden *n* con coeficientes en *A* (simbolizado por $M_n(A)$), es un anillo con respecto a las operaciones usuales de suma y producto de matrices. Si $n > 1$, entonces $M_n(A)$ no es conmutativo. Otros ejemplos de anillos de división no conmutativos son: los Octoniones, Sedeniones, Tessarines, cocuaterniones o bicuaterniones.

Definición: Cuaternión. Sea (A ; + ; .) un anillo conmutativo con unidad. Un cuaternión con coeficientes en *A* es un número hiper-complejo *q* de la forma: $q = a + b.i + c.j + d.k$ donde $a, b, c, d \in A$; i, j, k son unidades imaginarias que verifican que: $i^2 = j^2 = k^2 = -i \,.\, j \,.\, k = -1$ y además: $i \,.\, j = -j \,.\, i = k$; $j \,.\, k = -k \,.\, j = i$; $i \,.\, k = -k \,.\, i = j$.
Fueron creados en 1843 por William Hamilton [19]. Forman una estructura de Anillo de división no conmutativo. Tienen una notación compacta y resultan muy sencillos para trabajar. Son muy eficientes: requieren menor cantidad de operaciones básicas y menor espacio de almacenaje en comparación con la operación de matrices.

Operaciones Básicas con Cuaterniones. Sean los cuaterniones $q = (w , x , y , z)$, $q_1 = (w_1 , x_1 , y_1 , z_1)$ y $q_2 = (w_2 , x_2 , y_2 , z_2)$

Norma del Cuaternión q: $|q| = \sqrt{q.\bar{q}} = \sqrt{\bar{q}.q} = \sqrt{w^2 + x^2 + y^2 + z^2}$

Suma, Resta y Producto de un Escalar por un Cuaternión: Se realiza de la misma forma que con cualquier vector de 4 dimensiones.

Producto: $q_1.q_2 = (w_1.w_2 - x_1.x_2 - y_1.y_2 - z_1.z_2,\ w_1.x_2 + x_1.w_2 + y_1.z_2 - z_1.y_2,\ w_1.y_2 - x_1.z_2 + y_1.w_2 + z_1.x_2,\ w_1.z_2 + x_1.y_2 - y_1.x_2 + z_1.w_2)$. Notar que el producto entre cuaterniones no es conmutativo.

Cociente: $\frac{q_1}{q_2} = q_1.(q_2)^{-1} = q_1.\left(\frac{q_2}{|q|^2}\right)$ con $q_2 \neq (0,0,0,0)$.

Potencia: $q_1^n = |q_1|^n.\left(cos\left(n\frac{\alpha}{2}\right) + \check{v}.sen\left(n\frac{\alpha}{2}\right)\right)$.

2.3 Sistemas Compactos de Intercambio de Claves Diffie-Hellman Basados en Álgebra No Conmutativa

Esquema Diffie - Hellman Compacto con Matrices (DHCM). En [12] se presentó un sistema de intercambio de claves Diffie Hellman [2] sobre anillos de matrices de enteros con elementos en Z_{256}, que utiliza como clave privada un polinomio con coeficientes y exponentes en Z_{16}. Un par de ventajas surgen de ello. Primero: no se requiere el uso de librerías de precisión extendida, por lo tanto, puede ser usado en procesadores de pequeño porte. De allí la calificación de compacto. La otra ventaja se relaciona con el hecho que las matrices conforman estructura de anillo no conmutativo. Gracias a ello, el esquema es inmune a ataques cuánticos y de complejidad sub-exponencial. La clave resultante es una matriz de orden 4 con elementos en Z_{256} conformando así una clave de 128 bits, adecuada para su uso en varios cifradores simétricos seguros.

Resumen del Protocolo. Alice envía a Bob (a través de un canal público e inseguro) dos números enteros aleatorios m y n en Z_{16}, y dos elementos aleatorios A y B, matrices de orden 4 con elementos en Z_{256}. Elige como clave privada un polinomio entero $f(x)$ con coeficientes y exponentes en Z_{16} tal que $f(A) \neq 0$. Bob elige como su clave privada un polinomio entero $h(x)$ con coeficientes y exponentes en Z_{16} tal que $h(A) \neq 0$. Luego Alice y Bob calculan sus tokens: $r_A = f(A)^m \,.\, B \,.\, f(A)^n$ (Alice) y $r_B = h(A)^m \,.\, B \,.\, h(A)^n$ (Bob); y se los intercambian para el cálculo de las claves: $K_A = f(A)^m \,.\, r_B \,.\, f(A)^n$ (Alice), $K_B = h(A)^m \,.\, r_A \,.\, h(A)^n$ (Bob) con $K_A = K_B$.

Esquema Diffie-Hellman Compacto con Cuaterniones (DHCQ8). Diversas aplicaciones pueden implementarse con cuaterniones en lugar de matrices cuadradas, en menores tiempos de ejecución [14]. Ello inspiró el desarrollo de este esquema [13], que resultó en un ahorro de tiempo cercano al 50% bajo condiciones similares.

Consideraciones del Protocolo. Este protocolo es muy similar al de matrices [12]: En su lugar, los elementos A y B son cuaterniones en forma cartesiana con elementos en Z_{256}. Cada cuaternión, entonces, tiene 32 bits. Se incorporan dos instancias de normalización del cuaternión, y una modularización en Z_{256}. Gracias a esto se puede aprovechar la notable simpleza de potenciar cuaterniones normalizados, en comparación con la complejidad de potenciar matrices.

3. La Mejora Propuesta

3.1 Resumen de la Mejora Propuesta

El protocolo presentado en [13] trabaja con elementos de Z_{256}. En esta propuesta se trabaja con elementos en $Z_{2^{k*16}}$ para $k = 1$ (DHCQ16) y $k = 2$ (DHCQ32).

3.2 El Protocolo Propuesto

Alice elige dos números enteros aleatorios m y n en Z_{16}, y dos cuaterniones aleatorios A y B, con elementos de $Z_{2^{k*16}}$ (con $k = 1$ o bien $k = 2$) y calcula sus normalizaciones: q_A y q_B. Luego elige como clave privada un polinomio entero $f(x)$ con coeficientes y exponentes en Z_{16} tal que $f(q_A) \neq 0$ y envía a Bob por el canal inseguro los elementos m, n, q_A y q_B. Bob elige como clave privada un polinomio entero $h(x)$ con coeficientes y exponentes en Z_{16} tal que $h(q_A) \neq 0$. Alice y Bob realizan las normalizaciones de $f(q_A)$ y $h(q_A)$: $f'(q_A)$ y $h'(q_A)$. Alice calcula su token: $r_A = f'(q_A)^m \,.\, B \,.\, f'(q_A)^n$. Bob calcula el suyo: $r_B = h'(q_A)^m \,.\, B \,.\, h'(q_A)^n$; y se los intercambian para el cálculo de las claves: $k_A = f'(q_A)^m \,.\, r_B \,.\, f'(q_A)^n$ (Alice), $k_B = h'(q_A)^m \,.\, r_A \,.\, h'(q_A)^n$ (Bob), las cuales se modularizan: $K_A = k_A \,.\, 2^{k*16} \pmod{2^{k*16}}$, $K_B = k_B \,.\, 2^{k*16} \pmod{2^{k*16}}$ con $K_A = K_B$.

La clave obtenida para $k = 1$ posee 4 x 16 bits = 64 bits. Para lograr una clave de 128 bits, el proceso debe repetirse. Para $k = 2$, la clave posee 4 x 32 bits = 128 bits.

3.3 Un Ejemplo Numérico

El ejemplo numérico se realiza para $k = 1$ (DHCQ16), y a los fines de esta presentación, la cantidad de decimales se limita a 2.

Alice elige dos números enteros aleatorios $m=7$ y $n=10$, y dos cuaterniones aleatorios $A = (45606, 11140, 21549, 43028)$ y $B = (42679, 56493, 45062, 43484)$, con elementos de Z_{65536} y calcula sus normalizaciones: $q_A = (0.68, 0.17, 0.32, 0.64)$ y $q_B = (0.45, 0.60, 0.48, 0.46)$. Luego elige como clave privada un polinomio entero $f(x) = 14x^{15} + 12x^{14} + 13x^{13} + 8x^{12} + 5x^{11} + 4x^{10} + 9x^9 + 13x^8 + 7x^7 + + 11x^6 + 4x^5 + 9x^4 + 14x^3 + 7x^2 + 3x + 4$ y envía a Bob por el canal inseguro los elementos m, n, q_A y q_B.

Bob elige como su clave privada un polinomio entero $h(x)= 5x^{15} + 6x^{14} + 3x^{13} + 11x^{12} + 14x^{11} + 3x^{10} + 12x^{8} + 2x^{7} + 3x^{6} + x^{5} + 14x^{4} + 12x^{2} + + 9x$. Luego Alice y Bob realizan las normalizaciones de $f(q)$ y $h(q)$: $f'(q) = (0.78, -0.14, -0.27, -0.54)$ y $h'(q) = (-0.88, 0.11, 0.21, 0.41)$. Alice calcula su token: $r_A = (-0.37, -0.12, 0.36, 0.26)$, Bob calcula el suyo: $r_B = (-0.17, 0.26, 0.00, 0.39)$; y se los intercambian: $k_A = (-0.25, 0.06, 0.09, -0.11)$ (Alice), $k_B = (-0.25, 0.06, 0.09, -0.11)$ (Bob), las cuales se modularizan: $K_A = k_A \,.\, 65536 \pmod{65536} = (49061, 3662, 5778, 58342)$, $K_B = k_B \,.\, 65536 \pmod{65536} = (49061, 3662, 5778, 58342)$ con $K_A = K_B = K$: la clave.

3.4 Equipamiento Usado

El computador usado contiene un procesador AMD A10-5745M × 4 núcleos de 64 bits y 12Gb de memoria RAM. Se instaló una distribución Ubuntu 15.10, el cual tiene un núcleo Linux Debian. Los algoritmos fueron programados en Python 2.7.10.

3.5 Resultados Experimentales

Se presenta una comparación de los tiempos de ejecución del algoritmo DHCM [12], DHCQ8 [13] y la solución aquí propuesta: para 16 bits y 32 bits ($k = 1$ y $K = 2$) para la obtención de 10.000 claves de 128 bits con polinomios aleatorios con coeficientes y exponentes en Z_{16}. La tabla 1 muestra los resultados experimentales.

Tabla 1. Tiempos de ejecución para la obtención de 10.000 claves de 128 bits mediante las diferentes opciones analizadas.

	CPU Time (s)					CPU Time (s)			
Nº Test	DHCM 8-bits	DHCQ8 8-bits	DHCQ16 16-bits	DHCQ32 32-bits	Nº Test	DHCM 8-bits	DHCQ8 8-bits	DHCQ16 16-bits	DHCQ32 32-bits
1	37,8177	20,0350	11,3851	5,5569	14	37,8728	19,9016	11,6061	5,7219
2	36,4591	19,5770	11,1946	5,7522	15	36,7675	19,3920	11,1752	5,6757
3	37,6444	19,0540	11,3141	5,6158	16	37,7545	19,1195	11,6449	3,9712
4	36,6029	19,4511	11,3369	5,7398	17	37,0721	19,1790	11,3357	5,5914
5	37,3720	20,0530	11,2472	5,6139	18	36,4384	19,6326	11,4020	5,5797
6	37,2393	19,8001	11,0701	5,5895	19	37,0333	19,6599	11,4161	5,5830
7	37,0366	18,9289	11,2918	5,6067	20	37,5093	19,1302	11,1700	5,5174
8	37,3745	19,7008	11,1029	5,7065	21	36,5904	19,6875	11,3885	5,6230
9	37,3369	19,2519	11,2191	5,6430	22	36,8067	19,8909	11,6555	5,7127
10	36,4455	19,7068	11,6477	5,5709	23	37,9800	19,2684	11,4140	5,5620
11	37,6164	19,3665	11,4154	5,5396	24	37,2734	19,4936	11,1356	5,6469
12	37,6196	20,0801	11,1886	5,6568	25	36,9313	19,3517	11,2654	5,5139
13	37,3872	19,7883	11,4547	5,8716					

El tiempo promedio (en segundos) para la obtención de 10000 claves con DHCM fue 37,2, con DHCQ8 fue 19,54, con DHCQ16 fue 11,34 y con DHCQ32 fue 5,57. Como los coeficientes de variación (CV) son CV < 0,1 se acepta al promedio como indicador de tendencia central adecuado. Este experimento generó 1 millón de claves sin error.

3.6 Ventajas de la Solución

Velocidad del Cifrador. Menores tiempos de ejecución evidencian mayor velocidad del esquema de cifrado propuesto en comparación con los esquemas presentados en [12] y [13]: El esquema DHCQ8 logra la generación de 128 bits de clave en el 52% del tiempo en comparación con DHCM lo que implica 1,9 veces su velocidad; DHCQ16 lo hace en el 30% implicando más tres veces su velocidad, y DHCQ32 lo hace en el 15% lo que significa más de 6 veces su velocidad. Si la comparación se hace entre los esquemas propuestos y DHCQ8: DHCQ16 lo hace en el 58% del tiempo haciéndolo 1,72 veces más rápido. Sin embargo, DHCQ32 lo hace en el 29% del tiempo, es decir con 3,5 veces más rapidez.

Solución Apta para Procesadores de Pequeño Porte. La implementación de la solución aquí propuesta puede realizarse sin necesidad de uso de librerías de precisión extendida, lo cual lo hace apto para procesadores de pequeño porte.

Inmunidad Frente a Ataques de Complejidad Sub-Exponencial o de Computadora Cuántica. Los anillos de cuaterniones conforman estructuras algebraicas no conmutativas, y sobre este tipo de estructuras no se conocen aún ataques de estos tipos que hayan sido efectivos y que debiliten su seguridad.

4. Conclusiones

Ya se ha visto que usando cuaterniones se puede obtener aplicaciones más rápidas que con matrices. Y esto también es válido en criptografía. Puede aprovecharse la simpleza de la potencia cuaterniones normalizados para lograr mayor velocidad. Trabajar con los conjuntos numéricos adecuados permite hacer uso más eficiente de esta ventaja, logrando implementaciones criptográficas más veloces.
Al usarse estructuras de anillos no conmutativos, hace a este esquema inmune a ataques de complejidad sub-exponencial o de computadora cuántica.
El esquema propuesto es útil para procesadores de menor porte.

Referencias

1. Marrero Travieso, Yran: La Criptografía como elemento de la seguridad informática. ACIMED 11.6 (2003).
2. Diffie W., Hellman M.E: New directions in cryptography. IEEE Transactions on information theory, 22, 644-654, (1976).
3. Menezes, Alfred J., Paul C. Van Oorschot, and Scott A. Vanstone: Handbook of applied cryptography. CRC press (1996).
4. Rivest, Ronald L., Adi Shamir, and Len Adleman. A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM 21.2, 120-126. (1978)
5. Shor P.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput., 5, 1484-1509 (1997)
6. D-Wave-Systems Press Releases [en línea], (2016). Disponible en: <http://www.dwavesys.com/news/press-releases>. Fecha de consulta: 05/06/2016.
7. IBM: IBM Makes Quantum Computing Available on IBM Cloud to Acceletare Innovation [En Linea], (2016). Disponible en: <https://www-03.ibm.com/ press/ us/ en/ pressrelease/ 49661.wss>. Fecha de consulta: 05/06/2016.
8. Magliveras S.S., Stinson D.R., van Trung T.: New approaches to designing public key cryptosystems using one-way functions and trapdoors in finite groups, Technical Report CORR, 2000-2049 (2000)
9. Shpilrain V., Zapata G.: Combinatorial group theory and public-key cryptography, Preprint arXiv/math.gr, 0410068 (2004)
10. Barreto,P. et al: Introdução à criptografia pós-quântica, Minicursos do XIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais, SBSeg, (2013).
11. Gerritzen L. et al (Editors): Algebraic Methods in Cryptography, Contemporary Mathematics, AMS, Vol. 418, (2006)
12. Hecht J.: Un modelo compacto de criptografía asimétrica empleando anillos no conmutativos. V Congreso Iberoamericano de Seguridad Informática CIBSI, Montevideo (2009).
13. Kamlofsky J.A., Hecht J.P, Abdel Masih S., and Hidalgo Izzi, O.: A Diffie-Hellman Compact Model Over Non-Commutative Rings Using Quaternions. VIII Congreso Iberoamericano de Seguridad Informática CIBSI, Quito (2015).
14. Kamlofsky J., Bergamini L.: Cuaterniones en Visión Robótica. V Congreso de Matemática Aplicada, Computacional e Industrial MACI, Tandil (2015).
15. Elgamal, Taher. A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms. En Advances in cryptology. Springer Berlin Heidelberg, pp. 10–18 (1984).
16. Hecht, JP.: Fundamentos de Computación Cuántica. Editorial Académica Española. ISBN 978-3-8484-7529-2 (2005).
17. Cao Z., Xiaolei D., Wang L.: New public-key cryptosystems using polynomials over non-commutative rings, Preprint arXiv/cr, eprint.iacr.org/2007/009.pdf (2007).
18. Eftekhari, M.: A Diffie–Hellman key exchange protocol using matrices over noncommutative rings. Groups-Complexity-Cryptology, 4(1), pp. 167–176 (2012).
19. Hamilton, W. R.: Lectures on Quaternions: Containing a Systematic Statement of a New Mathematical Method, Hodges and Smith, (1853)

Loss of Votes in NIDC
Applying Storage in Parallel Channels

Pablo García[1], Germán Montejano[12], Silvia Bast[1], Estela Fritz[1]

1. FCEyN - Universidad Nacional de La Pampa - Argentina
pablogarcia@exactas.unlpam.edu.ar,
WWW home page: http://www.exactas.unlpam.edu.ar
2. FCFMyN - Universidad Nacional de San Luis - Argentina
gmonte@unsl.edu.ar,
WWW homepage: http://www.unsl.edu.ar

Abstract. Birthday Paradox states that in a group of 23 people, the probability that there are at least two who share the same birthday is very close to ½. This assertion is unacceptable for any scheme that proposes a vote storage method based on a vector of slots whose position is chosen at random. In this situation it may produce collisions.
A collision occurs when two or more votes are stored in the same slot. It produces the loss of the coincident votes. This is the original model of the Non - Interactive Dining Cryptographers (NIDC) protocol.
The actual paper shows new achieved results obtained by analyzing the behavior of a storage technique based on parallel channels. This scheme consists of replicating each vote in Q parallel channels, keeping the total number of slots (T) without variation.

Keywords: Parallel Channels, Storage Birthday Paradox, Non - Interactive Dining Cryptographers, Collisions.

1. Introduction

Within the scope of a research line that began at 2013 and which was formally presented in [1], the exact security level requested for anonymity in an electronic voting scheme was analyzed. Many of the proposed schemes (Mix Net based) give unconditional security to the votes' information and computational assurance to voter's privacy. However, it is easy to see that it is an erroneous proposal. In [2] it is concluded that it is necessary to give unconditional security for the privacy, because it must be protected indefinitely. Otherwise, votes must be kept for a finite period of time.

Consequently, those schemes, that include unconditional security as the main feature, acquire maximum interest. In this sense, one of the most interesting is Dinning Cryptographers (DC), which is described in detail in [3]. This protocol is resourceful and gives unconditional privacy.
The analysis is focused on a derivative of DC, called Non Interactive Dining Cryptographers (NIDC [4]), that relaxes the condition of concurrency online for all participants. This protocol is suitable to be applied to electronic voting scheme.
The original version of NIDC stores data in a vector of slots. This is observed in figure 1.

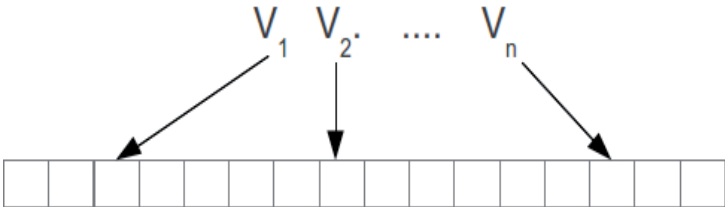


Fig. 1. Original NIDC Storage

If two or more votes are stored in the same slot, a collision occurs. That results in the loss of coincident votes. Simultaneously, the proposition of true randomness for the choice of position indicates that collisions may happen. It then seeks to ensure that the proportion of lost data is kept below a desired value with a certain probability. The proposed model in Figure 1 may be explained by Birthday Paradox ([5]). In those conditions, it is required a very significant number of slots to obtain suitable security levels.
Two interesting alternatives, aimed at improving the Birthday Paradox effect are presented in [6]. In that document an optimization for NIDC, applying multiple networks serially and in parallel, is proposed. In this case, however, it seeks to generalize the approach to storage in parallel channels, a matter that may be generalized to multiple real-world problems, including NIDC. The alternative proposal consists on implementing N parallel channels, replicating each vote in all channels, in potentially different random positions in each case, as outlined in Figure 2.

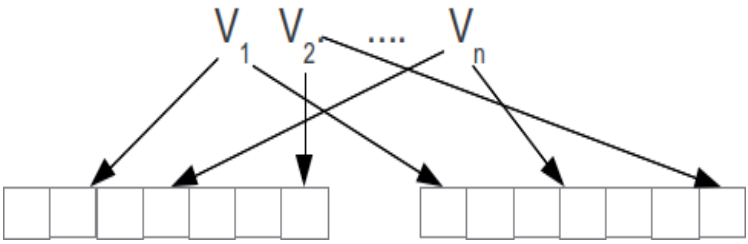


Fig. 2. Scheme based on Parallel Channels

It begins by describing the parameters involved:

T**:** # Total slots to implement. $T \in Z^+$.
S**:** # Parallel slots on each channel. $S \in Z^+ \wedge S\ d\ T$.
N**:** # Voters. $N \in Z^+$.
Q**:** # Parallel channels to implement. $Q \in Z^+$.
Q_{to}**:** # Parallel channels to implement (Theoretically Optimal). $Q_{to} \in R^+$.
Q_{po}**:** # Parallel channels to implement (Practically Optimal). $Q_{po} \in Z^+$.
R**:** # Replicas of a vote on the same channel. $R \in Z^+$.
*PLV***:** Percentage of Lost Votes.

Throughout previous papers ([7], [8] and [9]) the following relevant findings have been set forth (in addition, function *CEIL* will be used; it computes the nearest higher integer. That is necessary because Q_{to} could be non integer):

– For a fixed number of voters N, the recommended number of slots for each parallel channel (S) is given by the formula:

$$S = CEIL\left(\frac{N}{ln2}\right) \tag{1}$$

– For given values of T and N, there exists an optimal number of parallel channels. Such value is expressed:

$$Q_{to} = ln2\frac{T}{N} \tag{2}$$

That formula should be taken to the next integer.

$$Q_{po} = CEIL\ (Q_{to}) \tag{3}$$

– An appropriate lower bound for the probability of X ="*no vote is lost*" is obtained by applying equation:

$$Pr(X) > 1 - (\frac{1}{S}(N-1)^{Q})^{N} \quad (4)$$

Besides, concrete methods were published to obtain optimal values for all parameters using a spreadsheet ([10]) and the pseudo-code algorithm that must be applied for the same purpose was shown in [11].
In this document the variable *PLV* is analyzed. One equation is described to get an approximation of the expected value of that variable. The following section describes the deduction of such formula.

2. Expected Value of Percentage of Lost Votes (*PLV*)

By applying a parallel channels scheme, a question that quickly arises is: for a situation with N voters, and Q parallel channels of S slots (such that $T = SQ$), which is the expected Percentage of Lost Votes (*PLV)*.
At the beginning, it is considered the original Birthday Paradox proposal. The first thing we see is that even in the best case (the 23 people Birthdays on different dates), the number of slots that will not be used is 342, then we have approximately 6,3 % of occupied slots and 93,7 % of empty slots. Consequently, for each slot containing a vote, more than 15 receive no ballots.
The proposal is to divide all the slots in $Q>1$ parallel channels and to deposit an occurrence of each vote in each of the channels. In addition to what appeared in the simulations, the idea is related to the fact that a vote is lost on a given channel is independent of what happens in the other $(Q - 1)$ channels.
Independent events verify that:

$$Pr(A) \; B) = Pr(A)\, Pr(B) \quad (5)$$

Clearly a vote will be lost only if it collides on all the channels. The number of local collisions increases, since each channel will have a measure smaller than the single vector. However, an optimization based on replicas is obtained.
Let μ be:

$$\varepsilon = \frac{N}{S} \quad (6)$$

Initially the situation is analyzed if a single vector is implemented, therefore, $S = T$.
Several strategies based on analyzing the probability distribution are presented in [12]. Also, the approaches proposed by Feller [13], were mentioned.
These approaches improve their behavior when N ’ • y T ’ • . A tool

that may be useful is Stirling's approximation for calculating factorials:

$$N! = \sqrt{2\Sigma N}\left(\frac{X}{N}\right)^{N} \tag{7}$$

It is proposed another approach, which is simpler than the previous one because it only calculates expected values rather than probability distribution.
Considering the first vote, the probability that it falls into the slot 1 is:

$$p = \frac{1}{s} \tag{8}$$

Consequently, the probability that it does not fall into the slot 1 is:

$$q = 1 - p = \left(1 - \frac{1}{S}\right) \tag{9}$$

Generalizing to N votes, we get a binomial distribution with parameters N and p.
Let X_k be: *"Exactly k votes are stored in slot 1"* with $k \in Z^{+}$

$$\Pr(X_k\ ) = \binom{N}{k} p^k q^{N-k} \tag{10}$$

$$\Pr(X_k) = \binom{N}{k}\left(\frac{1}{S}\right)^{k}\left(1 - \frac{1}{S}\right)^{N-k} \tag{11}$$

Given that:

$$\lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^{n} = e \tag{12}$$

We can assure:

$$Pr(X_0) = \left(1 - \frac{1}{S}\right)^{N} \approx\ e^{-\varepsilon}$$

$$Pr(X_1) = N\frac{1}{S}\left(1 - \frac{1}{S}\right)^{N-1} \approx\ \varepsilon e^{-\varepsilon}$$

$$Pr(X_2) = \frac{N(N-1)}{2}\left(\frac{1}{S}\right)^{2}\left(1 - \frac{1}{S}\right)^{N-2} \approx\ \frac{1}{2}\varepsilon^2 e^{-\varepsilon}$$

These probabilities also represent the expected number of votes in slot 1. It is obvious that he same reasoning can be applied to any slot. Therefore, it is possible to find the expected frequency.

Given that $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$, for $N = S = 1000$, $\mu = 1$. Therefore:

$$Pr(X_0) = Pr(X_1) = e^{-1}\ H0.3678 \tag{13}$$

Similarly, for $N = 500$, $S=1000$, $\mu = ½$, in which case:

$$\Pr(X0) = \mathrm{e}^{-\frac{1}{2}} \approx 0{,}6065$$

$$\Pr(X0) = \frac{1}{2}\,\mathrm{e}^{-\frac{1}{2}} \approx 0{,}3032$$

Let *E(k)* be: #expected slots containing *k* votes. Its value is obtained as follows:

$$E(k) = Sp(X_k) \tag{14}$$

This fits together with the Poisson approximation stated in [13]:

$$\mathrm{E}[(\mathrm{Poisson}(\text{»}))] = \text{»} \tag{15}$$

$$\lambda[X0] = Se^{-\frac{N}{S}} \tag{16}$$

For $k = 1$:

$$\lambda|X_1| = \frac{\left(ne^{-\frac{r}{n}}\right)}{k!}\left(\frac{r}{n}\right)^k = \frac{Se^{(-\varepsilon)}}{k!}\varepsilon = S\varepsilon e^{-\varepsilon} \tag{17}$$

For $k = 2$

$$\lambda|X_2| = \frac{\left(ne^{-\frac{r}{n}}\right)}{k!}\left(\frac{r}{n}\right)^k = \frac{Se^{(-\varepsilon)}}{2}\varepsilon^2 = S\varepsilon^2 e^{-\varepsilon} \tag{18}$$

The Poisson approximation improves its quality when $S \to \infty$ and N ’ • . The previous formula is related to *S*. It is more interesting yet, to obtain a connection with the number of successful votes, ie for $k = 1$, $E(1)$ is divided into N and it is obtained:

$$\frac{S\varepsilon e^{-\varepsilon}}{N} = e^{-\varepsilon} \tag{19}$$

Consequently, for $Q = 1$

$$s[1] = 1000e^{-1}\ H368 \quad (20)$$

$$P\,r(successfulvote)\ H0.36 \quad (21)$$

$$Pr(lostvote)\ H1 - 0.36 = 0.64 \quad (22)$$

For $Q = 2$, one vote is lost if collides in the two channels:

$$Pr(successfulvote) = 1 - 0.39\ H0.61 \quad (23)$$

$$P\,r(lostvote) = 0.64^2\ H0.39 \quad (24)$$

The same scheme is generalized $\forall Q > 2$. Thus it is obtained a formula to calculate the expected value of the variable *Percentage of Lost Votes*.

$$|PLV| = \left(1 - e^{-\frac{n}{s}}\right)^q \quad (25)$$

2.1 Practical Verification of the Proposed Formula

Given formulas above, a simulator has been implemented which two main aims:

1. To verify the correctness of formulas.
2. To bear out that the approach of storing in parallel channels optimizes the results in terms of several variables which may be considered.

The simulator is implemented allowing the following inputs:

1. Total number of slots to implement (T).
2. Number f voters (N).
3. Quantity of parallel channels to implement (Q).
4. Quantity of election acts that will be simulated by session (R).

The simulator verifies that the total number of slots (T) is a multiple of quantity of parallel channel, because the quantity of slots in each channel (S) must be an integer number.
When the simulation is complete, the following information may be obtained:

1. Total of successful votes (SV).
2. Total of lost votes (LV).
3. Quantity of runs where at least one vote is lost (R)
4. Quantity of runs (Votings) without lost votes ($RWLV$)
5. Quantity of runs (Votings) with lost votes (RLV)
6. Best case, that is to say, how many votes were lost in the most successful run (BC).
7. Worst case, that is to say, how many votes were lost in the less successful run (WC).

Therefore, we will observe the behavior of the formula (25) based on the next ratio:

$$SPLV = \frac{LV}{SV+LV} \tag{26}$$

Table 1 shows the values that were obtained in different simulations and the difference between those and the analytical results obtained by application of equation (25). With this purpose, the following variables are introduced:

- FV: Values obtained by application of formula (25).
- SV: Values obtained by simulation.

Watching the values of Table 1, the difference between FV and SV remains at very low values. Specifically:

- The maximum one is 0,002452561.
- The minimum one is 2,80437E-07

– The average value is: 0,000227181

N	*T*	*S*	*Q*	*FV*	*SV*	*DIFFERENCE*
15	150	15	10	0,010185894	0,008466667	0,001719227
15	300	30	10	8,89424E-05	0	8,89424E-05
15	450	45	10	3,35005E-06	0	3,35005E-06
15	600	60	10	2,80437E-07	0	2,80437E-07
30	300	30	10	0,010185894	0,007733333	0,002452561
30	600	60	10	8,89424E-05	0	8,89424E-05
30	900	90	10	3,35005E-06	0	3,35005E-06
30	1200	120	10	2,80437E-07	0	2,80437E-07
60	600	60	10	0,010185894	0,0098	0,000385894
60	1200	120	10	8,89424E-05	0,00015	-6,10576E-05
60	1800	180	10	3,35005E-06	0	3,35005E-06
60	2400	240	10	2,80437E-07	0	2,80437E-07
120	1200	120	10	0,010185894	0,012025	-0,001839106
120	2400	240	10	8,89424E-05	0	8,89424E-05
120	3600	360	10	3,35005E-06	0	3,35005E-06
120	4800	480	10	2,80437E-07	0	2,80437E-07
240	2400	240	10	0,010185894	0,009129167	0,001056727
240	4800	480	10	8,89424E-05	0	8,89424E-05
240	7200	720	10	3,35005E-06	0	3,35005E-06
240	9600	960	10	2,80437E-07	0	2,80437E-07
360	3600	360	10	0,010185894	0,008391667	0,001794227
360	7200	720	10	8,89424E-05	1,11E-05	7,78313E-05
360	10800	1080	10	3,35005E-06	0	3,35005E-06
360	14400	1440	10	2,80437E-07	0	2,80437E-07

480	4800	480	10	0,010185894	0,00988125	0,000304644
480	9600	960	10	8,89424E-05	0	8,89424E-05
480	14400	1440	10	3,35005E-06	0	3,35005E-06
480	19200	1920	10	2,80437E-07	0	2,80437E-07

Table 1. Difference between *FV* and *SV*

Another aspect which should be highlighted is that the formula (25) works better when N<S. As both values approach, the behavior is worse. For example, Figure 3 shows the values of *FV* and *SV* with de following values for the parameters:

– $N = (7..15)$
– $T = 150$
– $Q = 10$
– $S = 15$

Fig. 3. *PLV*: Difference Between *FV* AND *SV*

3. Conclusions

The approach based on parallel channels optimizes the use of storage space intended to store data whose location is truly random. The formulas (1), (2), (3), (4) y (25) accurately describe the behavior of the model.
Specifically, the results obtained by the formula (25) are very close to the values obtained in the simulations, though the difference increases when N is close to S. Even in that case, the behavior of the formula is acceptable.

References

[1] Uzal R., van de Graaf J., Montejano G., Riesco D., García P.: "Inicio de la Línea de Investigación Ingeniería de Software y Defensa Cibernética". WICC 2013. Ps. 769 - 773. ISBN: 9789872817961. (2013).
[2] van de Graaf J., Montejano G., García P.: "Manejo de Colisiones en un Protocolo Non Interactive Dining Cryptographers". JAIIO 2013. ISSN: 18502776. WSegI 2013. ISSN: 23139110. Ps. 29 a 43. (2013).
[3] Chaum D.: "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability". Journal of Cryptology. (1988).
[4] van de Graaf J. : "Anonymous One Time Broadcast Using Non Interactive Dining Cryptographer Nets with Applications to Voting Towards Trustworthy Elections". Ps 231-241. Springer-Verlag Berlin, Heidelberg. ISBN: 9783642129797. (2010).
[5] Flajolet P., Gardy D., Thimonier L.: "Birthday Paradox, Coupon Collectors, Caching Algorithms and Self-Organizing Search". Discrete Applied Mathematics 39. Ps. 207-223. North-Holland. (1992).
[6] García P., van de Graaf J., Hevia A., Viola A.: "Beating the Birthday Paradox in Dining Cryptographer Network"s. The third International Conference on Cryptology and Information Security in Latin America, Latincrypt 2014. September 17-19, 2014. Florianopolis, Brasil. Lecture Notes in Computer Science, Springer (2014).
[7] García P., van de Graaf J., Montejano G., Bast S., Testa O.: "Implementación de Canales Paralelos en un Protocolo Non - Interactive Dining Cryptographers". JAIIO 2014. ISSN 18502776. WSegI 2014. ISSN: 23139110. (2014).
[8] García P., Montejano G., Bast S.: "Aspectos Optimizables en un Protocolo Non-Interactive Dining Cryptographers". CONAIISI 2014. ISSN: 23469927. (2014).
[9] García P., Montejano G., Bast S, Fritz E.: "Anonimato en Sistemas de Voto Electrónico: Últimos Avances". WICC 2016. ISBN: 9789506983772. (2016).
[10] García P., van de Graaf J. Montejano G., Riesco D., Debnath N., Bast S.: "Storage Optimization for Non - Interactive Dining Cryptographers (NIDC)". Information Technology New Generations (ITNG). Ps. 55 - 60. ISBN: 978-1-4799-8827-3. DOI: 10.1109/ITNG.2015.15. IEEE. (2015).
[11] García P., Bast S., Fritz E., Montejano G., Riesco D., Debnath N.: "A Systematic Method for Choosing Optimal Parameters for Storage in Parallel Channels of Slots". International Conference on Industrial Technology (ICIT)- Ps. 1700 - 1705. DOI:10.1109/ICIT.2016.7475019 IEEE – 2016.
[12] van de Graaf J., Montejano G., García P.: "Optimización de un Esquema Ocuppancy Problem Orientado a E – Voting". WICC 2013. Ps. 749 - 753. ISBN: 9789872817961. (2013).
[13] Feller W.: An Introduction to Probability Theory and its Applications.Volúmen I. Third Edition. John Wiley and Sons. New York. (1957).

Procedure for an empirical Detection of Anomalous or Unsafe Public Key Infrastructures

ANTONIO CASTRO LECHTALER[1,2], MARCELO CIPRIANO[1], EDUARDO MALVACIO[1]

{antonio.castrolechtaler; cipriano1.618; edumalvacio}@gmail.com

[1] CriptoLab, Escuela Superior Técnica, Facultad del Ejército, Universidad de la Defensa UNDEF - Cabildo 15, C1426AAA, Ciudad Autónoma de Buenos Aires, Argentina; [2] Facultad de Ciencias Económicas, Universidad de Buenos Aires - UBA, Córdoba 2122, C1120AAQ, Ciudad Autónoma de Buenos Aires, Argentina

Abstract: This work presents an empirical procedure with codification capabilities to create a PKI Auditing Software. It consists of an estimation of statistical parameters resulting from a theoretical PKI –hence, free of bias and errors– and the parameters obtained from a real PKI. It can perform a comparative analysis as well as determine whether such a system matches behavior expectations, detecting error prone or unsafe systems.

Keywords: PKI, RSA**,** Digital Certificates.

1. Introduction

The Public Key Cryptography or Asymmetric Cryptography has several uses and applications. For instance, RSA System that enables "confidentiality" and "authentication" through encryption and digital signatures, respectively. In this framework, a Public Key Infrastructure (PKI) can be implemented in networks and other services. Could these systems contain errors that would make them vulnerable? Trusting their reliability, users may compromise their safety and become easy prey of attacks.

In order to detect "unacceptable" behavior in these systems, this work and its precedents submit a ***Probabilistic-Statistical Procedure for the Experimental Determination of Anomalous Public Key Infrastructures***. We analyze the PKI's behavior according to the distribution of prime factors that each "*digital certificate"* contains in its '*public module"*, which can be codified and thus create a **Software Auditor**. Sections 2 and 3 will discuss the general concepts of Public Key Cryptography, the RSA system, and generalities regarding PKIs. We shall also present findings regarding their weaknesses and security breaches.

Section 4 presents the formula that calculates a *Probability Function*, if no collisions or repetitions of prime factors are/are not found in samples of digital certificates. Taking into account the requirements for the calculation of large factorials, we also include alternate estimation formulas, which are not as complex as in the original procedure. Finally, Section 5 presents the

empirical procedure for PKI detection together with the formulas for probabilistic reference values.

2. Public Key Cryptography

2.1 Origins

For thousands of years, known as *Classic Age of Cryptography*, only the Private or Symmetrical Key Cryptography existed: the same key was used to code and decode messages. The sender and the receiver shared the same key.

This caused an unavoidable hurdle: the Key Exchange Problem. At a particular moment, both parties had to agree on the key. However, how was it possible to agree upon a key without compromising its safety?

In 1976, **Whitfield Diffie** and **Martin Hellman** published a work in which they solve the Key Exchange Problem, establishing the "**Diffie-Hellman Key Exchange**" (DH) [1].

In 1977, **Ronald Rivest, Adi Shamir,** and **Leonard Adleman**[1] published a technical memorandum at the **Massachusetts Institute of Technology (MIT)**. In their work, they presented a system that permits to code/decode messages avoiding the key exchange problem. The same principle fostered the notion of "*Digital Signature*", which provided an "*Authentication*" service to the existing "*Confidentiality*" service. In the following year – 1978 -, they published the **RSA System** worldwide [2].

The **Association for Computing Machinery (ACM),** created in 1947, offers the Turing award annually, the highest recognition in the discipline. For their significant contributions to cryptography, Rivest, Shamir, and Adleman received the award in 2002. Recently, in 2015, Diffie and Hellman also received this recognition [3].

The groundbreaking work of DH and RSA fostered the development of Public Key Cryptography or Asymmetric Cryptography. This cryptographic method uses *Modular Arithmetic* – an area of mathematics dealings with two problems still unsolved to these days[2].

2.2 Recently Discovered Vulnerabilities

Since its origins, Public Key Cryptography (PKC) has been the target of attacks and the scientific community has focused research on its vulnerabilities. To this date, the mathematical problems behind the

[1] At the authors' request, family names are not in alphabetical order. The name for the system RSA follows their family names and in that specific order.

[2] The "Problem of the Discrete Logarithm" (PDL) for the procedure proposed by Diffie y Hellman and the "Problem of the Factorization of Numbers" (PFN) for Rivest, Shamir and Adleman.

aforementioned cryptographic processes remain without significant advances. From that viewpoint, PKC is a safe framework.

Many of the problems and weaknesses relate to the implementation stage. Among the most recent, in 2015, several vulnerabilities known as **LogJam** [4] were documented for DH[3].

The same occurs for the RSA [5}, both at the hardware [6, 7] and at the software levels. It is worth mentioning in particular the research performed by Dr. Lenstra [8] who, along with other researchers, evaluated more than a million public key certificates and discovered that approximately 5% shared prime factors. Given the number of certificates having compromised safety: is it a feasible value considering the size of the analyzed sample; or is the magnitude of the analyzed modules and the number of possible primes beyond consideration?

3. Public Key Infrastructure

3.1. What is a PKI?

Public Key Infrastructures (PKIs) are widely found in military and civil environments and systems, in public or private networks, in LANs or WANs and even in the internet. A PKI provides its users with "certificates" that, among other applications, can be used for user logins and authentication, coding and digital signatures, non-repudiation, and determination of session keys.

The certificates issued by a **PKI**[4] include, among others, a module and an *e* number (usually 65537) known as "*public key*" and a number *d* called "*private key*". The *m* value, of a size *t* (measured in bits) results from the multiplication of two prime values. The triplet *(m, e, d)* is calculated by the **PKI** when requesting the corresponding digital certificate and is delivered to a particular user.

A PKI internal or specific vulnerability[5] occurs when an anomaly is observed during the calculation of the *m* values. The awareness of this vulnerability could enable an "attacker" to breach the security of the RSA algorithm and to obtain private keys, enabling access to information or replacement of user identity.

[3] To solve part of the problem, researchers suggested the creation of a Diffie-Hoffman group of 2048 bits, among other things.

[4] X.509 is a standard from the International Telecommunications Union (ITU) establishing, among other things, types and format of certificates and the validation algorithm.

[5] The safety of the RSA System is based on the difficulty to factorize modules *m* within a reasonable timeframe (i.e., t=1024, 2048 or 4096 bits as currently used) and thus preserve the secret *d* key. Knowing one of the prime factors in a specific module allows to calculate easily the other factor and the *d* key.

3.2 Why Vulnerabilities Occur

Current systems are extremely complex, making it difficult to detect particular errors [9]. The reading and control of all the code lines that make up the PKI could be a difficult task.

The detection of errors that weaken the safety of these systems has several precedents - for instance, in **Debian's OpenSSL** [10] digital certificates.

Are these errors simply innocent "bugs" that passed the tests and filtered through only to be detected years after their creation, or were they "planted" in order to weaken safety?

4. Calculation of the Probability Function to find (or not find) Collisions of Prime Factors

4.1. Obtaining the Formula

We introduce the formula to calculate the probability of two or more digital certificates that share primes within a sample size *mu*. (A more detailed version is presented in [11].)

Given a size *t*, measured in bits, of public modules *m*, let *b* be the size in bits of *m* prime factors. For instance, if *t=1024* the prime values will have size *b=512* bits.

Let P be the set of prime numbers size *b*.

$$P=\{p \,/\, p \text{ prime};\ 2^{b-1} < p < 2^{b}\}. \tag{1}$$

We can calculate the cardinal or number of *P* elements – hereby called *p* – with a formula associated with the *Prime Number Theorem*[6]:

$$p = Card(P) \approx \grave{A}(2^{b}) - \grave{A}(2^{b-1}). \tag{2}$$

$$p \approx \frac{2^{b}}{\ln 2^{b}} - \frac{2^{b-1}}{\ln 2^{b-1}} = \frac{2^{b-1}}{\ln 2}\left(\frac{2}{b} - \frac{1}{b-1}\right) \tag{3}$$

Let *M* be the set of all public modules that can be determined from the elements *P*:

$$M = \{m / m = pq;\ p \neq q;\ p, q \in P\}. \tag{4}$$

[6] Conjectured by the German mathematician **Carl Gauss** (1777-1855) and independently proven by the Belgian mathematician **Charles-Jean de la Vallée Poussin** *(1866-1962)* and French mathematician **Jacques Hadamard** (1865-1963).

The cardinal of *M* (to be shown here as *m*) is the number of subsets of two elements of P, given that each public module is the product of two prime values and the order of multiplication is irrelevant from commutative property.

$$m = Card(M) = \binom{p}{2} = \frac{p\ (p-1)}{2}. \tag{5}$$

Finally, let *R* be the set of all samples of *mu* modules in which there is no collision of primes. Its cardinality is determined by:

$$Card(R) = \prod_{i=1}^{mu} m_i = \prod_{i=0}^{mu-1} \binom{p\ -2i}{2}. \tag{6}$$

$$Card(R\ ) = \frac{\prod_{i=0}^{2(mu-1)} (p-i).}{2^{mu}}. \tag{7}$$

Thus,

$$Card(R\ ) = \frac{p!}{2^{mu}(p-2(mu-2))!}. \tag{8}$$

The **Laplace**[7]classic definition of probability is used to calculate the **Probability Function**. Assume that the PKI does not store nor does it register prime numbers already used. Then:

$$p(R) = \frac{card(R\ )}{m^{\ mu}}. \tag{9}$$

$$p(R\ ) = \frac{\dfrac{p!}{2^{mu}\left[p-2(mu-2)\right]!}}{m^{\ mu}}. \tag{10}$$

Solving:

$$p(R\ ) = \frac{p\ !}{\left[p\ -2(mu-2)\right]!\left[p\ (p1-1)\right]^{mu}}. \tag{11}$$

Its complementary probability would be:

$$p(\overline{R}\ ) = 1 - \frac{p\ !}{\left[p\ -2(mu-2)\right]!\left[p\ (p-1)\right]^{mu}}. \tag{12}$$

[7] Laplace definition of probability: favorable events over total events.

4.2 Calculation of Large Factorials

Formulas to estimate the value of large factorials follow, such as those used in the calculation of the probability function in formulas (11) y (12). The complexity of the original formula undermines the viability of calculation. Hence, the use of estimation formulas is a reasonable alternative.

$$n! = e^{\ln n!} \approx e^{n(\ln -1)}. \tag{13}$$

$$n! \approx n^n e^{-n} \sqrt{2\pi n}. \tag{14}$$

$$n! \approx n^n e^{-n} \sqrt{\pi} \sqrt[6]{8n^3 + 4n^2 + n + \frac{1}{30}}. \tag{15}$$

Formulas (13) y (14) are known as the **Stirling**[8] **formulas** and formula (15) as **Ramanujan**[9] **formula**.

$$n! \approx \sqrt{2\pi} \left(\frac{n + \frac{1}{2}}{e} \right)^{n+\frac{1}{2}}. \tag{16}$$

$$n! \approx n^n e^{-n} \sqrt{\pi} \sqrt{2n + \frac{1}{3}}. \tag{17}$$

$$n! \approx n^n e^{-n} \sqrt{2\pi \left(n + \frac{1}{6} + \frac{1}{72n} - \frac{31}{6480n^2} - \frac{139}{155520n^3} + \frac{9871}{6531840n^4} \right)}. \tag{18}$$

The latter three are known as **Burnside**[10], **Gosper**[11] and **Batir**[12] **formulas**, respectively.

5. PKI´s Auditing Algorithm

5.1. Reference Values for the Unbiased Binomial Distribution of a PKI

Formula (11) allows us to calculate the probability to find at least two digital certificates sharing a prime value within a sample size *mu* of digital certificates. There will be samples with and without repeated primes. This

[8] **James Stirling** (1692-1770). Scottish mathematician.

[9] **Srinivasa Ramanujan** (1887-1920). Indian mathematician. He did not leave a proof of his formula, which was demonstrated in 2000 by the Russian mathematician **Ekatherina Karatsuba.**

[10] **William Burnside** (1852-1927). English mathematician.

[11] **Ralph Gosper, Jr.** (1943-). American mathematician and computer scientist.

[12] **Necdet Batir** (1959 -). Turkish mathematician.

represents a **Bernoulli experiment**: aleatory and independent tests or experiments with two possible and complementary results, called "*success*" and "f*ailure*".

Thus, we obtain the **Binomial Distribution**:

$$X \sim B(n, p(R)). \tag{19}$$

$$f(x) = \binom{n}{x} p(R)^x [1 - p(R)]^{n-x}. \tag{20}$$

Where x is a discrete aleatory variable representing the number of successes/failures[13] found in n tests or samples size *mu;* and *p(R)* is the probability of finding/not finding collisions of primes in a sample size *mu*, as shown by formula (11).

Some of the parameters of the probability distribution are the *media* and the *variance* (with which to calculate the standard deviation, or square root of the variance):

$$\overline{x} = np(R). \tag{21}$$

$$s^2 = np(R)[1 - p(R)]. \tag{22}$$

5.2 Experimental Detection of Anomalous Public Key Infrastructures

We have finally reached the main objective of this work and its supporting research: the experimental detection of anomalous public key infrastructures through empirical means. The procedure is susceptible to being automated and thus creating a PKI auditing algorithm.

We have accepted as a working hypothesis the existence of "*statistical permanence*": the experimental procedure applied to the PKI can reveal unknown behavior through statistical tools. Discrepancies found between the reference values and those obtained by "*direct experience*" imply an anomaly in the PKI behavior.

The statistical procedure is as follows:

1. Determination of the parameters to assess the **PKI:** size *m* of the modules and, therefore, size *b*, the size of its prime factors.
2. Calculation of the value *p*, according to formula (3).
3. Determination of the size of each *mu* sample.
4. Calculation[14] of the value *p(R)*, according to formula (11).

[13] An alternative procedure involves choosing the probability that the sample does have any repeated primes or their complementary. Both are calculated in (11) y (12), respectively.

[14] The selection of the approximation formulas for large factorials is possible, as shown in paragraph 4.2

5. Calculation of parameters, according to formulas (21) y (22).
6. Request of *digital certificates* from the PKI in order to group them together in *n* samples of *mu* size.
7. Count the samples that have/do not have repeated prime numbers (searching for collisions).
8. Use of the values obtained in the previous step to calculate the empirical median and variance parameters.
9. Comparison of the theoretical values shown in 5 with those obtained in 8.
10. Determination whether the **PKI** shows an anomalous behavior, according to the comparison performed in 9.

There are other alternatives to determine the existence of repeated primes in a sample. The PKI could show the prime values used to calculate the public module of each certificate. If no information on prime numbers is available, the procedure shown in [12] is viable.

6. Conclusions and Future Research

We have presented formulas to determine the behavior of an ideal and unbiased PKI, free from vulnerabilities. The statistical procedure presented herein can determine empirically the behavior of a real PKI.

The comparison of both values (the one expected theoretically and the one verifiable empirically) enables us to determine whether a behavior is anomalous or not in a specific **Public Key Infrastructure**.

Public Key Infrastructure Auditing Software shall make use of this procedure. Future research work shall aim to verify the procedure and to analyze concerns that not been addressed yet.

- Will estimation formulas for the calculation of large factorials create uncertainties that will interfere with the determination of anomalies?
- Which is the best formula to apply in the procedure, considering a computational cost/benefit criterion?
- What is the acceptable gap between theoretical and empirical parameters?

Acknowledgements. This work is part of a Technological and Social Development Project (PDTS, for its initials in Spanish) [13] of the Superior Technical School – Engineering School belonging to the National Defense University (UNDEF, for its initials in Spanish).
This project is supported by the Institute for Scientific and Technical Research for Defense (CITEDEF, for its initials in Spanish), and the Professional Council of Telecommunication, Electronics, and Computer Engineering Sciences (COPITEC, for its initials in Spanish). To them, the authors wish to express their deep appreciation.

References

[1] Diffie, W. y M.E.Hellman. "New directions in cryptography", IEEE Transactions on Information Theory 22 pp. 644-654. 1976.

[2] R. Rivest, A. Shamir, L. Adleman. “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems.” Communications of the ACM, Vol. 21 (2), pp.120–126. 1978.

[3] http://amturing.acm.org/byyear.cfm consultada el 20/6/2016.

[4] Adrian, D.; Bhargavan, K.; Durumeric, Z.; Gaudry, P.; Green, M.; Halderman, J.; Heninger, N.; Springall, D.; Thomé, E.; Valenta, L.; VanderSloot, B.; Wustrow, E.; Zanella-Béguelin, S.; Zimmermann, P. “Imperfect Forward Secrecy: How Diffie-Hellman Fails in Practice.” Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Pages 5-17. ACM New York, NY, USA. 2015.

[5] Boneh, D. Twenty Years of Attacks on the RSA Cryptosystem. Notices of the American Mathematical Society, Volume 46, Number 2. Providence, 1999.

[6] Chen, S.; Wang, R.; Wang, X.; Zhang, K. “Side-Channel Leaks in Web Applications: a Reality Today, a Challenge Tomorrow”. IEEE Symposium on Security & Privacy. Oakland, 2010.

[7] Pellegrini, A.; Bertacco, V.; Austin, T. Fault-based attack of RSA authentication. Proceedings Design, Automation & Test in Europe Conference & Exhibition. The IEEE Council. Dresden, 2010.

[8] Lenstra, A; Hughes, J; Augier, M and others. Ron was wrong, Whit is right. E-print International Association for Cryptologic Research. 2012. http://eprint.iacr.org/2012/064.

[9] Glass, Robert “Facts and Fallacies of Software Engineering”. Addison-Wesley Professional, 2003.

[10] Bello L, Bertacchini M. “Generator of Pseudo-Aleatory Numbers Predictable in Debian”. III International Conference on Computer Safety. Manizales, Colombia. October 2009.

[11] Castro Lechtaler, Antonio; Cipriano, Marcelo; Malvacio, Eduardo. “Detection of Anomalous Public Key Infrastructures”. XXI Argentine Conference on Computer Sciences CACIC 2015. Junín, Buenos Aires. October 2015.

[12] Cipriano, M. “Factorization of N: recovery of prime factors from public and private keys”. XIV Argentine Conference on Computer Sciences. CACIC 2008. Chilecito, La Rioja. October 2008.

[13] http://www.iese.edu.ar/investigacion.html#antecedentes consultada el 20/6/2016.

V

Innovation in Computer Science Education Workshop

Educational Software for a Discrete Event Simulation Introductory Course

DARÍO WEITZ

Facultad Regional Rosario – Universidad Tecnológica Nacional, Zeballos 1341, Rosario
dar.wtz@gmail.com

Abstract. Simulation is a required course in numerous engineering careers. Course contents usually have theoretical definitions and metodological issues, high level of abstraction and computationally intensive procedures. An educational software that includes animation elements is described. It is a didactic tool developed to facilitate the comprehension and learning of concepts, methods, computational procedures and modeling approaches included in a discrete-event simulation introductory course for an information systems engineering career. The educational software consists of four modules corresponding to to waiting-line systems, inventory systems, healthcare systems and a particular module developed to compare three different approaches in discrete-event simulation modeling: event-scheduling approach; activity scanning; process approach. Developed based on current theories of multimedia learning, principles for successful animated graphics and practical heuristics, the didactic tool proved to be appropriate for educational purposes at the undergraduate level. The educational software contributes to improve the short term exam success rate.

Keywords: simulation; discrete-event; educational software; animation; didactic tool

1. Introduction

Most real-world systems show such complexity that they inhibit their resolution through standard analytical models. For such reason, the simulation technique is used as an alternative approach to model the system under study and to calculate certain estimates of interest in order to gain some understanding of the system behavior. Simulation is the process of building a mathematical or logical model of a system or a decision problem, and experimenting with the model to obtain insight into the system behavior or to assist in solving the decision problem [1].

Simulation is widely used in operations research and management science; for this reason, it is included as a required course in numerous engineering and management science careers. The subject is a nexus between some mathematics, statistics, systems and operations research contents in Ingeniería en Sistemas de Información at the Universidad Tecnológica

Nacional (UTN) in Argentina. Course contents usually have theoretical definitions and methodological issues, high level of abstraction, computationally intensive procedures and demand students with a solid mathematical and statistics background.
Simulation software packages such as SIMUL8 or Arena are usually used during discrete-event simulation introductory courses. However, they exhibit significant limitations at the time the instructor, at the very beginning of the course, attempts to explain the basic concepts, time-advance mechanisms, event lists, different approaches (event-scheduling approach, activity scanning, process approach), computational procedures and the evolution of discrete-event simulation models. In this sense, prior to the use of simulation environments in application exercises, it is convenient to have some didactic tool in order to complement the teaching task when explaining the previously detailed topics.
Computers have provided new possibilities and strategies for innovative learning environments. Computer-based instruction, educational multimedia systems and animated graphics are frequently employed for teaching complex systems and abstract concepts. Educational software is defined as those computer programs developed in order to be used as facilitators of the teaching process, and the learning process as well [2]. Educational software share five essential characteristics: i) have a didactic purpose; ii) use the computer as support; (iii) are interactive allowing dialogue and information exchange; iv) have the possibility to individualize the speed of learning; v) are relatively easy to use [3].
Computers also provide the ability to integrate animation as part of the teaching-learning process. Animation is described as a pictorial display that changes its structure or other properties over time and which triggers the perception of a continuous change [4]. In educational environments, animations are often used to improve students' understanding of certain complex processes, particularly those that are not easy to describe verbally. Animation has been employed in teaching complex systems that change over time and space, systems affected by simultaneous influences or the sensitivit y of a process to changes in a particular variable [5]. In a system affected by simultaneous influences, animation helps to overcome our natural tendency to process information sequentially.
In the present paper, we describe an educational software which integrates animation as part of its learning strategy. It is a didactic tool developed to facilitate the comprehension and learning of concepts, methods, computational procedures and modeling approaches included in a discrete-event simulation university course for an information systems engineering career.

2. Didactic Strategy

Simulation is an undergraduate course (4th year) for an engineering degree in Information Systems in the UTN of the República Argentina. Simulation has been found a useful tool for design, analysis and evaluation of manufacturing

systems, service organizations and government institutions. At the end of the course, students should be able to predict, explain, train or identify appropriate solutions in waiting-line systems, inventory systems, manufacturing and assembly facilities, and in health, financial, educational and government organizations.

An educational software to be used as support in the classroom was developed. It is a didactic tool that integrates cognitive processes of analysis, synthesis, classification and deduction. It is student-oriented with control of learning contents, but it is also related to the needs of the teacher in what concerns to the curricular relevance [2].

Weitz [6] incorporated animation elements in modules designed for teaching various educational contents habitually included in a Classical Control Theory (CCT) course. Based on current theories of multimedia learning • cognitive theory of multimedia learning [7], epistemic fidelity theory [8]• , principles for successful animated graphics • congruence principle, apprehension principle [9], and practical heuristics [10], an efficient didactic tool for explaining theoretical definitions, underlying methodologies involved in diagrams construction and computational procedures related with feedback control systems was developed. Due to the success of the previously described strategy, and given the similarities between Classical Control Theory and Simulation (systems that evolve over time and space, rigorous and computationally intensive procedures, simultaneous influences in closed systems, etc.), it was decided to expand the conceptual idea and the gained experience during the development of the CCT educational software to certain Simulation contents which present some complexity for the instructor at the very beginning of the course.

3. Educational Software

The educational software was developed using Microsoft Visual C#. Microsoft Visual C# is a multi-paradigm, general purpose, object oriented programming environment used to create computer applications for the Microsoft Windows family of operating systems. It combines the C# language and the .NET Framework. Zedgraph was employed for graphics and drawings. ZedGraph is a class library, Windows Forms UserControl, and ASP web-accessible control for .NET, written in C# for drawing different kind of charts.

The educational software consists of four modules corresponding to computational procedures used in discrete-event simulation models to calculate measures of performance of systems such as those described in the didactic strategy. They correspond to waiting-line systems, inventory systems, healthcare systems and a particular module developed to compare three different approaches in discrete-event simulation modeling: event-scheduling approach; activity scanning; process approach. Modules are projected in the classroom in enlarged format using a laptop and an appropriate projector.

The module *waiting-line systems* shows the animation of a single-server queueing system where customers arrive according to a specific probability distribution of the interarrival times, and they are served by a dedicated server, also with a specific probability distribution of the service times. The server chooses a customer from the queue according to a FIFO discipline. The model assumes an infinite population of customers and a finite waiting area. Exponential, uniform or normal probability distributions can be chosen for interarrival or service times.
Clicking on a button named "Simular" marks the beginning of an animation of customers arriving to the queueing system (Figure 1). The display shows the value of the simulation clock, the server status, number of customers in queue and the departure time for the customer being served. Denial of service occurs when the number of clients in queue exceeds a preset value. Clicking on a button named "Pausa" stops the animation in order to access the display of two graphs: i) Q(t), number of customers in queue at time t; ii) B(t), utilization of the server at time t. The graphs are used to facilitate the explanation of two measures of performance which are usually computed in waiting-line systems: 1) q(n), expected average number of customers in the queue; 2) u(n), observed proportion of time during the simulation that the server is busy.

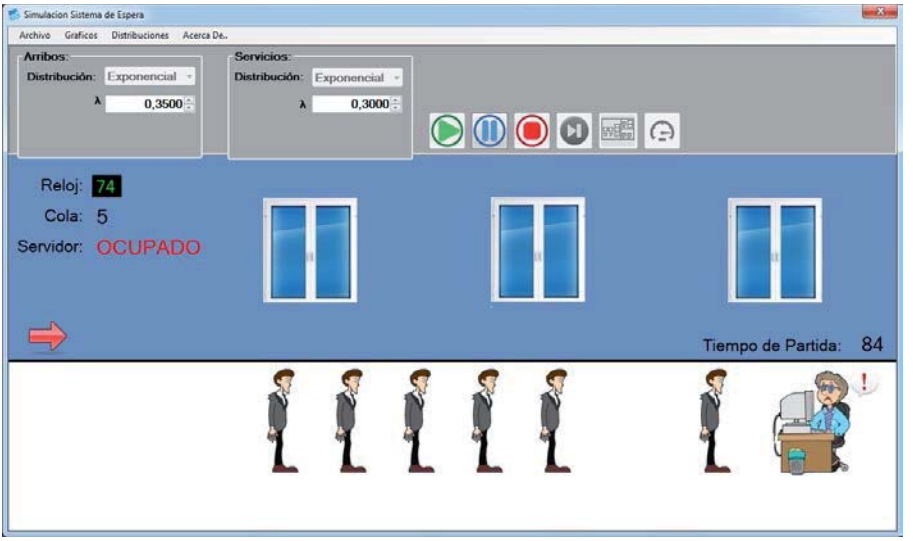


Fig. 1: Waiting-line system animation

Law & Kelton [11] in their book "Simulation Modeling and Analysis" propose an intuitive explanation of a single-server queueing system by means of a simulation exercise. The idea is to show snapshots of the system at succeeding event times, including data structures, state variables changes, statistical counters updates, and the interaction between the simulation clock and the event list. The waiting-line module allows the instructor to show

simultaneously the system physical representation, Q(t) and B(t) graphs, and the currently values of the state variables, simulation clock, event list and statistical counters at the end of each event (Figure 2).

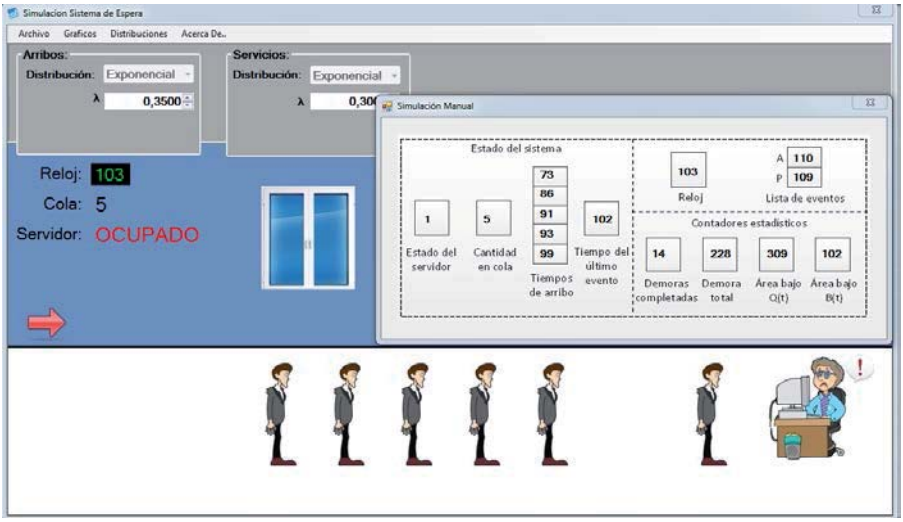


Fig. 2: Snapshot of the waiting-line system module

The module *inventory systems* computes the average monthly total cost as a performance measure in order to compare ordering policies for an inventory system. The display shows the animation of a warehouse where customers arrive according to a specific probability distribution of the interarrival times, and demand a variable quantity of items according to an empirical distribution. Additional data (initial inventory level, end of simulation month, distribution and parameters for the delivery lag) and figures for the computation of ordering, holding and shortage costs must first be entered (Figure 3). The ordering policy is a stationary one with parameters *s* (inventory lower level) and *S* (inventory upper level).

Clicking on a button named "Animación" marks the beginning of a representation of customers arriving to the warehouse and the satisfaction of the demand at least the inventory level is as large as the demand. Clicking on a Pause button stops the animation in order to access the display of three functions used to calculate holding and shortage costs: I(t), inventory level at time t; $I^{+}(t)$, number of items held in inventory at time t; $\Gamma(t)$, number of items in backlog at time t. The three functions and their corresponding area under the curves are shown (Figure 4). The module allows the carrying out of an inventory system simulation as an intuitive explanation of the simulation technique. The idea is to show snapshots of the system at succeeding event times, allowing the instructor to show simultaneously the system physical representation, I(t), $I^{+}(t)$ e $\Gamma(t)$ plots, and the currently values of the event list,

simulation clock, state variables and statistical counters at the end of each event.

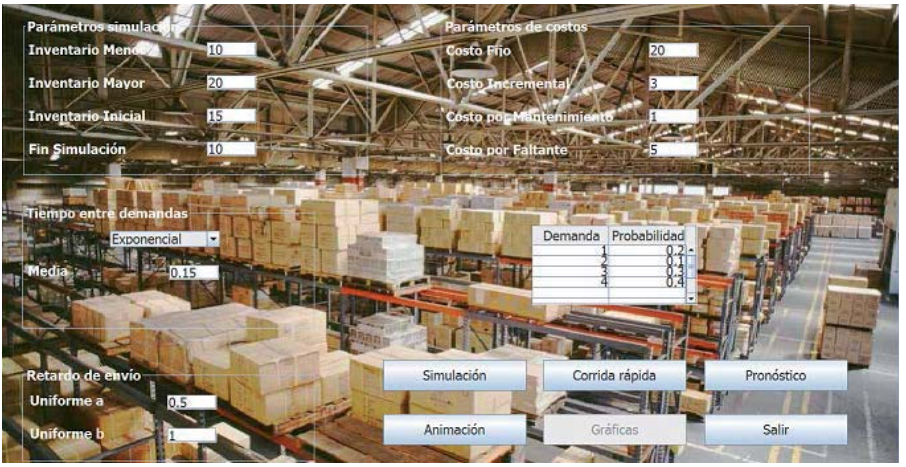


Fig. 3: Data entry screen for the inventory system

Fig. 4: Plots of I(t), $I^{+}(t)$ e $I^{-}(t)$

Both modules incorporate a button that allows a complete simulation • without animation• , where the parameter to be varied is the length of the simulation run. At the end of the run, estimates and confidence intervals of the performance measures and plots of the evolution of the estimates along the run are shown. The objective is to allow the instructor to explain the transient and steady-state behavior of the stochastic process which characterizes the output of waiting-line and inventory systems such as those previously described in the course. The modules also included a button named "Pronóstico" which allows the realization of 500 independent runs in

order to show frequency charts of different performance measures for an exhaustive analysis during the development of the topic Output Data Analysis.

The module for simulation of telemonitoring as a patient management approach attempts to assess different organizational alternatives in a healthcare institution willing to incorporate patients monitored remotely with wearable sensors [12]. Figure 5 describes the system and the objective of the simulation study.

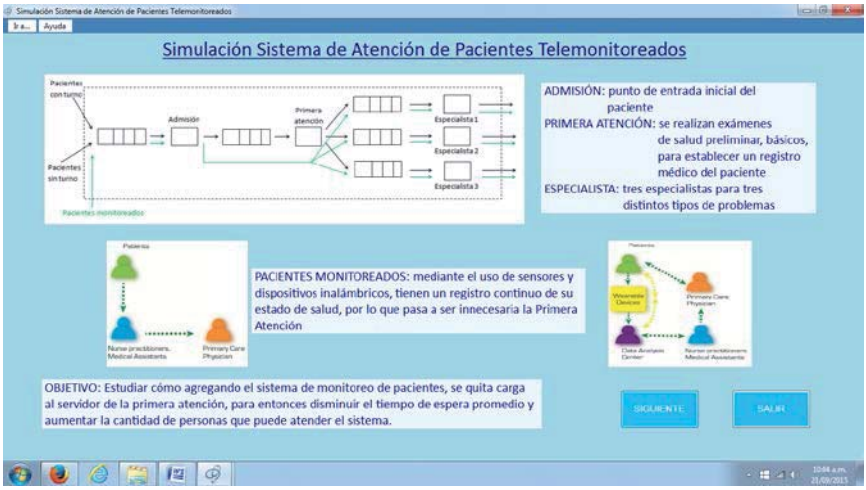


Fig. 5:: Telemonitoring healthcare system simulation model

Clicking on a button named “Siguiente” and after completing the data entry, begins an animation of the arrival and medical care of four different types of patients: with appointment, without appointment, telemonitored without appointment, telemonitored with appointment. The animation can be observed in two (2D) or three (3D) dimensions. Figure 6 shows the 3D version and it helps the teacher to show how the simulation technique can be used for training or for decision making in a very complex facility. The module also provides the recourse of answer a variety of *what-if* questions for every performance measure of the system under study.

Fig. 6: 3D view of the telemonitoring healthcare system

One of the content topics more difficult to explain in a simulation course is related with three different approaches commonly used in discrete-event simulation modeling: i) event-driven approach; ii) activity scanning; iii) process-driven approach. The first approach, also named event-scheduling approach, describes the changes that occur in the system at the instant of time that each event occurs. Events are sequenced in chronological order and may not correspond to a natural flow of entities [1]. Otherwise, process-driven simulation describes the flow of a typical entity through the whole system, and requires the use of special-purpose simulation software for coding the specific simulation model [11]. The activity scanning approach consists of sequences of activities waiting to be executed, and simulation proceeds from event to event executing those activities whose conditions are satisfied. An activity is defined by a couple of events: one to begin and the following to complete an operation that transforms the state of an entity.

We explain the event-driven approach by means of an animation that includes two instructional points: i) the flow of control showing the logical relationships among different components in the simulation model; ii) snapshots of the system at succeeding event times, including state variables, statistical counters, event list, and the simulation clock (Figure 7). The animation highlights the activity that is running in the invoked routine and the update of the system state to account for the fact that a particular type of event has occurred. We explain the process-driven approach by means of another animation that also includes two instructional points: i) the flowchart that describes the “experience” of a “typical” client as it flows through the system [11]; ii) a table that is being completed with the values needed to compute the desired measures of performance. The module includes specific animations for waiting-line systems and for inventory systems. It also includes equivalent animations for the activity scanning approach.

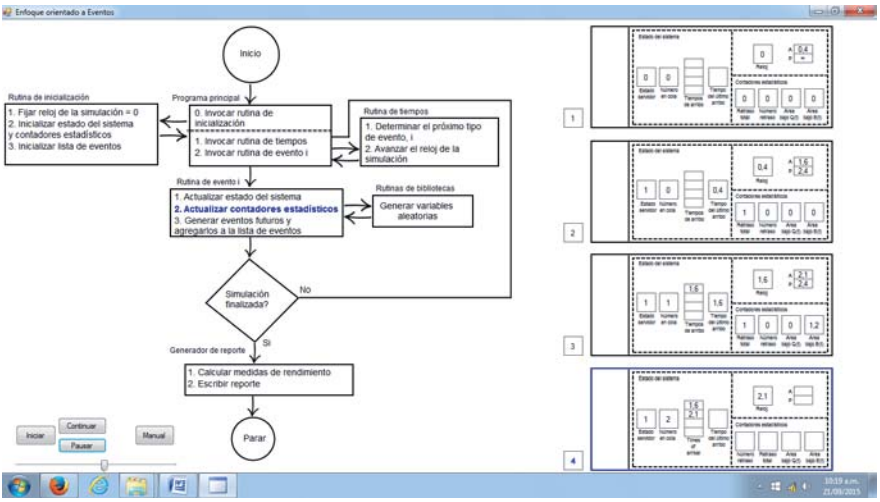


Fig. 7: Animation for the event-driven approach

4. Discussion

The development of an animated educational software for explaining theoretical definitions, underlying methodologies, computational procedures and modeling approaches such as previously described is justified for the following reasons: 1) systems under study evolve over time and space; 2) movement and trajectory are present on the subject matter; 3) content and format of the images • external representation• correspond with content and format of the concepts presented to the students (Congruence Principle); 4) the external representation portrays the instructor's mental model with high fidelity and clarity (Epistemic Fidelity Theory); 5) the educational software acts as a presentation function giving a visual context to ideas difficult to describe verbally; 6) in a second stage, it acts as a clarification function by improving understanding of new concepts or new methods with little additional textual information; 7) modules were specifically designed taking into account the Apprehension Principle (simple images and appropriate animation speed).

Course contents change significantly since 2005 when a new curriculum core was implemented. The educational software first stage (waiting-line systems, inventory systems) was incorporated in 2013. The second stage (telemonitoring healthcare system, modeling approaches) was included during the 2015 academic year. Short term exams average success rates during previous years (2005 – 2012) were 63.4%. After using the first stage of the instructional software (2013 – 2014), percentage rates increased to 76,7%, and to 84.6% during the 2015 academic year when the second stage was implemented.

5. Conclusions

Simulation is an university course for an engineering degree in Information Systems at the Universidad Tecnológica Nacional (UTN) of the República Argentina. Most university engineering and management science curricula have a simulation course as a required one. At the Facultad Regional Rosario of the UTN, engineering students struggle to understand certain topics taught at the very beginning of the course, which hindered the proper learning of the following contents.

Blackboard and textbooks are neither sufficient means to properly visualize the evolution of a dynamic system, nor to analyze the complexity of computationally intensive procedures and modeling approaches included in a discrete-event simulation university course. For this reason, an educational software that includes animation elements in modules designed for explaining waiting-lines systems, inventory systems, transient and steady-state behavior of a stochastic process, frequency charts, what-if analysis, and modeling approaches was developed.

Based on current theories of multimedia learning, principles for successful animated graphics and practical heuristics, an efficient didactic tool for explaining discrete-event simulation concepts and methods was developed. The educational software described in this paper contributes significantly to improve the short term exam success rate.

References

1. Evans, J.R., Olson, D.L.: Introduction to simulation and risk analysis. New Jersey: P. Hall. (1998).
2. Cataldi, Z.: Metodología de diseño, desarrollo y evaluación de software educativo. Tesis de Magister en Informática, Facultad de Informática, UNLP. (2000). http://laboratorios.fi.uba.ar/lsi/cataldi-tesisdemagistereninformatica.pdf
3. Marqués, P.: El software educativo. (1996). http://www.lmi.ub.es/te/any96/marques_software/
4. Schnotz, W., Lowe, R.K.: A unified view of learning from animated and static graphics. in: Lowe, R.K., Schnotz, W. (eds.) Learning with animation: Research implications for design. New York: Cambridge University Press, 304-356. (2008).
5. Rieber, L.P., Kini, A.: Theoretical foundations of instructional applications of computer-generated animated visuals. Journal of Computer-Based Instruction, 18(3), 83-88. (1991).
6. Weitz, D.A.: Effectiveness of Animation as a Learning Strategy in a Classical Control Theory Introductory Course. World Journal Control Science and Engineering, Volume 3, 1, 8-12. (2015).
7. Zoabi, W., Sabag, N., Gero, A. (2012). Using Animation to Improve the Student's Academic Achievement on Bipolar Junction Transistor. American Journal of Engineering Education – Fall 2012, 3, 2.
8. O'Donnell, F.: Simulation frameworks for the teaching and learning of distributed algorithms. Ph.D. Thesis, University of Dublin College, (2006). http://www.tara.tcd.ie/bitstream/handle/2262/1277/TCD-CS-2006-20.pdf?sequence=1&isAllowed=y

9. Tversky, B., Morrison, J.B., Betrancourt, M.: Animation: can it facilitate?. Int. J. Human-Computer Studies, 57, 247-262. (2002).
10. Weiss, R.E., Knowlton, D.S., Morrison, G.R.: Principles for using animation in computer-based instruction: theoretical heuristics for effective design. Computers in Human Behavior, 18, 465-477. (2002).
11. Law, A.M., Kelton, W.D.: Simulation Modeling and Analysis. New York: Mc Graw-Hill. (2000).
12. Weitz, D.A., Lianza, F., Nant, J.P., Schmidt, N., María, D.E.: Modelo de Simulación 3D para la Evaluación de Tecnologías de Monitoreo y Asistencia para Adultos Mayores. 3er Congreso Nacional de Ingeniería Informática / Sistemas de Información, CONAIISI 2015, Buenos Aires. (2015).

Experience with Augmented Reality. How It Affects Understanding of Control Structures

NATALI SALAZAR MESIA[1,2], GLADYS GORGA[2], CECILIA SANZ[2,3].

[1] Type A Fellow – National University of La Plata
[2]Institute of Research in Computer Science III-LIDI. School of Computer Science – National University of La Plata.
{nsalazar, ggorga, csanz}@lidi.info.unlp.edu.ar

Abstract. An experience using the hypermedia educational material EPRA, which incorporates Augmented Reality activities for using and applying control structures in beginner Programming courses, is presented. One of the aspects considered during the experience was the impact of control structures on learning. A diagnostic assessment was carried out with 192 students of the course Concepts Related to Algorithms, Data and Programs (CADP) and 84 students of the course Programming 1 of the Computer Science courses of studies of the UNLP to expose the difficulties students face in comprehending and applying control structures. Then the experience was carried out with the participation of 24 students from the course CADP and 30 students from the course Programming 1. The results obtained in this experience show an improvement in subject comprehension, while at the same time, Augmented Reality was successful at motivating students to learn.

Keywords: Augmented Reality, Impact on Learning, Digital Educational Material

1. Motivation

Augmented Reality (AR) appears as a technology with a significant potential for the educational scenario [1, 2]. The authors in [3] state that: "… [the contextualization that offers AR] allows students to acquire and learn, but also understand, how the concepts learned in the classroom can be applied to solve problems in real world situations. In these contexts, AR allows students gain a deeper understanding of their learning, by relating learning contents to their own experiences".

On the other hand, freshmen today are digital natives. In [4], the authors explain the following, after the research carried out by Prensky [5]: "… digital natives [are] the first generation that has been brought up with digital technologies and, as such, are 'natives' to computer language, video games and the Internet, while digital immigrants are those who have not been brought up in a digital world, but they have adopted some aspects of this

technology". The authors in [4] also characterize digital natives as individuals that do several tasks at the same time (multitasking), require immediate responses, prefer graphics to text, create videos, and use multimedia presentations, among other characteristics.

The experience presented here takes the statements of these research works as starting point, and proposes innovative changes to the teaching of basic programming concepts, in this case, in relation to the topic of control structures, for students who are mostly digital natives. Our proposal includes using Augmented Reality activities to experience and use control structures by solving simple problems with an immediate visual representation of the effects of choosing the different control structures, modifying the real scenario with virtual objects relevant to the problem being considered. To achieve this, experimental sessions were carried out in 2015 and 2016 as part of the courses Concepts Related to Algorithms, Data and Programs (CADP) and Programming 1 of the courses of studies given at the School of Computer Science of the UNLP.

In this paper, we present the results obtained in these experimental sessions. Our starting point is a diagnostic analysis of subject comprehension by 192 students taking CADP and 84 students taking Programming 1. Then, a reference group is selected from each of the courses to use the EPRA material and carry out the activities proposed. Finally, a post-test is administered to the group taking part in the experience. Initial results indicate that there is an improvement in the comprehension of the control structures included in the material developed. Additionally, it would seem that the use of Augmented Reality successfully motivates students to learn the topic selected [6].

This paper is organized as follows: Section 2 provides background in relation to the use of AR to learn abstract concepts; Section 3 describes the digital educational material EPRA; Section 4 presents the experience; Section 5 discusses the results obtained based on pre-test and post-test analysis; and finally, Section 6 presents our conclusions and future work.

2. Augmented Reality in Educational Scenarios

AR is a technology that allows integrating the real world with the digital world. As described by its name, the technology 'augments', or enhances, the real scenario through the addition of virtual objects [3]. Markers, QR codes, the position of the individual, and/or pieces of the scene can be used to record the virtual objects in the real scene.

The contribution of AR is increasingly acknowledged by education researchers, and it emphasizes the coexistence of virtual objects and real environments which allows students view abstract concepts and complex spatial relations. This situation is particularly valued in the context of Computer Science teaching [7]. In [8], the authors state that there is consensus among educators in relation to the need for student interaction with the contents being taught. From this standpoint, it seems clear that AR technology will be a positive contribution to the learning process of the

students due, among other reasons, to the high level of interaction it provides. Similarly, there is also agreement in relation to the already mentioned reference described in [3], which values the possibility that students have to carry out their own experiences about the topics to be learned. Below, some examples of how AR is used in various educational environments, mainly focused to the introduction of abstract concepts, are presented.

In [9], the authors describe an experience carried out in an informal learning environment, namely, a visit to a Mathematics exhibition at a museum that uses AR to provide additional information. There are two groups: one that participates in an interactive manner using AR, and another one that visits the sample using the traditional method. The goal set forth by these authors was comparing the mathematical knowledge acquired by both groups. To that end, they used a pre-test and a post-test that helped them analyze the knowledge of each participant. Based on the results presented, it is concluded that the addition of AR to an informal context (museum) is effective for acquiring Mathematics-related knowledge, and that it helps both content comprehension and long-term retention.

The authors in [10] also used this pre-test and post-test methodology for an experience carried out to learn how to fly a drone. All participants completed a pre-test about basic drone concepts, and then they were divided into two groups, one working with interactive contents with Augmented Reality and another one using a more traditional approach with text, graphics, audio and video. An analysis of post-test results indicate that 77% of the participants in the first group needed less time to fly a real drone, while only 33% of the members in the second group were able to do so in the same period of time. These results have been encouraging for the work we present here, and they have been a stepping stone for designing our experience.

3. EPRA Digital Educational Material

The educational material EPRA is a website[1] (with *Creative Commons* non-commercial license) that includes theoretical content and AR-based activities that use markers and detect the face of the person interacting to augment the scene [11]. It also includes an introductory section aimed at providing information about the educational material and a help section with tutorials. There are three types of AR activities: Exploratory Activities, Review Activities, and Integration Activities. The different types of activities included in EPRA are described below. A more detailed description of the educational material and its activities can be found in [6, 11, 12]:

A. Exploratory Activities: this type of activities is aimed at help students experience each of the control structures (If, While, Repeat and For) through real-life and game situations that are presented to them (fig.1a).

[1] Website: http://163.10.22.174

B. Review Activities: these activities will allow students observe and compare the behavior of the different control structures. Options are presented as follows: If *vs.* While, While *vs.* Repeat and For *vs.* While, where students, based on the instructions given, choose one control structure and then receives the corresponding feedback (fig 1b).
C. Integration Activities: these are designed so that students apply previously learned knowledge. They consist of a game that asks students to select a control structure based on certain personal characteristics. As a consequence of the selection, different virtual objects (glasses, hat) are added to the face of the student, creating an effect on the real scene a real that will change how the student looks (fig. 1c).

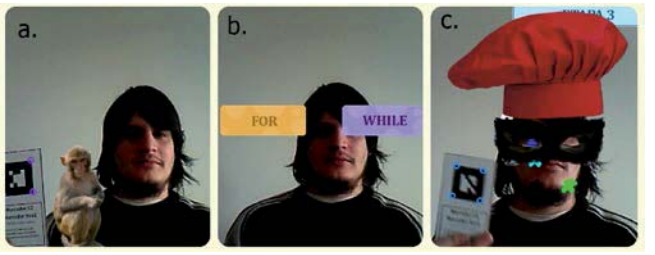


Fig. 1: Augmented Reality activities: (a) Exploratory Activity, (b) Review Activity, (c) Integration Activity.

4. Experience carried out

The experience carried out consisted in carrying out work sessions using EPRA material, both with students and educators from the already mentioned courses. For assessment, the focus was on student and educator satisfaction with EPRA, intrinsic student motivation when carrying out AR activities, and impact on learning measured by the use of the most suitable control structure to solve simple problems. This paper will focus only on this last aspect. The other assessment lines will be discussed in other publications. This section provides relevant background information for the experience, with a description of the courses and the profile of participating students. Then, the results of the general diagnostic test for all students are discussed, followed by a description of the test sessions carried out with the group involved in the experience with EPRA and the specific results for the students that participated in the experience.

4.1 Characteristics of the Courses

The experience is carried out with regular students from two beginner's courses of the School of Computer Science of the UNLP. To teach their respective curricula, both courses are organized into theoretical and practical

classes. Theoretical classes are usually large in attendance (around 200 students per band) and are given by the Head Professor, while practical classes are somewhat reduced groups (around 50 students per class) and are given by assistant educators who help students solve practical exercises designed to put into practice the contents discussed in the theoretical classes.
The course Concepts Related to Algorithms, Data and Programs (CADP) is one of the first year courses of the Bachelor's Degree in Computer Systems, Bachelor's Degree in Computer Science, and University Programmer Analyst. Students attending this course have previously taken the subject Expression of Problems and Algorithms (EPA), given as part of the entry course to the three university study courses mentioned above, and whose passing requirement is 80% attendance rate.
On the other hand, the course Programming 1 is one of the first year courses of Computer Engineering and, unlike the previous course, to be able to take it, students must have previously passed the course Introduction to Computer Science, given as part of the entry course to the university study course.

4.2 Characteristics of Participating Students and Diagnostic Test Results in Relation to Control Structure Learning

The students taking the courses mentioned above are mostly male with an average age of 19 years old in CADP and 21 years old in Programming 1. It should be noted that participating students had been previously introduced to the concept of control structure in each of the courses, CADP and Programming 1, (specifically, they had learned about If, While, Repeat and For), and they had also put that knowledge into practice to solve simple problems.
Students complete a diagnostic test aimed at establishing their major difficulties in relation to the topic at hand: use and application of control structures for solving simple problems. This test is administered during a practical class of the corresponding course, and is completed by all attending students on that particular day. This totaled 192 CADP students and 84 Programming 1 students. The test is a multiple-choice written test that presents five simple problems. Working individually, each student must choose the most suitable control structure to solve each of the problems.
The results indicated that there were no major difficulties in relation to the use of the IF control structure. The sample indicates that around 79% of the students answered correctly, while 21% chose a different control structure: 9% chose WHILE, 5% selected FOR, 2% selected REPEAT, and 5% did not choose an option. The same analysis was carried out with the other control structures; results are shown below in figure 2.

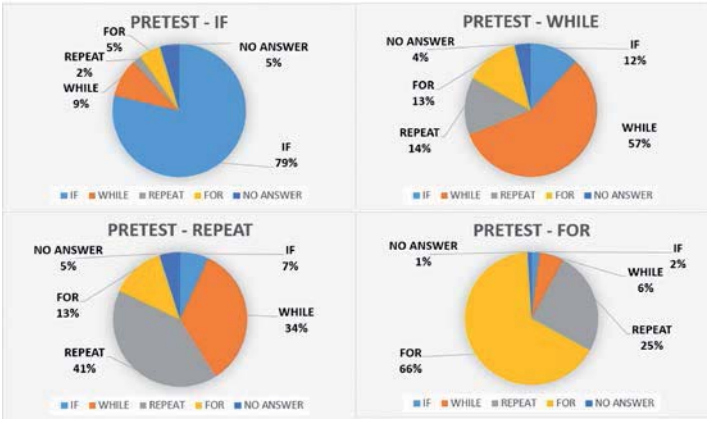


Fig. 2: Results of the diagnostic test (CADP)

Figure 2 shows that there was a high percentage of students who got the wrong answer when using the conditional iterative control structures WHILE and REPEAT. The most common error is mistaking WHILE for IF, with a similar percentage of students who have trouble distinguishing FOR and REPEAT, while REPEAT is directly associated to WHILE.

As regards the 84 pre-test completed by Programming 1 students, the results show that there are no major difficulties in relation to the pre-conditional iterative (WHILE), repetitive (FOR), or decision (IF) control structures. However, the post-conditional structure REPEAT presents the highest percentage of errors when applying it to a specific problem. For instance, the following problem is presented: "Student name, last name, and mark are read up to 'Juan Perez'. Choose the most appropriate control structure to calculate the average mark for all read students, including the last one." When responses were analyzed, the following percentages were found: 57% of the students answered correctly, 37% chose WHILE, 4% chose FOR, 1% chose IF, and 1% did not choose an option. The same analysis was carried out for the other control structures (fig. 3).

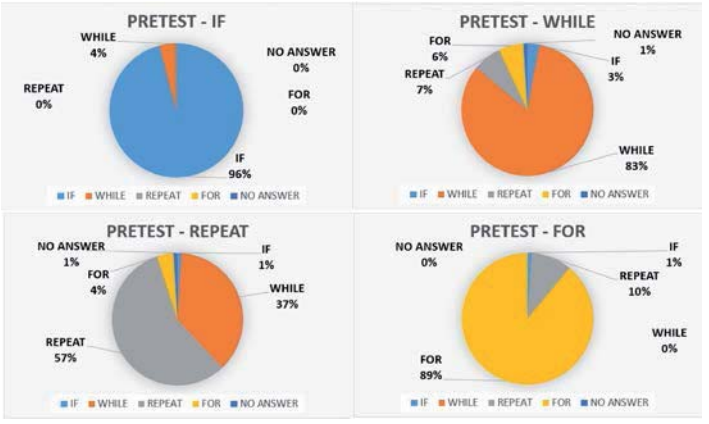


Fig. 3: Results of the diagnostic test (Programming 1)

The results discussed here show that students have some difficulty when they are asked to select the most suitable control structure to solve a simple problem, and these weaknesses are then reflected on midterms, where this type of problems is still present. These results have motivated the development and use of the educational material with Augmented Reality activities presented here.

4.3 Experimental Sessions

The group that carried out the experience with EPRA consisted of 24 CADP students and 30 Programming 1 students who had completed the diagnostic test. The sample was selected randomly, and is representative of the population under study. The number of participating students is conditioned by the availability of resources (classrooms, computers, webcams, etc.) required to carry out the experience, as well as by attendance to the practical class on the day of the experimental session [12]. Four sessions were carried out, with a maximum of 15 students and approximately 40 minutes each, in a suitable classroom of the School of Computer Science. When selecting the classroom, aspects such as availability of computers with access to the Internet and webcams, appropriate lighting and computer location to avoid marker detection issues were considered, among others. Students worked in pairs; and a computer was assigned to each pair of students. Each session followed these steps: a. Educators presented the educational material EPRA and explained how to use it. Students carried out the AR activities proposed using EPRA and then were asked to complete a post-test, which had a similar structure to that of the diagnostic test.

5. Result Analysis

The analysis carried out using the results of the tests completed by students (diagnostic test, now considered as a pre-test for participating students) and post-test, completed by participating students after carrying out AR activities) is presented in this section. Below, post-test results are discussed by comparing them with the diagnostic tests results obtained by the same students. The post-test was completed by the 24 CADP students and the 30 Programming 1 students that took part in the experimental sessions. Results are analyzed separately for each group, since the initial conditions for each group were different and this could affect the results.

As regards the results obtained with the CADP sample for control structure WHILE, it is observed that 92% of the students answered correctly, while only 8% chose an incorrect control structure. Four percent of the students chose decision control structure IF, and 4% chose FOR. These percentages show an improvement from what these same students had selected in their diagnostic test, which had an overall percentage for WHILE of 38%, 29% for IF, 21% for FOR, 8% for REPEAT, and 4% with no answer. Figure 4 shows

a graphic representation of the comparison between the pre-test and the post-test in relation to WHILE.

Fig. 4: Pre-test/Post-test comparison for WHILE (CADP).

As regards FOR, an improvement is also observed in the post-test versus the pre-test. In the pre-test, only 54% of the students selected this structure as the most suitable to solve the problem; the remaining 46% was distributed as follows: 30% chose REPEAT, 12% chose WHILE, and 4% did not answer (Figure 5).

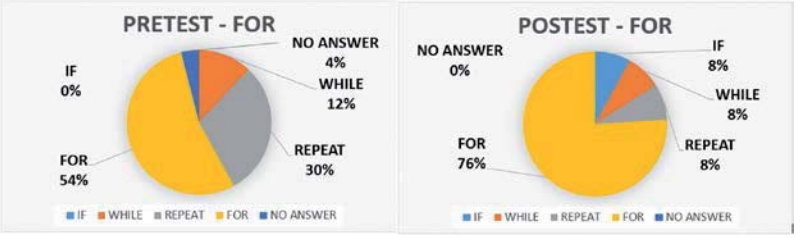


Fig. 5: Pre-test/Post-test comparison for FOR (CADP).

As regards IF, there is a slight improvement in the use of the structure - 80% of the students selected it correctly in the pre-test, versus 88% in the post-test. Finally, as regards REPEAT, even in the post-test students seem to struggle with it, although there is still a slight improvement (Figure 6).

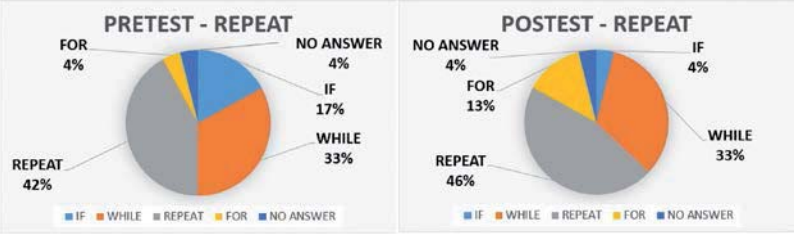


Fig. 6: Pre-test/Post-test comparison for REPEAT (CADP).

As for the post-test results of Programming 1 students, there are no significant improvements when compared to the pre-tests.
However, in some cases there is a slight improvement when selecting the correct control structure to solve a simple problem, while in other cases there are no significant differences or there is even a decrease in the percentage of correct answers. We present here only the results corresponding to control structure WHILE, for which an improvement was observed. The problem that involved answering by choosing control structure WHILE was responded with WHILE by 80% of the students in the pre-test, and 87% of the students in the post-test. These results are represented in Figure 7.

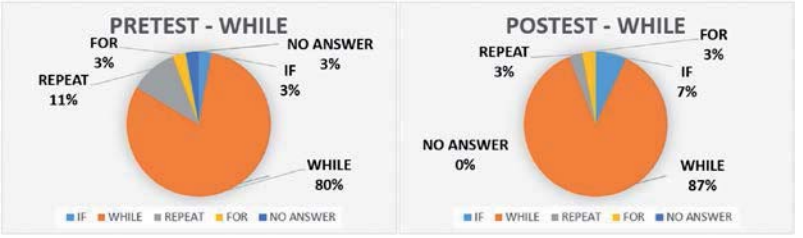


Fig. 7: Pre-test/Post-test comparison for WHILE (Programming 1).

As regards IF, there was a high percentage of correct answers both in the pre-test and the post-test (above 90%), REPEAT being the control structure with the lowest rate of correct answers - 60% in the pre-test and a decrease to 55% in the post-test. Finally, there are no significant difference observed for FOR between the pre-test and the post-test, with a rate of correct answers above 90% in both cases.

6. Conclusions and Future Work

An experience using the hypermedia educational material EPRA, which incorporates AR activities for applying control structures to solve simple problems in beginner Programming courses, was presented. To determine if AR activities have in any way changed student performance (incidence), a group of students was randomly selected from the courses mentioned above, and the results obtained in a pre-test were analyzed, then AR activities were carried out, and a post-test similar in nature to the pre.-test was administered.
The results obtained with CADP students differ from the results obtained with Programming 1 students. The initial conditions of these courses are different, since Programming 1 students have already passed (approved an exam) the entry course, which means that they have a better comprehension of its contents, including control structures, while CADP students may or may not have passed the entry course (the only requirement is attendance to the entry course). That is, this group is more heterogeneous. The only control structure that neither of the groups had learned during the entry course was

REPEAT, which is the one with which students from both groups had the most difficulty, although CADP students showed an improvement in the use of this structure after working with EPRA. The incidence of AR activities was higher among CADP students, who showed an improvement when choosing the most suitable control structure to solve a problem. This was observed with all control structures. Programming 1 students had already better results in their pre-tests, with the exception of REPEAT and WHILE. In the case of WHILE, a positive incidence of AR activities is observed, but no improvement was observed for REPEAT. This makes us think, on the one hand, that the use of control structure REPEAT and how it is different from the other structures, should be reinforced. On the other hand, the AR activities that were designed for this study are considered to be important in terms of student learning, especially those students who are just starting to learn about control structures, since it was the less advanced group of students the one that presented the greatest improvement. The activities also had a positive impact on student motivation (see [12]).

There are still more thorough analyses to be carried out with these results based on the combination of the three assessment lines proposed; that is, motivation, satisfaction, and incidence on learning. However, the information obtained so far is of interest in itself, and a starting point for other educators and researchers in the area.

References

1. Johnson, L., Adams, S., & Cummins, M.: The NMC Horizon Report: 2012 Higher Education Edition. Austin, Texas: The New Media Consortium (2012).
2. Chen, C. M., & Tsai, Y. N.: Interactive augmented reality system for enhancing library instruction in elementary schools. Computers & Education, 59, 638-652. doi:10.1016/j.compedu.2012.03.001 (2012).
3. Cabero, J., Barroso, J.: The educational possibilities of Augmented Reality. Journal of New Approaches in Educational Research, 5(1), 44-50. doi: 10.7821/naer.2016.1.140, (2016).
4. García García, F., Gértrudix Barrio, F., Durán Medina, J., Gamonal Arroyo, R., Gálvez de la Cuesta, M. C.: Señas de identidad del 'nativo digital'. Una aproximación teórica para conocer las claves de su unicidad. CDM Cuaderno de Documentación Multimedia. V.22 ISNN: 1575-9733, (2011).
5. Prensky, M.: Digital Natives, Digital Immigrants P1. On the Horizon. V.9. pp. 1-6, (2001).
6. Salazar Mesía, N., Sanz, C., Gorga, G.: Experiencia de enseñanza de programación utilizando Realidad Aumentada. XXII Jornadas sobre la Enseñanza Universitaria de la Informática, Actas de las XXII Jenui. pp. 213-220. Almería, (2016).
7. Tamim R. M., Bernard, R.M., Borokhovski, E., Abrami, P. C., Schmid, R. F.: What Forty Years of Research Says about the Impact of Technology on Learning a Second-Order Meta-Analysis and Validation Study. Review of Educational Research, V.81, No. 1, pp. 4-28, (2011).
8. Roussou, M.: Learning by Doing and Learning Through Play: An Exploration of Interactivity in Virtual Environments for Children.Computers in Entertainment

(CIE) - Theoretical and Practical Computer Applications in Entertainment, 2 (1), pp.1-23, (2004).
9. Sommerauer, P., Müller, O.; "Augmented reality in informal learning environments: a field experiment in a mathematics exhibition", Computers & Education, V.79, pp. 59-68, (2014).
10. Salamanca Díaz, D. M.: Creación de contenido educativo con realidad aumentada aplicando los principios de la teoría cognitiva del aprendizaje multimedia. Estudio comparativo para enseñar cómo volar un dron (cuadricóptero). Computing Colombian Conference (10CCC), 10th, Bogota, pp. 456-462, ISBN: 978-1-4673-9464 (2015).
11. Salazar Mesía, N., Gorga, G., Sanz, C.: EPRA: Herramienta para la Enseñanza de conceptos básicos de programación utilizando realidad aumentada. X Congreso de Tecnología en educación y Educación en Tecnología. ISBN 978-950-656-154-3, (2015).
12. Salazar Mesía, N., Gorga, G., Sanz, C.: Plan de evaluación del material educativo digital EPRA. Propuesta de indagación sobre la motivación intrínseca. XXI Congreso Argentino de Ciencias de la Computación CACIC. pp. 414-423. ISBN: 978-987-3724-37-4, (2015).
13. Adell, J., Castañeda, L.: Tecnologías emergentes, ¿pedagogías emergentes? En J. Hernández, M. Pennesi, D. Sobrino & A. Vázquez (Coords). Tendencias emergentes en educación con TIC. pp.18-63, Barcelona: Editorial espiral, (2012).

V

ETHICOMP LatinAmerica

Group Study Experience. First Introduction to Autonomous Weapons

FEDERICO OTARAN[1], LAUTARO DE LEON[1], JOAQUIN BOGADO[1], MARIA EMILIA CORRONS[1], MARIA BEATRIZ GARCIA[2], FRANCISCO JAVIER DIAZ[1]

[1] Laboratory of Investigation in New Information Technologies
LINTI - Computer Science School - UNLP
{ldeleon,ecorrons,federico.otaran}@cespi.unlp.edu.ar
{jbogado,jdiaz}@linti.unlp.edu.ar
http://linti.unlp.edu.ar/
[2] No affiliation
mariabeatriz.garcia@gmail.com

Translation and edition in English by Aldana Gómez Ríos - arios@mail.linti.unlp.edu.ar

Abstract This work is divided into two parts. In the first, we present a methodology that seems appropriate to address issues of professional ethics among peers, taken from other research groups. This technique, implemented, allowed us to choose and study a topic in some depth. The goal of the second part is to show the results of applying this methodology to the study of ethical issues related to autonomous weapons. These results are preliminary and part of a work in progress.

Keywords: computer ethics, autonomous weapons, moral machines

1. Introduction

"The ultimately unleashed Prometheus to whom science is hitherto giving unknown strengths and economics' unresting drive calls for ethics that detains its power by voluntary reins from causing harm to others." So begins the foreword, dated July 1979, to the original German version of the book by Hans Jonas titled "The Imperative of Responsibility: In Search of an Ethics for the Technological Age".[11]

To Jonas, it is the junction between science and technology that characterizes modernity. "Progress" tends, in fact, to change human life at an increasingly fast rate, which makes it necessary for us to rethink ethics. "What can serve us as compass? The anticipated danger itself!"[11].

With the idea of starting to study issues of professional ethics, in May 2016 we formed a study group, which through regular meetings, would allow us to investigate the way in which the emergence of new technologies influences society and, conversely, how society influences the emergence of new technologies and new uses for existing technology.

The mechanics proposed for the development of the study process was based

on the use of techniques associated with agile methodologies of development[20], although we also took ideas from study groups from other areas such as psychoanalysis.
This process was, however, tailored to our particular needs on the fly.
We decided to meet once a week for about 2 hours, in which each of us would present their new ideas regarding the research and, although we maintained some flexibility in schedules, the fact that the subjects were of interest to the group helped us keep motivated.
As a first step and during the first meetings, in addition to proposing and shaping the methodology, we made a first approach in brainstorming[21] style, in order to select a topic of study. The decision to address only one topic for the entire group and not one for each member was based, in principle, on the fact that the time and effort we could devote to the project was quite scarce and it was preferable to focus our efforts as much as possible. Among the proposed themes, we chose Autonomous Weapons for multiple different reasons, among which we can cite, without elaborating on the subject, that it seemed simpler, mainly because the scientific community has – for the most part – taken a stance on it[1][2][3]. Even Argentina has manifested itself to international organizations in response to the need to control this type of weapons[4][5].
From this, we follow a series of steps that constitute the backbone of this methodology. The first one was based on obtaining the necessary information about this subject from the various available bibliographical sources in order to have a clear idea of the state of the art of autonomous weapons today. Subsequently, based on the information collected and investigated, we tried to understand the arguments for and against the development of autonomous weapons. At the same time, during the face-to-face meetings, we debated the strengths and weaknesses of these arguments, and recorded the doubts that emerged when discussing and contrasting them with our ideas and subjectivities as regards autonomous weapon systems. It should be noted that many of them remained unanswered.
Far from being discouraged by the lack of answers, we are convinced that by showing that we can work on them, we will raise awareness of the need to discuss these issues in the community.

2. On our Background

The topic of professional ethics had been nurturing the curiosity of one of the members of this group and acquiring greater importance as an axis of investigation with the passage of time and informal conversations with people interested in this issue. In the context of CACIC 2011, the first Latin American ETHICOMP was held at the Computer Science School of the UNLP, showing the rise and importance of the subject at the national level.
In 2012, a first work was presented[22] also in the context of ETHICOMP II, and a second work [23] was presented in ETHICOMP III in 2013, seeking tocontinue some lines of work resulting from the work presented the previous

year, taking as reference the work of Matthew J. Sher. Participation and advancement in research on these subjects was always possible thanks to the unconditional support of the direction of the Laboratory of Investigation in New Information Technologies (LINTI).
It was necessary that some time elapsed for someone to take the initiative to form the research group we are presenting today in the scientific community, made up of four researchers from LINTI, students of the B.Sc. in Computer Science and the B.Sc. in Systems of the UNLP, two of them graduates and two in the final stage.
As Adela Cortina points out in her book [10] "Para que sirve realmente la etica?" (What is Ethics Really for?): "The world of the professions has a long history, usually told from the Western tradition, which is said to be born with the famous Hippocratic Oath, linked to a profession as valued as medicine. Two other professions would accompany him in the origins, that of the priests, and that of jurists, so that the three would deal with things as important to the life of a society as the good of the body, the good of the soul and that of the political community. Those who join a profession are committed to providing that good to their society, have to prepare for it by acquiring appropriate skills, and at the same time enter a community of professionals who share the same goal."
And a little later in this same book we can read something of the spirit that nests in this study or research group and marks its north: "In the face of the bureaucratic ethos of those who abide to the legal minimum, the professional ethos demands excellence, as his fundamental commitment is not that which binds him to the bureaucracy, but to concrete people, people of flesh and blood, whose knowledge gives meaning to any activity and social institution. It is therefore time not to disregard ordinary life, but to introduce into it the aspiration to excellence."

3. Autonomous Weapons

We understand by Autonomous Weapon, Lethal Autonomous Weapon System (LAWS) and Autonomous Robotic Weapons (ARW), any weapon system capable of making the decision to shoot at a target without the intervention of a human being, according to the definition of "Human-out-of-the-loop Weapons"[15]. So far, no such systems are known or are being used in service, however, there are highly automated defense systems such as the system marketed by Samsung, SGR-A1, used in the demilitarized zone that separates Korea from the North Korea. These systems require steps to become completely autonomous, and countries such as China, Israel, Russia and the United Kingdom have already expressed their interest in the development of this technology[16]. That an algorithm, i.e. a computer program should decide to shoot, thus ending the life of someone at the other end of the weapon, without another person pulling the trigger, presents some emerging problems relating to various factors involved in such a decision.

In his book "Drone Theory"[6], Chamayou wonders who would be responsible if a robot committed a war crime. The State who owns it? The military commander who uses it? The programmers who coded it? Müller[17] holds that LAWS do not mean substantial challenges to humanitarian laws or problems to determine liabilities, that the consequences of using this type of weapons would be generally positive, and therefore, that they should not be prohibited but regulated. However, he admits that some aspects of warfare could become worse[1] and that the International Humanitarian Law (Jus in Bello) principles of Distinction and Proportionality[2] could be compromised.

We in turn wonder, as software developers, whether it will be possible to design a machine that is incapable of committing a war crime[3] but which is nevertheless useful for its practical goals. Advantages that are constantly remarked in favor of autonomous weapons [7][8][9] include, among others:

- Replacing humans in the battlefield, thus reducing the number of victims.
- A machine is superior to a human in multiple aspects. The machine has a shorter reaction time and more precision, no need to rest and can operate with equal effectiveness in harsh environments and adverse situations.
- A machine does not experience feelings of rage, anger, fear, or craving for revenge.
- This would avoid possible violations of the rules of engagement and laws of war.
- Regarding the third point, in order for a robot to be able to follow the laws of war, it should be possible to tell them in some way, i.e., to translate them into an algorithm. This algorithm would make a distinction for the machine between what is morally correct and what is not. However, the work of M. Englert et al.[12] proposes limits to the moral action of a machine. In this work, the authors propose a *Gedankenexperiment*, a variant of the Trolley problem described by Thomson[13], in which:
- There is only one correct option[4] between two possible actions.
- All the information necessary to solve the problem is available.
- All the actions occur deterministically.
- And yet, it is fundamentally impossible to recognize the correct choice algorithmically.

---

[1] In particular, Müller says a reduction in the cost of war would make war more likely.

[2] The principle of Distinction refers to the ability to discriminate combatants from non-combatants and is crucial when selecting a target, while the principle of Proportionality requires that damage to civilians be proportional to the military aim.

[3] A war crime is a violation of the laws of war and the rules of engagement described in various international treaties.

[4] The authors understand "correct" to mean the morally acceptable choice and contrast it with another which is clearly not.

- The proposed scenario can be summarized in that the program should choose the more morally acceptable alternative based on determining whether the source code of a program contains malicious instructions that will actually be executed during the program run. Using theory of computation and program verification methodologies, the authors demonstrate that this problem cannot be solved for all inputs. In other words, there will be cases, particular situations for which an algorithm cannot decide which is the morally correct option, despite the points listed above.

Here, it seems reasonable to distinguish clearly two concepts that some authors use in an interchangeable way, such as W. Wallach and C. Allen in their book Moral Machines[14] – the concepts of Morals and Ethics. It seems possible, to these authors, to design a machine that distinguishes good from evil by following an algorithm, i.e., a moral machine. This behavior should be discussed in advance by the programmers of said algorithm. This task, which seems to be extremely difficult, would also be limited, as demonstrated by the formal methods described above. Nonetheless, we believe that acting ethically, and here lies the distinction from morality, is a strictly human and individual activity, involving other intrinsically human activities that are not programmable, such as the search for wellbeing, understanding and deliberation [24].

4. Drones and Autonomous Weapons

We also note a parallelism between some arguments in favor of the use of combat drones with military goals, and arguments in favor of autonomous weapons. Combat drones or Unmanned Aerial Vehicles (UAVs) differ from autonomous weapons in that the former have a human operator (two in the case of MQ-9 Reaper, a pilot and a sensor operator). On the one hand, the drone can project force without projecting vulnerability[6], i.e., the drone is able to inflict lethal force while guarding the lives of human opponents usually found in the command room. Thus, the use of drones can be understood to be able to "save lives". However, the lives saved are those of the combatants of the side that uses them, while the lives in peril are those of the military and civilians on the other side. In the same way, Autonomous Weapons can save lives, as, like the drones, they eliminate the combat and defense ability of the enemy. The war stops being war to become plainly a massacre, an annihilation of the enemy.

War, a state of usually open and declared armed hostile conflict between states or nations, [19] is the only one in which a person is allowed to take the life of another without this being considered a crime. Improving weapons with the goal of making war cheaper would lower the threshold to start a war rather than resolving differences by other means.

The use of drones as war devices by the Obama administration has been strongly questioned[18] and its various nuances are reflected in the film Eye

in the Sky directed by Gavin Hood and starring Helen Mirren, Aaron Paul, and Alan Rickman. In the film, it is clear how the rules of engagement are modified to end the lives of five well-known terrorists despite the possibility of collateral damage to the civilian population of an allied country. While many of the situations in the film pose the same questions if the drone was replaced by an autonomous weapon, we find that in the case of the autonomous weapon, there appear to be more sinister possibilities. Although these weapons may make life-or-death decisions, they will act on orders that may be in conflict with the laws of war, but if the weapon has no mechanism to refuse an order, it might be used to commit a crime.

Let us assume that the crime was committed, either accidentally or deliberately, and that the weapon will make the decision to shoot a person and end their life. Chamayou[6] argues that, on the one hand, there is the possibility that those responsible for the deployment or the programming, or the owners of such systems (e.g., the State or a private company) would point at each other, and on the other hand, the only person clearly identified as a responsible party would be the victim, since, through voluntary or involuntary acts, they activated the mechanisms for the weapon to fire.

5. The Argentinean Position regarding Autonomous Weapons

The official position of Argentina regarding Autonomous Weapons can be summarized in the following excerpt from a document presented at a debate before the Human Rights Council in Geneva on May 30, 2013[4]: “We wish to emphasize our concern on the references to the fact that these systems may lead to a 'normalization of the conflict' and a possible arms race that would create divisions among States and weaken International Law; the possible encouragement of reprisals, retaliation and terrorism; and its impact on human rights and international humanitarian law. As a way to avoid these negative consequences, the report concludes that there must be an international body in charge of assessing the situation and articulating the longer-term options.”

We understand that arms development is a global business and that total censorship to the research and development of this kind of devices could make many States uneasy. However, there are too many questions around autonomous weapons to encourage their development without there being guarantees, among which we can emphasize access to justice for victims of abuse of this type of weapon or lack of certainty as to the advantages that they provide.

6. Conclusions and Future Work

The methodology described allowed us to approach a topic of study as a group. In future iterations, we must improve aspects like systematizing note-

taking, references and extracts. We also found difficulties in supporting some arguments, probably due to ignorance of the main philosophical currents. We believe that a study of these currents would be beneficial to be able to apply them in the future to specific topics concerning computer ethics.

In addition, we consider that the position taken by Argentina on the subject of autonomous weapons is weak – it is not enough to have a body that regulates their use. It seems to us that Argentina should discourage their development instead of investing public funds in the investigation of these technologies.

As a future work, we also believe it necessary to share our concerns and reflections with our colleagues and with the community in general, as an integral part of the activities of the group and at regular intervals. In this way, the research group would be contributing to one of the main goals of the University, which is to raise awareness among future generations of professionals. These presentations should include works such as the one presented here, but also lectures, panels and discussion forums.

References

1. Stop Killer Robots Campaign - http://www.stopkillerrobots.org
2. International Committee for Robot Arms Control - http://icrac.net/
3. Autonomous Weapons: an Open Letter from AI & Robotics Researchers - http: //futureoflife.org/open-letter-autonomous-weapons
4. Document delivered by Mr. Mariano Alvares Wagner on behalf of GRULAC. 2013 http://www.stopkillerrobots.org/wp-content/uploads/2013/05/HRC_Argentina_ 09_30May2013.pdf
5. Country Statements about Killer Robots. 2014 – http://www.stopkillerrobots.org/wp-content/uploads/2013/03/KRC_CountryStatus_14Mar2014.pdf
6. Gregoire Chamayou. Drone theory. Kindle Edition, Penguin 2015.
7. William M. Fleischman. Why We Should Not Build Autonomous Robotic Weapons. ETHICOMP 2013.
8. Ronald Arkin. Lethal Autonomous Systems and the Plight of the Non-combatant. 2013.
9. Ronald Arkin. Ethical Robots in Warfare. www.cc.gatech.edu/ai/robot-lab/online-publications/arkin-rev.pdf
10. Adela Cortina. Para que sirve realmente la etica?. Paidos 2013.
11. Jonas Hans. Principio de Responsabilidad - Ensayo de una etica para la civilizacion tecnologica. Herder 1995.
12. M. Englert, S. Siebert, M. Ziegler. Logical Limitations to Machine Ethics with Consequences to Lethal Autonomous Weapons. CoRR 2014.
13. Judith J Thomson. The Trolley Problem. Yale Law Journal 1985.
14. Wendell Wallach, Colin Allen. Moral Machines: Teaching Robots Right from Wrong. Kindle Edition, Oxford University Press 2008.
15. Losing Humanity, a case against killer robots. Humans Right Watch 2012.
16. Shaking The Foundations: The Human Rights Implications of Killer Robots. Humans Right Watch 2014.
17. V. Müller, T. Simpson. Autonomous Killer Robots Are Probably Good News. PhilPapers, 2015.

18. The Forgotten Victims of Obama's Drone War. The New York Times 2013 http://www.nytimes.com/2013/05/23/opinion/the-forgotten-victims-of-obamas-drone-war.html
19. War. Merriam Webster online dictionary. https://www.merriam-webster.com/dictionary/war
20. Martin Fowler. The New Methodology. Blog post 2005. http://martinfowler.com/articles/newMethodology.html
21. Alex Faickney Osborne. Applied Imagination: Principles and Procedures of CreativeProblem Solving. Scribner 1953.
22. J. Bogado, B. García. Reflexiones iniciales acerca de la validez ética de la utilización de técnicas de minería de datos sobre datos personales en la búsqueda de terroristas. ETHICOMP 2012. 8 Group Study Experience.
23. M. B. Garcia, W. Fleischman, J. Bogado. Una conversacion con Matthew Sher sobre privacidad y la amistad. ETHICOMP 2013.
24. Aristoteles. Etica a Nicomaco. Gredos 2010.

Esta edición de 100 ejemplares
se terminó de imprimir en Estudiocentro,
Bolívar, Buenos Aires, Argentina,
en el mes de junio de 2017.

Its objectives are:
"Coordinate academic activities related to the improvement of the teachers' training as well as the curricular update and the use of shared resources to assist the development of both the Computer Sciences careers and the Technology careers in Argentina" and "To establish a cooperative framework for the development of Postgraduate activities in Computer Sciences and Technology, in order to optimize the assignation and use of the resources".

RedUNCI:

This Network was formally created through an Agreement signed in November 1996 by five National Universities (UNSL, UBA, UNLP, UNS y UNCPBA), during the second edition of CACIC.
Actually 58 argentine Universities are active members of this network.

Regular Activities of the RedUNCI

- Arrangement of an Annual Congress on Computer Science (CACIC) since 1995.
- Arrangement of an Annual Workshop for Researchers on Computer Science (WICC) since 1999.
- Meetings for university professors of Computer Science, for Postgraduate Dissertators and for specialists in certain areas, to promote the debate of common interest topics.
- Publication of *the Journal on Computer Science &Technology* by agreement with ISTEC (Iberoamerican Science and Technology Education Consortium).
- Annual Congress on Technology in Education and Education in Technologies (TE&ET) since 2006.
- Publication of the *Iberoamerican Journal of Technology in Education and Education in Technology*, since 2007.

ISBN 978-987-4127-28-0