Decisioning 2022

Collaboration in knowledge discovery and decision making: Applications to sustainable agriculture

UNLP, Facultad de Informática

La Plata, Argentina

June 30th to July 1st, 2022

Editors:

Alejandra Beatriz Lliteras

Facultad de Informática-LIFIA-UNLP-CICPBA, Argentina

Vanessa Agredo Delgado

Corporación Universitaria Comfacauca-Unicomfacauca, Colombia

Decisioning 2022

Collaboration in knowledge discovery and decision making: Applications to sustainable agriculture

UNLP, Facultad de Informática

La Plata, Argentina

June 30th to July 1st, 2022

Editors:

Alejandra Beatriz Lliteras

Facultad de Informática-LIFIA-UNLP-CICPBA, argentina

Vanessa Agredo Delgado

Corporación universitaria comfacauca-unicomfacauca, colombia

Decisioning 2022 : Collaboration in knowledge discovery and decision making : Applications to sustainable agriculture / Leandro Antonelli ... [et al.] ; Editado por Alejandra Beatriz Lliteras ; Vanessa Agredo Delgado. - 1a edición especial - La Plata : Universidad Nacional de La Plata. Facultad de I informática, 2023.
Libro digital, PDF

Archivo Digital: descarga
ISBN 978-950-34-2303-5

1. Agricultura Sustentable. 2. Metadatos. 3. Sistemas de Gestión de Bases de Datos. I. Antonelli, Leandro II. Lliteras, Alejandra Beatriz, ed. III. Agredo Delgado, Vanessa, ed.
CDD 630.2

Index

General Chairs

- Dr. DIEGO TORRES, Universidad Nacional de La Plata – Universidad Nacional de Quilmes, Argentina
- Dr. LEANDRO ANTONELLI, Universidad Nacional de La Plata, Argentina

Organizing Committee

- Dr. ALEJANDRO FERNÁNDEZ, Universidad Nacional de La Plata, Argentina
- Dr. MARIO LEZOCHE, Université de Lorreine, France
- Dr. PASCALE ZARATE, Université de Toulouse, France
- Dr. CÉSAR COLLAZOS, Universidad del Cauca, Colombia
- Dr. HERNÁN ASTUDILLO, Universidad Técnica Federico Santa María, Chile
- Dra. REGINA MOTZ, Universidad de la República, Uruguay

Program Committee Chairs

- Mg. ALEJANDRA BEATRIZ LLITERAS, UNLP-Facultad de Informática-LIFIA-CICPBA, Argentina
- Ing. VANESSA AGREDO DELGADO, Corporación Universitaria Comfacauca-Unicomfacauca, Colombia

About

Sustainable agriculture is one of the Sustainable Development Goals (SDG) proposed by UN (United Nations), but little systematic work on Knowledge Discovery and Decision Making has been applied to it.

Knowledge discovery and decision making are becoming active research areas in the last years. The era of FAIR (Findable, Accessible, Interoperable, Reusable) data science, in which linked data with a high degree of variety and different degrees of veracity can be easily correlated and put in perspective to have an empirical and scientific perception of best practices in sustainable agricultural domain. This requires combining multiple methods such as elicitation, specification, validation, technologies from semantic web, information retrieval, formal concept analysis, collaborative work, semantic interoperability, ontological matching, specification, smart contracts, and multiple decision making.

Decisioning 2022 is the first workshop on Collaboration in knowledge discovery and decision making: Applications to sustainable agriculture. It has been organized by six research teams from France, Argentina, Colombia and Chile, to explore the current frontier of knowledge and applications in different areas related to knowledge discovery and decision making. The format of this workshop aims at the discussion and knowledge exchange between the academy and industry members.

Selected papers

Defining the Language of the Software Application Using the Vocabulary of the Domain

Leandro Antonelli[1], Julio Leite[2], Alejandro Oliveros[3] and Gustavo Rossi[1]

[1]Lifia, Fac. de Informática, UNLP, La Plata, Bs As, Argentina
[2]Dep. Informática, PUC-Rio, Gávea, RJ, Brasil
[3]INTEC-UADE, Bs As, Argentina
{lanto,gustavo}@lifia.info.unlp.edu.ar
julio@inf.puc-rio.br
oliveros@gmail.com

Abstract. Requirements engineering is one the most critical stages in software development. If the requirements are not correct the software development team will produce an artifact that will not satisfy the needs, wishes and expectations of the client. Requirements (and knowledge in general) are spread among many stakeholders. Natural language is widely used since it is an adequate tool considering non-technical stakeholder. Nevertheless communication problems arise with the use of natural language. The software development team members need to learn about the application domain and this process of learning means focusing on the features to be included in the software application, while leaving apart the elements out of the boundaries of the application. This process is not easy when people face a new application domain. Thus, this paper proposes an approach to define the software application language from a vocabulary of the application domain.

Keywords. LEL, vocabulary, requirements, application domain, software application.

1 Introduction

Requirements engineering is a critical stage of software development. Errors made at this stage can cost up to 200 times to repair when the software is delivered to the client [7]. Requirements described as Use Cases or as User Stories define the goals, the scope and the functionality of the software system. Nevertheless, software applications are "packed knowledge about the domain" [11]. This knowledge needs to be captured in a complementary artifact to Use Cases and User Stories, for example in business rules [23] or given-then-when scenarios [27]. While goals and requirements for the software application can be elicited from a small group of people (the client or the sponsor) the knowledge of the domain relies in a wider group of stakeholder (the domain experts) who generally has a different and complementary point of view of

the domain. Thus, it is important to involve as many experts as possible to collaboratively [20] acquire their knowledge.

Experts and development team belong to different worlds and use different languages [26]. The experts use the language of the domain while development team uses a computer science language. In order to cope with this communication gap it is important to use artifacts in natural language that are readable by both parties [20]. Nevertheless, the use of natural language is not enough since both parties need to share a common language. Particularly the development team should adopt the language used in the application domain. This adoption is not easy because the application domain is broader than the software application. Hence, the knowledge in the application domain (and its representation through its language) sometimes is overwhelming for the development team members. Moreover, considering the amount of stakeholder in application domain (clients, users, sponsor, experts, etc.). Thus, it is hard for the development team to decide what is important regarding the boundaries of the software application. This paper proposes an approach to consider the language of the application domain (captured through its vocabulary) in order to reduce it to obtain the language limited to the boundaries of the software application.

The LEL is glossary [23] that has the aim of understanding the language of an application domain without worrying about the application software. The LEL categorizes terms in four categories (subjects, objects, verbs and states) and uses two attributes (notion and behavioral responses) to describe the terms. We believe that the LEL is a convenient tool because of three characteristics that we found in our experience: it is easy to learn, it is easy to use, and it has good expressiveness. We have used the LEL in many domains, some of them very complex, and we had good results. Cysneiros et al. [12] report the use of LEL in a complex domain as the health domain.

Our proposed approach uses the glossary LEL as input and obtains a new glossary LEL as output. Although the language used in the software application can omit, change or add concepts from the application domain, our proposed strategy only consider removing the elements of the application domain that are not relevant for the software. Thus, the proposed approach is a kind of "filter" with the aim to reduce the language of the domain to a subset that belong the boundaries of the application software. Hence, the proposed approach can also be considered as a process to define the scope of the application.

The rest of the paper is organized in the following way. Section 2 describes some preliminary knowledge needed to understand the approach. Section 3 describes the proposed approach. Section 4 provides evidence about the applicability and usability of the approach. Section 5 discuses some related works. Finally, section 6 presents some conclusion and future work.

2 Language Extended Lexicon

The Language Extended Lexicon (LEL) is a glossary that describes the language of an application domain, where not necessarily there is a definition of a software application. The LEL is tied to a simple idea: "understand the language of a problem

without worrying about the problem" [19]. The language is captured through symbols that can be terms or short expressions. They are defined through two attributes: notion and behavioral responses. Notion describes the denotation, that is, the intrinsic and substantial characteristics of the symbol, while behavioral responses describe symbol connotation, that is, the relationship between the term being described and other terms (Fig. 1). Each symbol of the LEL belongs to one of four categories: subject, object, verb or state. This categorization guides and assists the requirements engineer during the description of the attributes. Table 1 shows each category with its characteristics and guidelines to describe them.

Category: symbol
Notion: description
Behavioral responses:
Behavioral response 1
Behavioral response 2

Fig. 1. Template to describe a LEL symbol

Table 1. Template to describe LEL symbols according to its category

Category	Notion	Behavioral Responses
Subject	Who is he?	What does he do?
Object	What is it?	What actions does it receive?
Verb	What goal does it pursue?	How is the goal achieved?
State	What situation does it represent?	What other situations can be reached?

3 The proposed approach

This section describes the proposed approach in a general way, and after that it describes every step.

The approach in a nutshell

The proposed approach has the goal to analyze the glossary LEL used as input and select a subset of symbols and their descriptions, in order to provide a new glossary LEL as output. This output glossary LEL will describe the elements that would be inside the boundaries of a new software application that would be developed to provide support to the application domain.

It is important to mention that this proposed approach only considers the reduction of symbols and their description from the input glossary LEL to the output glossary LEL. And the proposed approach does not consider the modification of the descriptions or the additions of new ones.

The approach is based mainly in the relationship of the categories of the glossary LEL and some key elements in a software application design. The categories of the glossary LEL are: subject, objects, verbs and states. For the proposed approach, states are not used. Thus, symbols of category subject of the LEL are related with user roles in a software application, verbs of the LEL are related with functionality of the

application, and finally, objects of the LEL are related with databases of the application [2] [16].

The proposed approach consists basically of a succession of four steps: (i) classification of subjects, (ii) classification of behavioral responses of subjects, (iii) classification of behavioral responses of verbs, and (iv) classification of objects. Fig. 2 summarizes the steps.

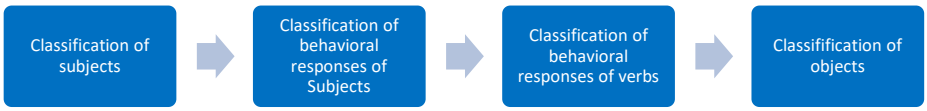


Fig. 2. Our approach in a nutshell

Classification of subjects

The classification of subjects pursues the goal of dividing the universe of discourse in three parts: (i) the software application to be developed, (ii) other software applications already developed, and (iii) the information system or behavior of the application domain that will remain manual, that is with no automatization.

Our approach is concerned about (ii) the software application to be developed. Nevertheless, the others two subsets are important because they will interact with the software application to be developed.

Thus, subjects of the glossary LEL should be categorized in one of the following categories: (i) subjects / users of the intended software application, (ii) subjects / users of another software application, and (iii) subjects / actors that will keep performing activities manually.

The rest of the paper will use an agriculture domain in order to provide examples of the proposed approach. Thus, we consider a farm that has the objective of growing fruits as business. The farmer is the person who has the technical agriculture knowledge. There are many field laborers who help the farmer. And there is an administrator who is in charge of taking the strategic decisions for the business.

The glossary LEL for this situation includes three subjects: farmer (Fig. 3), administrator (Fig. 4), and field laborer (Fig. 5). The farmer should be categorized as (i) subject / user of the intended software application, because the goal is to automatize some of their tasks. The administrator should be categorized as (ii) subject / user of another software application, because he already has a software application to manage the needs of the markets, the sales, the cash flow, etc. Finally, the field laborer should be categorized as (iii) subjects / actors that will keep performing activities manually, since he is in charge of cultural activities that consist in activities that cannot be automatized with machines. Fig. 6 summarizes this procedure.

Subject: farmer (user of the intended software application)
Notion: responsible to grow the fruits.
Behavioral responses
The farmer fertilizes spraying.
The farmer fertilizes watering.

Fig. 3. Subject farmer

Subject: administrator (user of another software application)
Notion: responsible to maintain a positive balance in the cash flow of the business.
Behavioral responses
The administrator decides the fruits to plant.

Fig. 4. Subject administrator

Subject: field laborer (actor that will keep performing activities manually)
Notion: responsible to labor tasks in the field.
Behavioral responses
The field laborer performs cultural activities.

Fig. 5. Subject field laborer

for each subject s with the glossary LEL
categorize s as
(i) user of the intended software application
(ii) user of another software application
(iii) actors that will keep performing activities manually

Fig. 6. Procedure for Subject categorization

Classification of behavioral responses of subjects

The behavioral responses of the subjects denote the actions (activities, tasks) that subjects perform within the application domain. Thus, subjects categorized as "subject / user of the intended software application" will be users of the software application and some of their behavioral responses would be functionality that will be included in the software application to be developed. This second steps of the approach, consists in analyzing the behavioral responses of the subject previously categorized as "subject / user of the intended software application", and each behavioral response should be categorized as: (i) functionality of the intended software application, (ii) functionality of another software application, and (iii) activities to keep performing manually. In some situations, all the activities of the subjects selected could be included in the new software application. But, some other times it is necessary this second step to analyze every activity in order to define the scope of the new software application.

Regarding the example, the farmer fertilizes in two different ways. One technique consists in using a spraying back pack, and another one consists in using an irrigation pipe. The procedure to fertilize using the backpack is manual, so it will be outside the

new software application. Nevertheless, the irrigation pipe can be adapted in order to automatize the fertilization. Thus, the behavioral impact “The farmer fertilizes spraying” is categorized as (iii) activities to keep performing manually. While the behavioral impact “The farmer fertilizes watering” is categorized as (i) functionality of the intended software application (Fig. 7). The procedure is summarized in Fig. 8.

Subject: farmer
Notion: responsible to grow the fruits
Behavioral responses
The farmer fertilizes spraying. (activities to keep performing manually)
The farmer fertilizes watering. (functionality of the intended software application)

Fig. 7. Categorization of behavioral responses of subject farmer

for each subject s categorized as user of the intended software application
for each behavioral response b that belong to s
categorize b as
(i) functionality of the intended software application
(ii) functionality of another software application
(iii) activities to keep performing manually

Fig. 8. Procedure for behavioral responses of the subject categorization

Classification of behavioral responses of verbs

The behavioral responses of the verbs describe how the activity represented by the verb should be carried out. The behavioral responses are a kind of work breakdown of the verb that describes. Although the step 2 of the approach classifies the behavioral responses of the subjects according to their inclusion in the intended software application, it could happen that some activities will not be completely automatized. Thus, each behavioral response of the subjects (categorized as “functionality of the intended software application”) that in turn are described as verbs should be analyzed. In this step, the behavioral responses of these verbs should be categorized as (i) functionality of the intended software application, (ii) functionality of another software application, and (iii) activities to keep performing manually.

The process of fertilizing through watering with the irrigation pipe is composed of several steps. First, some calculus of the mixture of the minerals to use to fertilize should be done. Then, the mixture should be prepared. After that, the mixture should be poured into the irrigation pipe. Finally, it should be decided which sectors of the layout of the field should be fertilized. Thus, Fig. 9 summarizes the categorization of every behavioral response, and the procedure is summarized in Fig. 10.

Verb: fertilize watering

Notion: activity that pursue the aim of adding nutrient to the plant.

Behavioral responses:

The farmer plans the mixture of minerals. (functionality of another software systems)

The farmer prepares the mixture of minerals. (activities to keep performing manually)

The farmer pours the mixture into the irrigation pipe. (activities to keep performing manually)

The farmer plans the layout to fertilize. (functionality of the intended software application)

The farmer activates the irrigation pipe. (functionality of the intended software application)

Fig. 9. Categorization of the verb "Fertilize watering"

for each subject s categorized as user of the intended software application

for each behavioral response b that belong to s categorized as functionality of the intended software application

for each behavioral response v of the verb that describes b

categorize v as

(i) functionality of the intended software application

(ii) functionality of another software application

(iii) activities to keep performing manually

Fig. 10. Procedure for behavioral responses of the verb categorization

Classification of objects

The behavioral responses should have the structure: subject + verb + object [3] where the object describes the element (material, resource, data) on which relies the action of the verb. Thus, if the verb (that is the behavioral response of the previous step) is categorized as "functionality of the intended software application", the object that receive the action, is probably an "object of the intended software application". Nevertheless, it could happen that the object is within the border of two different software applications. Thus, every object should be analyzed an categorized as (i) object within the boundaries of the intended software application, (ii) object within the boundaries of another software application, and (iii) object shared by several software applications.

The process of fertilizing through watering with the irrigation pipe should have access and control of the layout of the field. This is necessary to open and close the hatches to make the water (with the fertilizes) flows to the desired sector of the field. And of course, it should also have access to the pump to activate the irrigation pipe. Fig. 11 summarizes the categorization of every object, and Fig. 12 summarizes the procedure.

Verb: fertilize watering
Notion: activity that pursue the aim of adding nutrient to the plant.
Behavioral responses:
The farmer plans the mixture of minerals.
The farmer prepares the mixture of minerals.
The farmer pours the mixture into the irrigation pipe.
The farmer plan the layout to fertilize. (object within the boundaries of the intended software application)
The farmer activates the irrigation pipe. (object within the boundaries of the intended software application)

Fig. 11. Categorization of the objects belonging to the "Fertilize watering"

for each subject s categorized as user of the intended software application
for each behavioral response b that belong to s categorized as functionality of the intended software application
for each behavioral response v of the verb that describes b
for each object o of v
categorize o as
(i) object within the boundaries of the intended software application
(ii) object within the boundaries of another software application
(iii) object shared by several software applications.

Fig. 12. Procedure for objects categorization

This step could also be used as a revision phase, since if a verb is considered to be "functionality of the intended software application", and the object is considered "object within the boundaries of another software application", it could be analyzed why the verb (functionality) is within the boundaries while the object (data) is outside.

In the example, as a result of the process applied, it is stated that the farmer will be the user of the software application that will provide the functionality to control the hatches of the field to fertilize as well as switching on and off the pump.

4 Evaluation

The framework proposed was applied to an application to manage sanitary resources related to covid-19. The system manages doctors, rooms, beds and patients. The system also manages the evolution of a patient and provides alerts according to certain workflow to follow the evolution of the patient.

Participants were 25 students of a degree course divided in 11 groups. The objective of the course is to provide a realistic experience in software development. In particular, the course emphasizes requirements practices. It is important to mention that most of the students have experience in industry since in Argentina, students generally begin to work in industry in second year of their undergraduate studies.

Participants received a glossary LEL already prepared and they had to apply the proposed approach. One of the professors of the course is a Medical Doctor, and he played the role of the client providing the information about what should be included in the intended software application. Another professor of the course checked the categorization of the elements of the glossary LEL.

The evaluation was focused on the applicability of the approach. The System Usability Scale (SUS) [9] [10] was used to assess the usability and applicability of the approach. Although SUS is mainly used to assess usability of software systems, it was probe to be effective to assess products and processes [6]. The System Usability Scale (SUS) consists of a 10-item questionnaire; every question must be answered in a five-option scale, ranging from "1" ("Strongly Disagree") to "5" ("Strongly Agree"). Although there are 10 questions, they are related by pairs, asking the same question but in a complementary point of view in order to obtain a result of high confidence.

The calculation of the SUS score is performed in the following way. First, items 1, 3, 5, 7, and 9 are scored considering the value ranked minus 1. Then, items 2, 4, 6, 8 and 10, are scored considering 5 minus the value ranked. After that, every participant's scores are summed up and then multiplied by 2.5 to obtain a new value ranging from 0 to 100. Finally, the average is calculated. The approach can have one of the following results: "Non acceptable" 0-64, "Acceptable" 65-84, and "Excellent" 85-100 [22]. The score obtained was 71,17. Thus, the approach can be considered as "acceptable".

5 Related work

Lee et al. [18] use domain knowledge information (and requirements documents) in natural language to create a richer knowledge base used to produce artifacts specific of the application domain. Their approach produces UML diagrams and source code. They propose a broader approach that ours. Nevertheless, they do not describe how to define the boundaries of the application to produce artifacts so specific like source code. Voelter et al. [28] use domain specific languages in product line engineering as a middle ground between feature modeling and programming. This approach pursues the same concern of our approach to define the boundaries of the software application even in a specific domain as software product lines.

Wang et al. [29] are concerned about reducing the gap between natural language requirements and Architecture Analysis and Design Language models. Although their objective is different from the one of our approach, in some way both concerns are related. Their approach is mainly based on data dictionaries and glossaries, which are the elements used by us. Borelli et al. [8] also propose an approach for architectural design. In particular, they work with IoT, and they propose a tool to analyze a Domain Specific Language (DSL) in order to obtain a new language called: A Buildout IoT Application Language (BIoTA).

Mukhtar et al. [24] consider the importance of identifying the vocabulary of a software application. They use general dictionaries to identify compound words that contain some atomic words. Then, experts analyze the terms and their definition to

finally consider those terms. This approach could be used in a previous stage of our approach, where relevant concept analysis should be identified.

Bai et al. [4] [5] propose a strategy to build domain specific language models from general domain data. They search for similar topics in documents of related domains while we work with a specific domain to provide a description of a specific application. Moreover, Bai et al. work with natural language documents while we work with a semi structured natural language model as Language Extended Lexicon. Demsky et al. [13] developed an application called Bristlecone that relate high-level specification with the low level application's conceptual operations, which can be considered as a kind of vocabulary.

Doorn et al [15] propose a strategy to understand the future universe of discourse through the use of Scenarios [16]. Their strategy relies on constructing future scenarios (and requirements) in order to obtain the future universe of discourse captured by LEL. This strategy is more complex that our approach since they try to define the future universe of course, while our approach only limit the application domain universe of discourse. We plan to develop an extension of our proposed approach to obtain the language of the future application system, but we believe that defining the limits of the universe of discourse beforehand is more important. Haj et al. [17] propose an approach based on natural language processing that to obtain Semantic of Business Vocabulary and Rules (SBVR). Lie et al. [21] proposed an approach to develop the software application language with a semantic support. They use User Stories as input and analyze them in order to obtain relevant concept which are linked with wordnet in order to have a semantic support. It is an interesting idea, but the wordnet dictionary is a general way, so it could be hard to relate the concept with the correct definition.

Dilshener et al. [14] performed an analysis about the relationship between the concept that appear in the source code and the use of those concepts in a more abstract and conceptual artefact. This works is a kind of verification that both vocabularies should be synchronized in order to make easy the software development process. Amatriain et al. [1] performed a similar analysis showing that they were able to develop a framework with minimal overhead thanks to the use of vocabularies, in particular a domain specific language (DSL). Nascimento et al. [25] performed an analysis of the vocabulary used in the source code in order to assess the level of understanding of the students that has written the source code. It is interesting to emphasize how the use of language can reduce the effort and it also can be used to assess the understanding of the domain.

6 Conclusions

This paper presented an approach to define the language of the software application from the language of the application domain. The process of acquiring and identifying the relevant requirements and knowledge to develop a new software application can be overwhelming in some situation of a complex domain where a lot of stakeholders are involved. Thus, this paper proposes an approach that uses the LEL glossary to capture the vocabulary of the application domain and obtain the LEL glossary of the

software application in a straightforward way with different steps of analyzing the concepts defined by the language and categorizing them according to their situation in the boundaries of the software application. This process should be considered as a framework that filter the elements of the application domain and define the scope of the software application at the same time. The proposed approach only consider the reduction of the initial LEL glossary, but the analyst who applies the proposed approach will acquire more information during the process and he could add this new knowledge to the LEL glossary obtained as a result. This specific improvement of the proposed approach is a future work. We are currently analyzing scenarios where the language changes or is enriched in order to propose a new approach that also consider this situations.

Acknowledgement

This paper is partially supported by funding provided by the STIC AmSud program, Project 22STIC-01.

References

1. Amatriain, X., Arumi, P.: “Frameworks Generate Domain-Specific Languages: A Case Study in the Multimedia Domain,” in IEEE Transactions on Software Engineering, vol. 37, no. 4, pp. 544-558, July-Aug., doi: 10.1109/TSE.2010.48 (2011).
2. Antonelli, L., Rossi G., Leite, J.C.S.P., Oliveros, A: “Deriving requirements specifications from the application domain language captured by Language Extended Lexicon”, Workshop in Requirements Engineering (WER), Buenos Aires, Argentina, April 24-27 (2012).
3. Antonelli, L., Leite, J.C.S.P., Oliveros, A., Rossi G.: “Specification Cases: a lightweight approach based on Natural Language”, Workshop in Requirements Engineering (WER), Brasilia, Brazil, August 23 – 27 (2021).
4. Bai, S., Zhang, N., Li, H.: “Semi-supervised Learning of Domain-Specific Language Models from General Domain Data”, in Proceeding of the International Conference on Asian Language Processing (IALP '09), ISBN 978-0-7695-3904-1, pp 273-279 (2009)
5. Bai, S., Huang, C.L.,Tan, Y.K., Ma, B.: Language models learning for domain-specific natural language user interaction, in proceeding of the IEEE International Conference on Robotics and Biomimetics (ROBIO), ISBN 978-1-4244-4774-9, pp. 2480-2485 (2009).
6. Bangor, A., Kortum, P. T., Miller, J. T. "An empirical evaluation of the system usability scale." Intl. Journal of Human–Computer Interaction 24.6, pp. 574-594 (2008).
7. Boehm, B.W.: Software Engineering, Computer society Press, IEEE, 1997.
8. Borelli, F. F., Biondi, G. O., Kamienski, C. A.: "BIoTA: A Buildout IoT Application Language," in IEEE Access, vol. 8, pp. 126443-126459, doi: 10.1109/ACCESS.2020.3003694 (2020).
9. Brooke, J. “SUS-A quick and dirty usability scale. Usability evaluation in industry”, 189(194), 4-7, (1996).
10. Brooke, J. "SUS: a retrospective." Journal of usability studies 8.2, pp.29-40, (2013).
11. Brooks, F., The Mythical Man-Month: Essays on Software Engineering, Addison-Wesley Professional, 2 edition 1995.

12. Cysneiros, L.M., Leite, J.C.S.P.: Using the Language Extended Lexicon to Support Non-Functional Requirements Elicitation, in proceedings of the Workshops de Engenharia de Requisitos, Wer'01, Buenos Aires, Argentina, 2001.
13. Demsky B., Sundaramurthy, S.: "Bristlecone: Language Support for Robust Software Applications," in IEEE Transactions on Software Engineering, vol. 37, no. 1, pp. 4-23, Jan.-Feb., doi: 10.1109/TSE.2010.27 (2011)
14. Dilshener, T. Wermelinger, M.: "Relating developers' concepts and artefact vocabulary in a financial software module," 2011 27th IEEE International Conference on Software Maintenance (ICSM), pp. 412-417, doi: 10.1109/ICSM.2011.6080808 (2011).
15. Doorn, J.H., Hadad, G.D.S., Kaplan, G.N.: "Comprendiendo el Universo de Discurso Futuro", WER'02 - Workshop en Ingeniería de Requisitos, Valencia, España, pp.117-131, Noviembre 2002.
16. Hadad, G., Kaplan, G., Oliveros, A., Leite, J.C.S.P.: Construcción de Escenarios a partir del Léxico Extendido del Lenguaje, in Proceedings SoST, 26JAIIO, Sociedad Argentina de Informática y Comunicaciones, Buenos Aires (1997)
17. Haj, A., Jarrar, A., Balouki, Y. Gadir, T.:"The Semantic of Business Vocabulary and Business Rules: An Automatic Generation From Textual Statements," in IEEE Access, vol. 9, pp. 56506-56522, doi: 10.1109/ACCESS.2021.3071623 (2021).
18. Lee, B. S., Bryant, B.R.: Automation of software system development using natural language processing and two level grammar, In Proceeding of the Workshop Radical Innovations of Software and Systems Engineering in the Future, Monterey, 244-257 (2002)
19. Leite, J.C.S.P., Franco, A.P.M.: A Strategy for Conceptual Model Acquisition, in Proceedings of the First IEEE International Symposium on Requirements Engineering, San Diego, California, IEEE Computer Society Press, pp 243-246 (1993)
20. Lim, S. L., Finkelstein, A.: "StakeRare: Using Social Networks and Collaborative Filtering for Large-Scale Requirements Elicitation", IEEE transactions on software engineering, Volume 38, Issue 3, May-Jun 2012, doi: 10.1109/TSE.2011.36, pp 707-735, 2012
21. Liu, Y., Lin, J., Cleland-Huang, J., Vierhauser, M., Guo, J., Lohar, S.: "SENET: A Semantic Web for Supporting Automation of Software Engineering Tasks," 2020 IEEE Seventh International Workshop on Artificial Intelligence for Requirements Engineering (AIRE), pp. 23-32, doi: 10.1109/AIRE51212.2020.00011 (2020).
22. McLellan, S., Muddimer, A., Peres, S. C. "The effect of experience on System Usability Scale ratings." Journal of usability studies 7.2, pp. 56-67 (2012).
23. Meservy, T. O., Zhang, C., Lee, E. T. and Dhaliwal, J.: "The Business Rules Approach and Its Effect on Software Testing," in IEEE Software, vol. 29, no. 4, doi: 10.1109/MS.2011.120, pp. 60-66, July-Aug, 2012.
24. Mukhtar, T., Afzal, H. Majeed, A.: "Vocabulary of Quranic Concepts: A semi-automatically created terminology of Holy Quran," 2012 15th International Multitopic Conference (INMIC), pp. 43-46, doi: 10.1109/INMIC.2012.6511467 (2012).
25. Nascimento, M., Araújo, E., Serey D., Figueiredo, J.: "The Role of Source Code Vocabulary in Programming Teaching and Learning," 2020 IEEE Frontiers in Education Conference (FIE), pp. 1-8, doi: 10.1109/FIE44824.2020.9274137 (2020).
26. Potts, C.: "Using schematic scenarios to understand user needs," in Proceedings of the 1st conference on Designing interactive systems: processes, practices, methods, & techniques, (1995).
27. Rose, S., Nagy, G.: Formulation: Document examples with Given/When/Then, Independently published, 979-8723395015, (2021).
28. Voelter, M., Visser, E.: Product Line Engineering Using Domain-Specific Languages in proceeding of the 15th International Software Product Line Conference (SPLC) ISBN 978-1-4577-1029-2, pp 70-79 (2011).

29. Wang, F. et al.: "An Approach to Generate the Traceability Between Restricted Natural Language Requirements and AADL Models," in IEEE Transactions on Reliability, vol. 69, no. 1, pp. 154-173, March, doi: 10.1109/TR.2019.2936072 (2020).

Students' Accommodation Allocation: A Multicriteria Decision Support System

Rôlin Gabriel Rasoanaivo [1][0000-0002-2496-2672] and Pascale Zarate [1][0000-0002-5188-1616]

[1] IRIT, Université de Toulouse, CNRS, INP, UT3, UT1, UT2, Toulouse, France
rolin-gabriel.rasoanaivo@ut-capitole.fr,
pascale.zarate@ut-capitole.fr

Abstract. The social life of students at university has an impact on their educational success. The allocation of accommodation is part of this aspect. This article presents our proposal to improve students' allocation accommodation. We aim to support university administrative departments for the selection of students for housing. Therefore, we propose a decision support system based on multi-criteria decision support methods. To calculate the weights of the criteria, we use the Analytic Hierarchy Process (AHP) method. Then, to rank the students, Analytic Hierarchy Process, Weighted Sum Method (WSM) and Preference Ranking Organisation Method for Enrichment Evaluation methods (PROMETHEE) are used. The aim is to find the most adequate method to rank the students. The result is achieved because the AHP can calculate the weight of criteria and the AHP, WSM and PROMETHEE are able to rank the students.

Keywords. Student housing allocation, decision support system, AHP, WSM, PROMETHEE.

1 Introduction

To successful manage universities, the institution must demonstrate its priorities in providing students housing because sharing accommodation and participating in campus life are part of the social and intellectual benefits of students [1]. According to researches carried out by Owolabi, on-campus students are more successful in their studies than off-campus students [2]. Therefore, following our research, several authors have been interested in the subject of student housing. Several studies collected the opinion of the students about their accommodation with the aim of bringing a new era of living on campus [3–10]. Then other researches made the proposal of work tools to help the person in charge of housing within the universities, as well as the students to facilitate the search for housing [11–14]. These works justify the importance of considering the management of student accommodation within university life.

However, each university has its own procedures for allocating student accommodation. The most common is the consideration of criteria which vary in number according to each institution. Thus, we have identified thirteen criteria for the allocation of accommodation used by universities, namely: dependent children of parents, students with scholarships, distance from the place of study, admission to the examination, educational enrolment, age, nationality, level of study, date of application for accommodation, physical capacity of the student, orphan of a parent, parent working at the university, and year of undergraduate [15–21].

Our research has focused on Decision Support System (DSS) for students' accommodation allocation [22]. Let us note that the first DSS appeared in the 1960s [23]. Since then, several authors have investigated the application of these systems such as Eom et al. from 1971 to 2001 [24–27] and Papathanasiou et al. from 1989 to 2019 [28]. Authors like Carlsson and Walden [29] have noted that Decision Support Systems are currently innovative and among the technological challenges. Keenan [30] has shown that they are spread over several areas including life sciences and biomedicine, physical sciences, and social sciences. Nevertheless, we did not find any DSS for student's accommodation allocation.

Thus, our aim is to propose a decision support system based on AHP, WSM and PROMETHEE to facilitate the selection of student during the housing allocation process.

2 Issues

Regarding the criteria for housing allocation, each student is considered as a specific case. The idea is to select those students who meet them. However, there is much more demands than offers, but sometimes there are many more students who meet the criteria than there is housing to be allocated [17]. So, the question which remains is: what housing applications should be admitted or rejected? And which of these students will be housing allocated?

To do this, a method needs to be designed to fairly distribute this accommodation.

3 Methodology

Multi-criteria analysis methods are used to formulate real problems, according to three basics formulations: the choice problem, noted P_α, the sorting or allocation problem noted P_β and the ranking problem noted P_γ [31]. The allocation of accommodation to students belongs to the ranking problem (P_γ). This situation led us to choose the following three methods: AHP, WSM and PROMETHEE to solve this problem. These three methods are selected because of their popularity and their usefulness [31–33].

Analytic Hierarchy Process (AHP)

AHP was developed by Saaty in 1970 [34]. According to its founder [35], the method is based on three concepts: hierarchical structure, priority structure and logical consistency. To find the logical consistency, the following calculations have to be made:

Medium consistency: $\lambda_{max} = a_{ij}\frac{w_j}{w_I}$ (1)

Consistency index: $CI = \frac{\lambda_{max}-n}{n-1}$ (2)

Consistency ratio: $CR = \frac{CI}{RI}$ (3)

Finally, when comparing pairs, the consistency ratio (CR) must be within 0.1. Otherwise, the results could be inconsistent.

Weight Sum Method (WSM)

The weighted sum method combines all criteria into one scalar composite objective function using the weighted sum [36]. The steps to follow are normalisation of all alternatives, normalisation of the weights whose sum must be equal to 1 and implementation of the weighted sum. The result is obtained by [37]:

$R(a_{ij}) = \sum_{j=1}^{n} w_j a_{ij}$ (4)

Preference Ranking Organisation METHods for Enrichement Evaluation

Brans initiated the PROMETHEE method in 1982 [38]. It is a multi-criteria method for defining the relationships of outranking, indifference, and incomparability between alternatives. Two concepts are to be considered, namely the preference index and the outranking flows. The method of calculating these flows [39] is presented below. The result is obtained by comparing the outflows, inflows, and net flows of the alternatives.

Preference index: $\begin{cases}\pi(a_1,a_2) = \sum_{j=1}^{k} P_j(a_1,a_2)w_j \\ \pi(a_2,a_1) = \sum_{j=1}^{k} P_j(a_2,a_1)w_j\end{cases}$ (5)

Outflow: $\Phi^{+}(a) = \frac{1}{n-1}\sum_{x\in A}\pi(a,x)$(6)

Inflow: $\Phi^{-}(a) = \frac{1}{n-1}\sum_{x\in A}\pi(x,a)$ (7)

Net flow: $\Phi(a) = \Phi^{+}(a) - \Phi^{-}(a)$ (8)

Prototype

Decision Support System (DSS) is structured by three main components which are: the model management system, the user interface, and the knowledge base [40]. We propose a DSS prototype based on the architecture defined by Sprague [41]. The DSS will be composed by a Data Base, a Model Base, and a Human/Interface module.

Database

The developed DSS consists of allocating housing to students through their applications. We have implemented a relational database management system using three methods AHP, WSM and PROMETHEE. The figure 1 below shows us the main menu of the system administrator. While the figure 2 shows the possibility of choosing the method allowing to arrange the students.

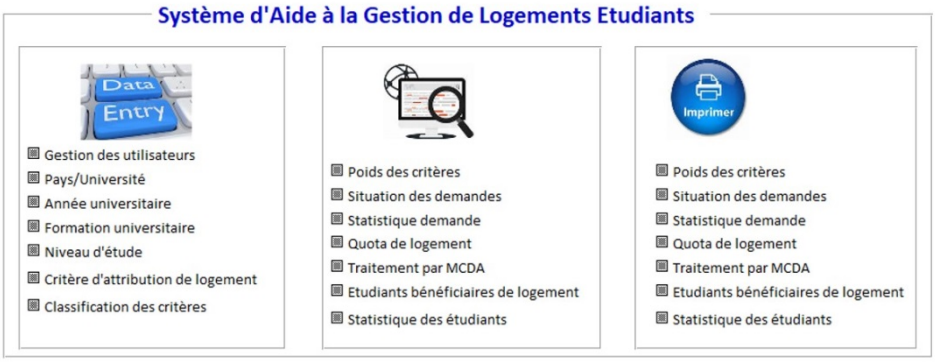


Fig. 13. System administrator main menu

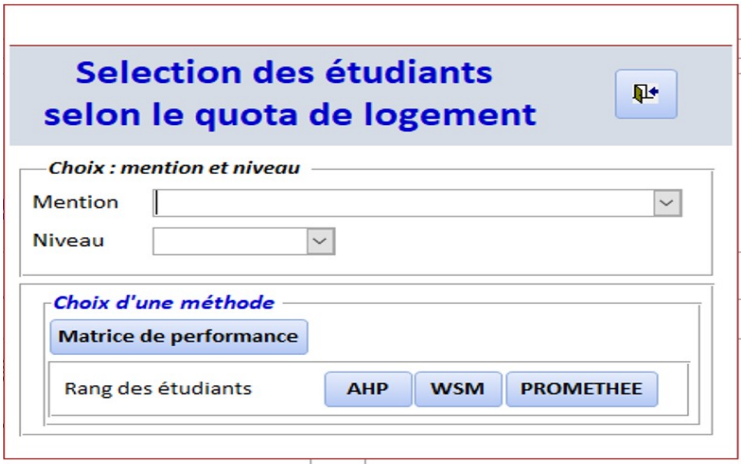


Fig. 14. Choice of methodologies

Process

The application concerns a university that has eleven criteria for evaluating students [17]. These criteria are grouped into two sets, of which the first six are basic criteria and the last five are social criteria. The following table 1 describes these evaluation criteria.

Table 2. Criteria for evaluating students at a university

Basic criteria	Admission requirements
Age	By level
Year of Baccalaureate	By academic year
Administrative registration	Enrolled
Examination result	Successful
Nationality	According to the case
Professional situation	Not employed

Social criteria	Value
Physical capacity (CP)	Normal = 5 ; Disability = 10
Orphan of parent (OP)	None = 0; Father or Mother = 5 ; Father and Mother = 10
Parent's place of work (LTP)	University = 5 ; Other = 0
Dependent child of parent (EC)	By number
Distance from home (DD)	By mileage

In relation to these two groups of criteria, the assessment procedure proceeds in two stages. Firstly, students must meet the basic criteria, otherwise their applications for housing will be rejected. Then, those who pass the basic criteria will move on to the second assessment where the social criteria are applied. Each social criterion has its own value to rank the students. This is where the multi-criteria decision support

method comes in. At this stage, the processing is done by specialty and by level (bachelor/master/PhD). The following figure 3 summarizes the procedure for evaluating students by criteria.

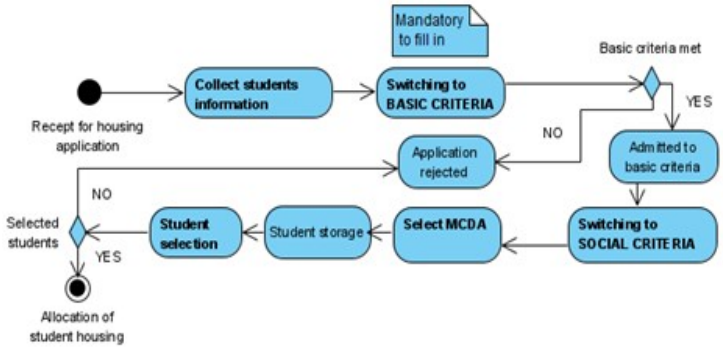


Fig. 15. Processing applications for student accommodation

Result

Assessment according to the basic criteria

This concerns the housing applications of students in the first year of the Computer Science and Law degrees. The result is shown in figure 4 below.

Récapitulatif des demandes par Mention et Niveau

Mention	Niveau	Reçue	Acceptée	Rejetée
Droit	L1	101	78	23
Mathématiques, Informatique et Applications	L1	35	26	9

Fig. 16. Result of the processing applications for student accommodation

Assessment according to social criteria

This second phase will deal with the 26 and 78 students meeting the basic criteria. We show for each of the three methods chosen, the result for the 26 students of the Computer Science major. To begin with, the initial judgement matrix and the judgement matrix normalized to scale 10.

Code ét	CP	DD	EC	LTP	OP
L1MIA16	5	100	4	0	5
L1MIA05	5	102	2	0	5
L1MIA06	5	100	3	0	0
L1MIA07	5	100	5	0	5
L1MIA08	5	100	4	0	5
L1MIA11	5	100	2	0	0
L1MIA12	5	100	1	0	0
L1MIA13	5	923	1	0	10
L1MIA15	5	100	3	0	0
L1MIA18	5	100	2	0	0
L1MIA21	5	100	4	0	0
L1MIA22	5	100	1	0	5
L1MIA23	5	350	2	0	10
L1MIA24	5	100	6	0	0
L1MIA25	5	100	5	0	0
L1MIA26	5	100	2	0	0
L1MIA27	5	102	1	0	0
L1MIA28	5	100	4	0	5
L1MIA29	5	100	5	0	5
L1MIA30	5	399	1	0	0
L1MIA31	5	100	3	0	10
L1MIA32	5	923	4	5	5
L1MIA34	5	399	2	0	10
L1MIA35	5	100	2	0	0
L1MIA02	5	100	5	0	5
L1MIA04	5	100	6	0	0

Fig. 17. Initial judgement matrix

ID_ETUDIAN	CP	DD	EC	LTP	OP
L1MIA16	5	0,68	5,71	0	5
L1MIA05	5	0,7	2,86	0	5
L1MIA06	5	0,68	4,29	0	0
L1MIA07	5	0,68	7,14	0	5
L1MIA08	5	0,68	5,71	0	5
L1MIA11	5	0,68	2,86	0	0
L1MIA12	5	0,68	1,43	0	0
L1MIA13	5	6,3	1,43	0	10
L1MIA15	5	0,68	4,29	0	0
L1MIA18	5	0,68	2,86	0	0
L1MIA21	5	0,68	5,71	0	0
L1MIA22	5	0,68	1,43	0	5
L1MIA23	5	2,39	2,86	0	10
L1MIA24	5	0,68	8,57	0	0
L1MIA25	5	0,68	7,14	0	0
L1MIA26	5	0,68	2,86	0	0
L1MIA27	5	0,7	1,43	0	0
L1MIA28	5	0,68	5,71	0	5
L1MIA29	5	0,68	7,14	0	5
L1MIA30	5	2,72	1,43	0	0
L1MIA31	5	0,68	4,29	0	10
L1MIA32	5	6,3	5,71	10	5
L1MIA34	5	2,72	2,86	0	10
L1MIA35	5	0,68	2,86	0	0
L1MIA02	5	0,68	7,14	0	5
L1MIA04	5	0,68	8,57	0	0

Fig. 18. Normalised judgement matrix

Result according to AHP

The application of the comparison scale according to Saaty [40], allowed us to obtain the criteria judgement matrix and the result of the priorities presented in figure 7. Thus, the values of the elements of logical consistency are given below.

Comparaison	CP	DD	EC	LTP	OP
CP	1	3	4	4	3
DD	0,33	1	2	2	1
EC	0,25	0,5	1	1	0,5
LTP	0,25	0,5	1	1	0,5
OP	0,33	1	2	2	1

Critères	Poids
CP	0,45
EC	0,1
LTP	0,1
DD	0,18
OP	0,18

Cohérence logique

λmax	5,0244
Indice de cohérence	0,0061
Ratio de cohérence	0,0054

Fig. 19. Priority of criteria

Subsequently, establish the student judgement matrices for each criterion from the normalized judgement matrix. After various calculations, the students' priorities for each criterion are shown in Figure 8 and the ranking of students in Figure 9 below.

Weight	L1MIA05	L1MIA06	L1MIA07	L1MIA08	L1MIA11	L1MIA12	L1MIA13	L1MIA15	L1MIA16	L1MIA18	L1MIA21	L1MIA22	L1MIA23	L1MIA24	L1MIA25	L1MIA26	L1MIA27	L1MIA28	L1MIA29	L1MIA30	L1MIA31	L1MIA32	L1MIA34	L1MIA35	L1MIA02	L1MIA04
CP	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04
DD	0,04	0,02	0,02	0,02	0,02	0,02	0,17	0,02	0,02	0,02	0,02	0,02	0,06	0,02	0,02	0,02	0,04	0,02	0,02	0,06	0,02	0,17	0,06	0,02	0,02	0,02
EC	0,02	0,03	0,07	0,05	0,02	0,01	0,01	0,03	0,05	0,02	0,05	0,01	0,02	0,10	0,07	0,02	0,01	0,05	0,07	0,01	0,03	0,05	0,02	0,02	0,07	0,10
LTP	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,17	0,03	0,03	0,03	0,03
OP	0,05	0,01	0,05	0,05	0,01	0,01	0,11	0,01	0,05	0,01	0,01	0,05	0,11	0,01	0,01	0,01	0,01	0,05	0,05	0,01	0,11	0,05	0,11	0,01	0,05	0,01

Fig. 20. priority of student

Student	L1MIA03	L1MIA13	L1MIA04	L1MIA03	L1MIA01	L1MIA07	L1MIA29	L1MIA02	L1MIA05	L1MIA08	L1MIA16	L1MIA22	L1MIA24	L1MIA04	L1MIA30	L1MIA22	L1MIA25	L1MIA21	L1MIA28	L1MIA20	L1MIA26	L1MIA01	L1MIA18	L1MIA35	L1MIA23	L1MIA12
Note	0,076	0,072	0,054	0,053	0,047	0,039	0,039	0,039	0,037	0,037	0,037	0,037	0,036	0,036	0,035	0,033	0,033	0,031	0,030	0,029	0,029	0,028	0,028	0,028	0,028	0,027
Rank	1	2	3	4	5	6	6	6	9	10	10	10	13	13	15	16	17	18	19	20	20	22	22	22	22	26

Fig. 21. AHP student ranking

Result according to WSM

The weights of the criteria considered are those calculated according to AHP. Figure 10 shows the application of the weighted sum to the normalized judgement matrix and Figure 11 gives the final result.

Student	L1MIA05	L1MIA06	L1MIA07	L1MIA08	L1MIA11	L1MIA12	L1MIA13	L1MIA15	L1MIA16	L1MIA18	L1MIA21	L1MIA22	L1MIA23	L1MIA24	L1MIA25	L1MIA26	L1MIA27	L1MIA28	L1MIA29	L1MIA30	L1MIA31	L1MIA32	L1MIA34	L1MIA35	L1MIA02	L1MIA04
CP	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262	2,262
DD	0,124	0,121	0,121	0,121	0,121	0,121	1,118	0,121	0,121	0,121	0,121	0,121	0,424	0,121	0,121	0,121	0,124	0,121	0,121	0,483	0,121	1,118	0,483	0,121	0,121	0,121
EC	0,276	0,414	0,688	0,551	0,276	0,138	0,138	0,414	0,551	0,276	0,551	0,138	0,276	0,826	0,688	0,276	0,138	0,551	0,688	0,138	0,414	0,551	0,276	0,276	0,688	0,826
LTP	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,964	0,000	0,000	0,000	0,000
OP	0,887	0,000	0,887	0,887	0,000	0,000	1,774	0,000	0,887	0,000	0,000	0,887	1,774	0,000	0,000	0,000	0,000	0,887	0,887	0,000	1,774	0,887	1,774	0,000	0,887	0,000

Fig. 22. implementing the weighted sum

Student	L1MIA03	L1MIA13	L1MIA04	L1MIA03	L1MIA01	L1MIA07	L1MIA29	L1MIA02	L1MIA05	L1MIA08	L1MIA16	L1MIA22	L1MIA24	L1MIA04	L1MIA30	L1MIA22	L1MIA25	L1MIA21	L1MIA28	L1MIA20	L1MIA26	L1MIA01	L1MIA18	L1MIA35	L1MIA23	L1MIA12
Note	5,781	5,292	4,794	4,736	4,570	3,958	3,958	3,958	3,820	3,820	3,820	3,549	3,407	3,208	3,208	3,071	2,933	2,882	2,796	2,796	2,658	2,658	2,658	2,658	2,524	2,520
Rank	1	2	3	4	5	6	6	6	9	9	9	12	13	14	14	16	17	18	19	19	21	21	21	21	25	26

Fig. 23. WSM student ranking

Result according to PROMETHEE

For the preferences: all the criteria are to be maximised, the weights applied are always those calculated according to AHP and the preference function for all the criteria are of the same Usual type. After setting the preferences and considering the assessments, the result will be given in figure 12.

Flux	L1MIA0332	L1MIA0334	L1MIA0133	L1MIA0233	L1MIA0005	L1MIA0331	L1MIA0002	L1MIA0007	L1MIA0229	L1MIA0008	L1MIA0116	L1MIA0228	L1MIA0330	L1MIA0004	L1MIA0224	L1MIA0222	L1MIA0225	L1MIA0227	L1MIA0221	L1MIA0066	L1MIA0015	L1MIA0111	L1MIA0118	L1MIA0226	L1MIA0235	L1MIA0112
Phi	0,370	0,260	0,241	0,238	0,124	0,107	0,082	0,082	0,082	0,046	0,046	0,046	0,046	-0,037	-0,052	-0,052	-0,073	-0,075	-0,080	-0,111	-0,143	-0,143	-0,182	-0,182	-0,182	-0,182	-0,230
Phi+	0,422	0,334	0,328	0,326	0,248	0,204	0,172	0,172	0,172	0,152	0,152	0,152	0,157	0,095	0,095	0,093	0,079	0,135	0,059	0,048	0,048	0,020	0,020	0,020	0,020	0,020	0,000
Phi-	0,052	0,074	0,087	0,088	0,124	0,097	0,090	0,090	0,090	0,106	0,106	0,106	0,194	0,147	0,147	0,166	0,155	0,215	0,170	0,190	0,190	0,202	0,202	0,202	0,202	0,202	0,230
Rank	1	2	3	4	5	6	7	7	7	10	10	10	13	14	14	16	17	18	19	20	20	22	22	22	22	22	26

Fig. 24. PROMETHEE student ranking

Discussion

In 2018, Saare et *al.* were designing a mobile system for managing and mitigating the accommodation problems at the Universiti Utara Malaysia [12]. In 2019, Podunavac et *al.* proposed a web portal managing registration for student accommodation in a dormitory [13]. In 2020, Magambo et *al.* have created an online portal for locating Students' private rental Accommodation in Tanzania [14].

For our part, we developed a DSS focusing on student housing allocation using three multi-criteria decision support methods (AHP, WSM and PROMETHEE) to better compare the results obtained. And to make a choice between the method used by the university, we illustrate in the following figure 13 the comparison of the results obtained and in terms of number percentage in the table 2 the similarity of student ranks.

		L1MIA0332	L1MIA0334	L1MIA0133	L1MIA0233	L1MIA0005	L1MIA0331	L1MIA0002	L1MIA0007	L1MIA0229	L1MIA0008	L1MIA0116	L1MIA0228	L1MIA0330	L1MIA0004	L1MIA0224	L1MIA0222	L1MIA0225	L1MIA0227	L1MIA0221	L1MIA0066	L1MIA0015	L1MIA0111	L1MIA0118	L1MIA0226	L1MIA0235	L1MIA0112
AHP		1	3	2	4	9	5	6	6	6	10	10	10	15	13	13	16	17	19	18	20	20	22	22	22	22	26
WSM	Rang	1	3	2	4	12	5	6	6	6	9	9	9	18	14	14	13	16	25	17	19	19	21	21	21	21	26
PROMETHEE		1	2	3	4	5	6	7	7	7	10	10	10	13	14	14	16	17	18	19	20	20	22	22	22	22	26

Fig. 25. Comparison of results

From the comparison of these three results, we can draw the following three situations:

7.69% of students have the same rank on the three methods: two students;

15.38% have different ranks on the three methods: four students;

76.92% have the same rank on two methods: the remaining twenty.

To make a choice between the method used by the university, we illustrate the two-by-two comparison of the results obtained.

Table 3. Similarity of student ranks on two methods

	WSM		PROMETHEE	
	Number	%	Number	%
AHP	8	30.77	13	50
WSM			4	15.38

4 Conclusion and Perspective

To decide is certainly to take risks but it is to be in reaction to a strategic choice to be made. The same applies to the person in charge of university work when receiving requests for student accommodation, a decision must be made whether to accept or reject an application. But, in any case, the question must always be asked: why was it accepted or rejected?

This is the reason for our analysis, which proposes the use of multi-criteria decision-making methods to classify students in relation to their situation, regarding the criteria for allocating housing. Three methods were studied: AHP, WSM and PROMETHEE.

The result showed us that each method was able to rank the students. And after comparing the results, we found that the rankings of the students are different for each method. First, the ranking carried out by AHP showed us that compared to WSM the similarity of ranks is 30.77%, and compared to PROMETHEE, it becomes 50%. Secondly, as for WSM, this similarity of ranks with PROMETHEE is 15.38%.

However, since the purpose of the decision analysis is to clarify the choice of the decision maker and not to replace the decision maker, the manager of university services has the choice of the method to use according to the analysis of the results that we carried out.

Nevertheless, we can report that a difficulty was encountered when using the AHP method. This being when the criteria or the alternatives are more numerous. For example, in the case of the 78 students in the Law stream and the other applications accepted in the other streams, the students are even more numerous. It will be difficult to handle such a large square matrix.

So, in our future research, we will use the AHP method only to calculate the criteria weights. On the other hand, WSM and PROMETHEE for ranking students. And in addition, we will look for another method that is easier to handle and combine with these two methods for storing students. Then, we will not stop with a simple application of the methods, but we will look for a trick allowing to recover all the ranks of the students carried out by the methods implemented in the system and to reorder the students from all their ranks.

Finally, our objective in this first analysis is then to integrate all these ranking methodologies into the developed prototype, offering end users the possibility of having several methodologies to use. And in the perspective, it will be to find the

possibility of using all the results of these methods to obtain a new ranking of the students.

References

1. Shattock, M. (2010). *EBOOK: Managing Successful Universities*. McGraw-Hill Education (UK).
2. Oluwaseyi, B. (2015). The Effects of Students' Housing on Academic Performance at the University of Ibadan in Nigerian, *6*(3), 15.
3. Simpeh, F., & Shakantu, W. (2020). An on-campus university student accommodation model. *Journal of Facilities Management*, *18*(3), 213–229. https://doi.org/10.1108/JFM-03-2020-0017
4. Mira, R. G., Uzzell, D. L., Real, J. E., & Romay, J. (Eds.). (2017). *Housing, Space and Quality of Life* (1st ed.). Routledge. https://doi.org/10.4324/9781351156363
5. Macintyre, C. (2003). New models of student housing and their impact on local communities. *Journal of Higher Education Policy and Management*, *25*(2), 109–118. https://doi.org/10.1080/1360080032000122598
6. Property, J. (2021, December 3). Opportunities In Student Accommodation. *Just Property*. Text, Just Property. Retrieved April 26, 2022, from https://just.property/blog/opportunities-in-student-accommodation/
7. Worsley, J. D., Harrison, P., & Corcoran, R. (2021). The role of accommodation environments in student mental health and wellbeing. *BMC Public Health*, *21*(1), 573. https://doi.org/10.1186/s12889-021-10602-5
8. Abbey∗, B. (1994). Student housing: A retrospect and some prospects. *Journal of Tertiary Education Administration*, *16*(2), 195–204. https://doi.org/10.1080/1036970940160205
9. Parameswaran, A., & Bowers, J. (2014). Student residences: from housing to education. *Journal of Further and Higher Education*, *38*(1), 57–74. https://doi.org/10.1080/0309877X.2012.699515
10. Simpeh, F., & Akinlolu, M. (2021). A Scientometric Review of Student Housing Research Trends. *IOP Conference Series: Earth and Environmental Science*, *654*(1), 012015. https://doi.org/10.1088/1755-1315/654/1/012015
11. Muhsen, Z. A. (2011). *Developing an Online Student Accommodation Registration in UUM* (masters). Universiti Utara Malaysia. Retrieved from https://etd.uum.edu.my/2760/
12. Saare, M. A., Ta'a, A. B., Lashari, S. A., & Sari, S. A. (2018). Mobile System for Managing and Mitigating the Accommodation Problems. *Journal of Physics: Conference Series*, *1019*, 012045. https://doi.org/10.1088/1742-6596/1019/1/012045
13. Podunavac, I., Crnjac Milić, D., & Nenadić, K. (2019). Proposal for a web portal managing registration for student accommodation in a dormitory. *Tehnički glasnik*, *13*(1), 75–80. https://doi.org/10.31803/tg-20180629110602
14. Magambo, L., Dida, M. A., & Kaijage, S. F. (2020). Towards an Online Portal for Locating Students' Private Rental Accommodation in Tanzania. *Journal of Information Systems Engineering and Management*, *5*(3), em0118. https://doi.org/10.29333/jisem/8388
15. Hammad, D., Musa, J., Rishi, A., & Ayuba, I. (2013). Criteria for the Selection of Students Accommodation Model in Nigeria Tertiary Institutions Using Analytic Hierarchy Process. *Academic Research International*, *4*, 550–556.
16. ALLOCATION-CRITERIA1920.pdf. (n.d.). Retrieved from https://www.ulster.ac.uk/__data/assets/pdf_file/0010/403759/ALLOCATION-CRITERIA1920.pdf

17. Rasoanaivo, R. G. (2020, January 10). La méthode PROMETHEE pour l'aide à la décision multicritère : application à la gestion d'attribution de logements universitaires (Cas de l'Université de Toamasina) (Mémoire de DEA). Université de Toamasina.
18. student-accommodation-allocation-policy-201920.pdf. (n.d.). Retrieved from https://intranet.royalholloway.ac.uk/students/assets/docs/pdf/student-accommodation-allocation-policy-201920.pdf
19. Allocation-Policy-version-26-2020-21_0.pdf. (n.d.). Retrieved from https://www.chi.ac.uk/app/uploads/2021/11/Allocation-Policy-version-26-2020-21_0.pdf
20. Hall Allocation Policy for Undergraduate Students (for Residential Year residences) | Student Housing and Residential Life. (n.d.). Retrieved March 25, 2022, from https://shrl.ust.hk/Hall-Allocation-Policy-for-Undergraduate-Students
21. Accommodation guarantee at ANU. (2020, August 17). *ANU*. The Australian National University. Retrieved March 25, 2022, from https://www.anu.edu.au/study/accommodation/accommodation-application-advice/accommodation-guarantee-at-anu
22. Rasoanaivo, R. G., & Zaraté, P. (2022). Decision Support System for student's room allocation, 8th International Conference on Decision Support System Technology – ICDSST 2022, University of Macedonia, Department of Business Administration Thessaloniki, Greece, 23-25/may/2022, https://icdsst2022gr.files.wordpress.com/2022/07/icdsst-2022-proceedings.pdf, 30–36.
23. Gorry, G., & Scott-Morton, M. (1971). A framework for management information systems,. *Sloan Management Review*, *13*.
24. Eom, H. B., & Lee, S. M. (1990). A Survey of Decision Support System Applications (1971–April 1988). *INFORMS Journal on Applied Analytics*, *20*(3), 65–79. https://doi.org/10.1287/inte.20.3.65
25. Eom, H. B., & Lee, S. M. (1990). Decision support systems applications research: A bibliography (1971–1988). *European Journal of Operational Research*, *46*(3), 333–342. https://doi.org/10.1016/0377-2217(90)90008-Y
26. Eom, S. B., Lee, S. M., Kim, E. B., & Somarajan, C. (1998). A survey of decision support system applications (1988–1994). *Journal of the Operational Research Society*, *49*(2), 109–120. https://doi.org/10.1057/palgrave.jors.2600507
27. Eom, S., & Kim, E. (2006). A survey of decision support system applications (1995–2001). *J. Oper. Res. Soc.* https://doi.org/10.1057/palgrave.jors.2602140
28. Papathanasiou, J., Zaraté, P., & Freire de Sousa, J. (Eds.). (2021). *EURO Working Group on DSS: A Tour of the DSS Developments Over the Last 30 Years*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-70377-6
29. Carlsson, C., & Walden, P. (2021). Decision Support Systems: Historical Innovations and Modern Technology Challenges. In J. Papathanasiou, P. Zaraté, & J. Freire de Sousa (Eds.), *EURO Working Group on DSS: A Tour of the DSS Developments Over the Last 30 Years* (pp. 1–14). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-70377-6_1
30. Keenan, P. B. (2021). Thirty Years of Decision Support: A Bibliometric View. In J. Papathanasiou, P. Zaraté, & J. Freire de Sousa (Eds.), *EURO Working Group on DSS: A Tour of the DSS Developments Over the Last 30 Years* (pp. 15–32). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-70377-6_2
31. Roy, B. (1996). Decision Problems and Processes. In B. Roy (Ed.), *Multicriteria Methodology for Decision Aiding* (pp. 3–6). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4757-2500-1_1

32. Papathanasiou, J., & Ploskas, N. (2018). PROMETHEE. In J. Papathanasiou & N. Ploskas (Eds.), *Multiple Criteria Decision Aid : Methods, Examples and Python Implementations* (pp. 57–89). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-91648-4_3
33. Solikhun, S. (2017). PERBANDINGAN METODE WEIGHTED PRODUCT DAN WEIGHTED SUM MODEL DALAM PEMILIHAN PERGURUAN SWASTA TERBAIK JURUSAN KOMPUTER. *KLIK - KUMPULAN JURNAL ILMU KOMPUTER*, *4*, 70. https://doi.org/10.20527/klik.v4i1.75
34. Saaty, R. W. (1987). The analytic hierarchy process—what it is and how it is used. *Mathematical Modelling*, *9*(3), 161–176. https://doi.org/10.1016/0270-0255(87)90473-8
35. Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234–281. https://doi.org/10.1016/0022-2496(77)90033-5
36. Cheng, Y.-P., & Chou, A. N.-G. (2011). Bi-level weights sum method for shock diagnosis. *Expert Systems with Applications*, *38*(4), 4497–4504. https://doi.org/10.1016/j.eswa.2010.09.123
37. Weighted Sum Method - Multi Criteria Decision Making. (2020, February 20). *GeeksforGeeks*. Retrieved from https://www.geeksforgeeks.org/weighted-sum-method-multi-criteria-decision-making/
38. Brans, J. P., Vincke, Ph., & Mareschal, B. (1986). How to select and how to rank projects: The Promethee method. *European Journal of Operational Research*, *24*(2), 228–238. https://doi.org/10.1016/0377-2217(86)90044-5
39. Brans, J.-P., & De Smet, Y. (2016). PROMETHEE Methods. In S. Greco, M. Ehrgott, & J. R. Figueira (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys* (pp. 187–219). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3094-4_6
40. Power, D. J. (2008). Decision Support Systems: A Historical Overview. In F. Burstein & C. W. Holsapple (Eds.), *Handbook on Decision Support Systems 1: Basic Themes* (pp. 121–140). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-48713-5_7
41. Sprague, R. H. (1980). A Framework for the Development of Decision Support Systems. *MIS Quarterly*, *4*(4), 1–26. https://doi.org/10.2307/248957

Generic Software for Benchmarking Formal Concept Analysis: Orange3 Integration

Nicolas Leutwyler[1,2,3], Mario Lezoche[1,2], Hervé Panetto[1,2], and Diego Torres[3,4]

[1] Université de Lorraine, Nancy 54000, France
[2] CRAN, Nancy 54000, France
[3] LIFIA, CICPBA-Facultad de Informática, UNLP, La Plata 1900, Argentina
[4] Dto. CyT, UNQ, Bernal, Argentina

Abstract. Thanks to the internet of things (IoT) and cyber physical systems (CPS), we face an incremental growth of the available data, either on the internet or in private databases. This resulted in data mining techniques becoming an essential piece in the information retrieval process. Moreover, trends like the industry 4.0 encourages its usage to support data driven decisions, for instance. Formal Concept Analysis (FCA) is one of the most used techniques in the unsupervised data mining field due to its inherent ability to find patterns between concepts. As a consequence, many applications need the use of fast algorithms to perform the calculations to retrieve either the lattice or the association rules related with the data at their disposal. Due to this, scientists often rely on manually crafted benchmarks to compare how certain algorithms perform under different circumstances. In this work, we propose the architecture of a software to generalize these benchmarks independently of the algorithms, to be integrated in the open source data analysis software Orange3.

Keywords. Formal Concept Analysis, benchmarking, metaprogramming, open source.

1 Introduction

Data mining techniques are widely used to support data driven decisions [19], to infer knowledge *automatically* in contexts such as software modelling, or artificial intelligence [25,24]. Furthermore, the internet of things and cyber physical systems could use some gains the data mining provide through semantic interoperability [3,20]. However, these fields need precise and fast processing of the information since they usually work in real time [22]. In this context, the effort towards finding fast data mining algorithms and measuring their performance is understandable [13,14].

In conventional agriculture, pesticides, antimicrobials and other pest control products are undesired, thus the need of alternative solutions. The European Green Deal is one of the most important actions in Europe to overcome the challenges of climate change and environmental degradation, sometimes caused by the usage of pesticides. In the state of the art, there are numerous descriptions of active plant-based products used as bio-pesticides. The Knomana (KNOwledge MANAgement on pesticide plants in Africa) project's goal is to gather data about these bio-pesticides and implement methods to support the exploration ofknowledge by the potential users (farmers, researchers, retailers, etc.). Considering the needs expressed by the domain experts, information retrieval is needed to obtain relevant insight on the matter. In addition, data clustering into similar groups is helpful when it comes to understanding key differences (or similarities) of objects in general. Formal Concept Analysis (FCA) appears as a suitable approach, due to its inherent qualities for structuring and classifying data through conceptual structures that provide a relevant support for data exploration.

Additionally, the trend driven by the Industry 4.0 [4,15] is to increase the usage of the available data in order to increase the performance and the efficiency of processes. In particular, regarding the Agriculture 4.0, impulsed alsoby the Green Deal in the EU, several works have been carried out using the data mining method called Formal Concept Analysis (FCA) [23,9], and also itsmulti-relational data mining [7] extension Relational Concept Analysis (RCA)[21,11]. There are several good results about the time complexity in the worst case from the main FCA algorithm, which is the one that calculates the set of *formal concepts* (see Section 2). Nevertheless, according to some previous experiments, it is known that some algorithms with worse time complexity than others perform better under certain circumstances [13,14].

Consequently, while developing new algorithms in this area, it is important to also perform a good benchmarking suit of tests to understand in which situations the algorithms strives. This is something that usually takes extra effort since it is for the most part a manual process. In the current literature, one way to ease the manual work required to perform these tests is approached by providing generic benchmarking tools for the particular application [17,26]. Particularly, we could not find any work in this area applied particularly to the algorithms for FCA and its extensions.

In this work, we introduce a software tool to benchmark, and another to use FCA in data pipeline. Both of them are thought to be added to the architecture of the open access data analysis software Orange3 [18]. It is worth mentioning that some works have been published in the field of generically benchmarking algorithms [6,5], and that our goal in this paper is to present a tool that, while generic, it still provides specific functions for the FCA use case.

The paper is organized as follows: Section 2 explains the notation and concepts we will use throughout the document. Section 3 presents a use case ofthe Formal Concept Analysis as a motivation for the creation of the generic tool. In Section 4, we discuss the state of the art of generic benchmarking tools for specific applications. Section 5 presents the software model, the context in which it is integrated, and the main algorithms. Finally,

Section 6 summarizesthe contribution and discusses the possible future work.

2 Preliminaries

2.1 Formal Concept Analysis

Formal Concept Analysis (FCA) is a clustering method whose input is a triple $K = (O, A, I)$, where O is a set of objects, A is a set of attributes, and I is an incidence matrix, indicating whether each object has an attribute or not

$$\text{e.g., } K = (\{o_1, o_2\}, \{a_1, a_2\}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix})$$

is a formal context in which o only has the attribute a_2, and o_2 has both. Alternatively we can see it as a *bipartite graph* i.e., O and A are the disjoint sets of *nodes*, and I is the set of *arcs*. The derivative operation $'$ on objects in the set $X \subseteq O$ is defined as the intersection of attributesof each object $o \in X$.

$$X' = \{a \in A \mid \forall o \in X : I_{o,a}\} \tag{1}$$

Analogously, we can define the derivative of a set of attributes as follows,

$$Y' = \{o \in O \mid \forall a \in Y : I_{o,a}\} \tag{2}$$

Having this in mind, a *formal concept* is a pair $C = (X, Y)$ where $X \subseteq O$, $Y \subseteq A$ such that $X' = Y$, $Y' = X$. X is called the extent and Y the intent. Put it into bipartite graph notation, a formal concept is a *bi-clique* i.e., a complete bipartite subgraph. For readability purposes, we note $C.E$ to the extent, and $C.I$ to the intent. The set of all the formal concepts and the relation of inclusion of extents form the so-called *concept lattice*, which is a partially ordered set, and is often noted with the letter L.

2.2 Common algorithms and their differences

Many reviews about algorithms for computing formal concepts have been made in the past [2,14,27]. There are many nuances to how they are implemented and also to their output. Some of them compute only the formal concepts, whilst others also calculate their underlying lattice diagram. For the purpose of this work, we will consider computing all the formal concepts and calculating their diagram to be separate problems, although they can be solved at the same time. As we mentioned in Section 1, and as we can see in the mentioned reviews,

there are many approaches on how to deal with the repetition of results in the calculation of the concepts, which occurs mainly because different subsets $X \subseteq O$, might yield the same $Y \subseteq \quad \subseteq A$ when the derivative operation is applied, but only the largest of them is present in a concept. Algorithms deal with this problem in different ways, from which in this work we aim to consider two: having a clever structure that allows to rapidly finding repeated results (e.g., Linding's algorithm, etc), or by traversing the context in a certain order that ensures that some results will not be repeated (e.g., Andrews' Inclose algorithm, etc).

2.3 Orange3 software

The *Orange3* software [18] is an open source machine learning and data visualisation tool whose aim is to make data analysis accessible to the end user in an intuitive way. To achieve this, it provides a way to pipeline data through "boxes" with certain input and output each, allowing to reuse them whenever necessary. This structure also allows simplifying the way to contribute to the project, since boxes can be thought as independent programs that define how to interact with their input and how to export the output. Additionally, Orange3 allows the development of separate plugins or *add-ons*, or in other words: external pieces of software that can be added to the main application. The scope of this paper is to introduce the architecture of an add-on with its components and interactions, and explaining how it would help to the analysis and benchmarking of formal concept analysis algorithms.

3 Motivation

To have a closer idea to what the Knomana dataset contains, an extract can be found in the Table 1. The formal context's objects are names of organisms composed by three parts: species, genus, and family. Even though they could be considered as different types, i.e., crops, pests, and protection species, in this example, we put them in the same table because they share the same attributes.

Table 1. Plants, crops and bio-aggressors formal context

K	Food	Medical
Abies sibirica/ Abies/ Pinaceae		
Acanthospermum hispidum/ Acanthospermum/ Asteraceae		X
Anticarsia gemmatalis/ Anticarsia/ Noctuidae		
Allium sativum/ Allium/ Amaryllidaceae	X	X
Spodoptera frugiperda/ Spodoptera/ Noctuidae		
Spodoptera littoralis/ Spodoptera/ Noctuidae		
Spodoptera litura/ Spodoptera/ Noctuidae		
CropS/ CropG/ CropF	X	X

CropFabaS/ CropFabaG/ Fabaceae	X	
Zanthoxylum rhetsa/ Zanthoxylum/ Rutaceae		X
Zingiber officinale/ Zingiber/ Zingiberaceae	X	X

Particularly, in the work [10], the FCA extension RCA is used to extract patterns in the data related to some plants being natural pesticides to other ones. Moreover, RCA needs to perform the algorithm to calculate the set of formal concepts many times in a loop until it converges, as explained in the Figure 1. Each iteration, using the calculated lattices and the relations in the input, increases the size of formal contexts in terms of their attributes, i.e., adds columns. This results on the possibility of the number of formal concepts increasing greatly, hence the need of fast algorithms to do it. In addition, available algorithms to calculate formal concepts perform differently according to the type of data, for instance, some of them perform better when formal contexts are sparse, and others when they are dense.

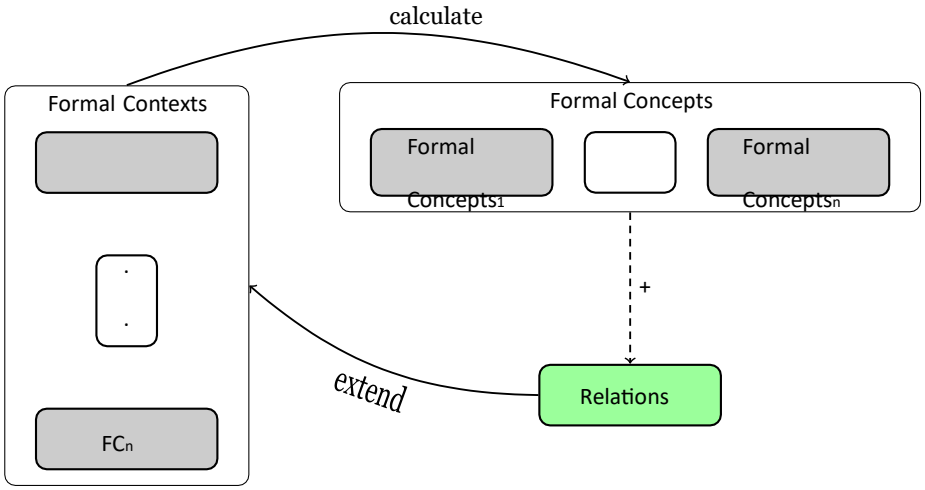


Fig. 1. RCA extension algorithm main loop

This leads us to a state in which making all kinds of experiments on these algorithms is necessary to better understand when which of them are more suitable according to the situation. However, it is rather tedious, since algorithms are typically different between themselves and there is not a common pattern to test them all the same way. Therefore, this work aims to smoothen the effort needed to perform benchmarking experiments on particularly formal concepts calculation algorithms. Additionally, the paper is intended to serve as a first approach guide on how to tackle generic denotational benchmarking frameworksusing metaprogramming techniques.

4 Related Work

For the purpose of this work, we consider that benchmarking tools are divided in two categories. The first one, is a one in which the tool must provide a set of options to reliably test a specific process that never changes. For example, a tool to automatically test REST API's load. Regardless of the specific endpoints, the benchmarking part would always follow the pattern of reaching the endpoints, measuring the time between request and response, and so on and so forth [8,28]. The second group is the one that involves giving to the user a generic set of functions to test something that we do not know in advance, e.g., benchmarking algorithms, functions in general. The challenge of this category is the fact that the process we want to test is not known beforehand, and thus, the techniques used to solve them usually involve metaprogramming or reflection [16]. A commonly applied method to tackle this type of tools, is to develop a *domain specific language* (DSL) in order to provide the users a flexible way to define what or how to test their functions [1,12].

In the case of benchmarking the calculation of formal concepts in FCA, the problem belongs to both categories. On the one hand, the main process will include one step that will always be a part of it: calculating derivatives (explained in Section 2). On the other hand, how or when the algorithm will do it is unknown and hence it belongs to the second category. Therefore, the solution we propose includes a part that takes advantage of the common pattern the method will always follow, and one in which the user is given the possibility to manually choose what and how to test. The disadvantage of the solution, is that it will not completely remove the manual effort required, but since the complexity is encapsulated in the provided functions, it will reduce it.

5 Software model

As explained in the Section 2.3, the units that users have to deal with in the platform are represented by *boxes* that are in fact algorithms with defined *inputs* and *outputs*. In that regard, the first input that concerns us is the name of a file representing a Formal Context, which, in our particular case, will be a *csv*. For the mentioned input, there should be a box called Formal Context that outputs the parsed formal context $K = (O, A, I)$ so that other algorithms do not have to deal with the parsing task over and over again. Then, to provide visualization to what we are parsing, there will be a box whose purpose is to show the bipartite graph representation of K (see Section 2). Furthermore, and continuing with the visualization, the add-on will provide a box for visualizing the Hasse diagram (and leave the door open to implement other visualizations such as the Iceberg concepts lattice).

Regarding the core and therefore the most important part of the architecture, the plugin will include a box that computes the list (or stream) of formal concepts, and that will act as the entry point for the generic benchmarking

abstract public interface (a.k.a., API). This box (red one in the Fig 2) will allow executing a default algorithm (Inclose), or to choose a user defined one. There is where, depending on what are the metrics the user specified to measure, the box will show them in a table format. The goal is to provide the user a set of generic decorators that allow to annotate specific functions, or specific parts of the algorithm to be measured in different ways: how many times a certain function is executed, how much memory it consumes, how much time in total or in average it spent during the execution, etc. And some specific ones related to formal context analysis, such as the times a derivative was calculated repeatedly.

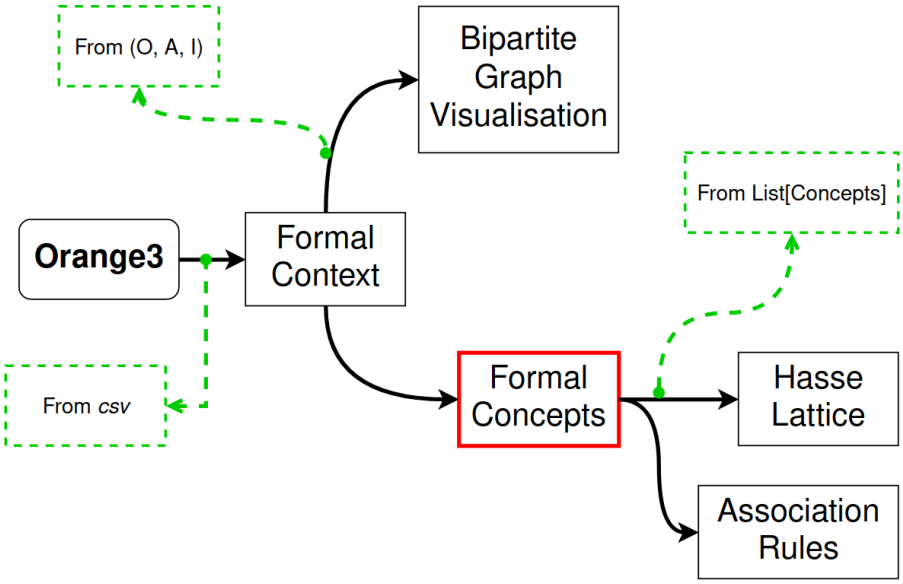


Fig. 2. Architecture diagram representing the expected components and theirinteractions

5.1 Benchmarking API

The benchmarking API will heavily rely on introspection and meta-programming patterns [16]. Particularly, it will use decorators to add meta information to the algorithm that will later be used for the *runner* to gather the data and be able to output it in some fashion. The notation we will use for the decorators is the same *Python* uses, and it consists of adding an @ to the beginning of each of them in order to identify them. Additionally, in the pseudocode, we will use *args as a way to say "any number of parameters". The generic decorators will be the following,

1. *@measure_time()*
2. *@measure_times_executed()*
3. *@measure_memory()*

while the FCA specific ones will be,

4. *@object_derivative()*
5. *@attribute_derivative()*

Specifically, the decorator 1 is expected to be applied to any function the user would want to measure the time it takes. In addition, the decorator 2 will count how many times a function is executed. And finally, the 3rd decorator will take note of the memory usage during the execution of the function and output the maximum usage of it. Notice that the *runner* will measure every decorated function, even if it is being called recursively, meaning that depending on what the user wants to do, sometimes it would be better to separate recursive functions in the *first call* and then the recursive one. Decorators 4 and 5, both describe a function that, each time is called, produces an object or an attribute derivative respectively. This provides essential information to the *runner* to measure how many times the overall algorithms repeats calculations.

Implementation Firstly, we will use the words *decorator* and *wrapper* interchangeably. On the one hand, the three first decorators are very similar in structure: they will wrap the function with a specific type of function or class that hints the runner to call it in a particular way and also output their specific type of information, e.g., a float representing time, an integer representing the amount of times the function has been executed, etc. When the wrapper is called, it initializes the necessary objects to gather the information, then it calls and returns the same as the function it wraps. As the execution ends, the wrapper will havethe information saved in a dictionary as an instance variable to be collected bythe *runner*.

On the other hand, the fourth and fifth decorators are different from the rest in the sense that they have to keep track of what are the structures[1] already generated. To do that, it is necessary to implement a way to tell whether two structures are the same or not.

[1] i.e., some implementation of the abstract data type *set.*

Algorithm 1: *measure_time* decorator

Input: *f*, a function or a *callable* class
Output: A callable class responsible for measuring the execution time of *f*
Def *wrapper(*args, collector)*:
1 *start_timer()*
2 res ← *f*(*args)
3 *time_passed* ← *stop_timer()*
4 *collector.add_measure_time(f, time_passed)*
5 **return** res
6 **end**
7 **return** *wrapper*

5.2 Runner

This component is the responsible for running the user provided code as its name suggests, but also for reporting the information at the end of the run. Thanks to the fact that decorators handle the complexity of knowing *when* the functions

Algorithm 2: *measure_times_executed* decorator

Input: *f*, a function or a *callable* class
Output: A callable class responsible for measuring how many times *f* is called
1 **Def** *wrapper(*args, collector)*:
2 *res* ← *f*(*args)
3 *collector.add_times_executed(f)*
4 **return** *res*
5 **end**
6 **return** *wrapper*

Algorithm 3: *measure_memory* decorator

Input: *f*, a function or a *callable* class
Output: A callable class responsible for measuring the maximum memory *f* consumes during its executions
1 **Def** *wrapper(*args, collector)*:
2 *profiled_function* ← *profile*(*f*)
3 *res, profiling_data* ← *profiled_function*(*args)
4 *collector.function_executed_with_memory(f, profiling_data.memory)*
5 **return** *res*
6 **end**
7 **return** *wrapper*

are executed, and also *what* to do in each case, for the running part, the runner should only execute the algorithm as it is. The challenge comes in the collecting part because the runner does not have control over when *some* of the functions are called. In fact, most of them will be instantiated to be executed and then discarded, thus, the runner would not get their results. This would lead to lose some data that the framework needs to report interesting metrics.

To solve this problem, all decorators will expect one more parameter besides the function to wrap, being an object whose purpose is to save each measurable function call. Then, our runner will run a modified version of the abstract syntax tree (AST) that provides this parameter, having access to this new object, and thus having access to the information after its execution.

Particularly, the Algorithm 1 starts a timer, runs the *original function* with its parameters, after, it measures the time that passed between the call and the end of the function, and finally, it tells the *collector* to add it to the total time spent for that specific function. Following the same pattern, the Algorithm 2, simply executes the function with its parameters, and afterwards it adds one to the total times executed for the specific function. And somehow more complex, the Algorithm 3 wraps the function to be executed to a profiling wrapper, and after executing it, sends the memory statistic to the *collector*. Lastly, the Algorithm 4 calls the collector's *object_derivative_calculated* function, to add 1 to the amount of times that particular derivative has been calculated. The algorithm for *attribute_derivative* would be exactly the same as 4 but calling

Algorithm 4: *object_derivative* decorator

Input: f, a function or a *callable* class
Output: A callable class responsible for measuring the amount of times a sethas been calculated

```
1 Def wrapper(*args, collector):
2 res ← f(*args)
3 collector.object_derivative_calculated(res)
4 return res
5 end
6 return wrapper
```

the function that adds to the attribute derivatives instead. It is important to notice that all these algorithms return a function defined inside it, meaning it is considering high order functions, i.e., functions as first class citizen values.

On top of each wrapper, the collector is expected to be an object with the following methods

Algorithm 5: *add_measure_time collector*'s method

Input: f, a function or a *callable* class, and t, an integer representing the time spent by f

Output: A mehod responsible for adding the time spent by f

```
1 times_table[f] += t
```

Algorithm 6: *add_times_executed collector*'s method

Input: f, a function or a *callable* class

Output: A mehod responsible for adding 1 to f times executes

```
1 times_executed_table[f] += 1
```

Algorithm 7: *function_executed_with_memory collector*'s method

Input: f, a function or a *callable* class, and m an integer representing the memory spent by f

Output: A mehod responsible for recording the most memory spent by the executions of f

```
1 memory_spent_table[f] ← max(memory_spent_table[f], m)
```

All three algorithms assume the existence of a mapping between functions and their specific value. In particular, the Algorithm 5 adds t to the mapping, here the algorithm assumes that the table has been previously initialized with 0, resulting in a semantic that will maintain the total amount of time spent by a function. In the same line, the Algorithm 6 adds 1 to the current value, meaning that it correctly counts how many times a function has been executed. Finally, the Algorithm 7 always remembers the maximum between what it had previously and the new m.

6 Conclusion and Future Work

In this work, we presented a generic tool to allow benchmarking certain aspects of FCA formal concepts generation algorithms instead of handcrafting them each time. The tool is currently being developed on their PIDIR, by Soukayna Ouabi and Loïc Chaillot, two students at TELECOM Nancy Engineering School. The advantages of the tool are not only the encapsulation and the centralization of the benchmarking complexities, but also the fact that it provides the programmers a denotational way to mark the parts of the code they want to benchmark, i.e., they write *what* instead of *how*.

Furthermore, the PIDIR project is expected to be extended in the coming months, so the tool can be upgraded to also include FCA extensions such as the widely used Relational Concepts Analysis (RCA). This will come with its own challenges, mainly in the area of software modelling. On top of that, many

challenges are still open in the development of a tool whose main goal is to be versatile and to provide an easy data visualization to the user. Combining that, plus the fact that both FCA and RCA produce an output with an exponential size in terms of their input, we realize that much work could be done in order to guarantee that all data can be explored, without the need of having everything loaded in memory.

Finally, the work as a whole could be considered as an approach to tackle problems involving the generation of tools for programmers and scientists working on algorithms creation and benchmarking in general, since the form to proceed should be in the lines of: understanding what needs to be measured, generating the denotational API, adding the collector, and when the time to benchmark is needed, the runner should always run a modified version of the AST adding the extra parameter to all the places needed, e.g., all functions decorated with the API decorators.

References

1. Albonico, M., Benelallam, A., Mottu, J.M., Sunyé, G.: A DSL-based approach for elasticity testing of cloud systems. In: Proceedings of the International Workshop on Domain-Specific Modeling. pp. 8–14. DSM 2016, Association for Computing Machinery, New York, NY, USA (Oct 2016). https://doi.org/10.1145/3023147.3023149
2. Arévalo, G., Berry, A., Huchard, M., Perrot, G., Sigayret, A.: Performances of Ga- lois Sub-hierarchy-building Algorithms. In: Kuznetsov, S.O., Schmidt, S. (eds.) Formal Concept Analysis. pp. 166–180. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg (2007). https://doi.org/10.1007/978-3-540-70901-5_11
3. Camarinha-Matos, L.M.: Challenges in IoT Applications and Research. Internet of Things. Technology and Applications. IFIPIoT 2021 (Jan 2022), https://www.academia.edu/73315599/Challenges_in_IoT_Applications_and_Research
4. Cardin, O.: Contribution à la conception, l'évaluation et l'implémentation de systèmes de production cyber-physiques. Habilitation à diriger des recherches, Univer- sité de nantes (Dec 2016), https://tel.archives-ouvertes.fr/tel-01443318
5. Darmont, J., Schneider, M.: Benchmarking OODBs with a Generic Tool. J. Database Manag. **11**, 16–27 (Jul 2000). https://doi.org/10.4018/jdm.2000070102
6. Digalakis, J., Margaritis, K.G.: On benchmarking functions for genetic algorithm. International Journal of Computer Mathematics **77** (Jan 2001). https://doi.org/10.1080/00207160108805080
7. Džeroski, S.: Multi-relational data mining: an introduction. ACM SIGKDD Explorations Newsletter **5**(1), 1–16 (Jul 2003). https://doi.org/10.1145/959242.959245
8. Fertig, T., Braun, P.: Model-driven Testing of RESTful APIs. In: Proceedings of the 24th International Conference on World Wide Web. pp. 1497–1502. WWW '15 Companion, Association for Computing Machinery, New York, NY, USA (May 2015). https://doi.org/10.1145/2740908.2743045, https://doi.org/10.1145/2740908.2743045
9. Ganter, B., Stumme, G., Wille, R.: Formal Concept Analysis: Foundations and Applications. Lecture Notes in Computer Science, Springer Berlin Heidelberg (2005), https://books.google.fr/books?id=oyb6BwAAQBAJ
10. Keip, P., Gutierrez, A., Huchard, M., Le Ber, F., Sarter, S., Silvie, P., Mar-tin, P.: Effects of Input Data Formalisation in Relational Concept Analysis for a Data Model with a Ternary Relation. In: Cristea, D., Ber, F.L., Sertkaya, B. (eds.) ICFCA 2019 -

15th International Conference on Formal Concept Analysis. Lecture Notes in Computer Science, vol. 11511, pp. 191–207. Springer Interna- tional Publishing, Frankfurt, Germany (Jun 2019). https://doi.org/10.1007/978-3-030-21462-3_13, https://hal-lirmm.ccsd.cnrs.fr/lirmm-02092148
11. Keip, P., Gutierrez, A., Huchard, M., Le Ber, F., Sarter, S., Silvie, P., Martin, P.: Effects of Input Data Formalisation in Relational Concept Analysis for a Data Model with a Ternary Relation. In: Cristea, D., Le Ber, F., Sertkaya, B. (eds.) Formal Concept Analysis. pp. 191–207. Lecture Notes in Computer Science, Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-21462-3_13
12. King, T.M., Nunez, G., Santiago, D., Cando, A., Mack, C.: Legend: an agile DSL toolset for web acceptance testing. In: Proceedings of the 2014 International Sym- posium on Software Testing and Analysis. pp. 409–412. ISSTA 2014, Association for Computing Machinery, New York, NY, USA (Jul 2014). https://doi.org/10.1145/2610384.2628048
13. Kuznetsov, S., Obiedkov, S.: Algorithms for the Construction of Concept Lattices and Their Diagram Graphs, vol. 2168 (Sep 2001). https://doi.org/10.1007/3-540-44794-6_24, journal Abbreviation: Principles of Data Mining and Knowledge Discovery - Lecture Notes in Computer Science Pages: 300 Publication Title: Principles of Data Mining and Knowledge Discovery - Lecture Notes in Computer Science
14. Kuznetsov, S.O., Obiedkov, S.A.: Comparing performance of algorithms for generating concept lattices. Journal of Experimental & Theoretical Artificial Intelligence 14(2-3), 189–216 (Apr 2002). https://doi.org/10.1080/09528130210164170, http://www.tandfonline.com/doi/abs/10.1080/09528130210164170
15. Lezoche, M., Hernandez, J.E., Alemany Díaz, M.d.M.E., Panetto, H., Kacprzyk, J.: Agrifood 4.0: A survey of the supply chains and tech- nologies for the future agriculture. Computers in Industry **117**, 103187 (May 2020). https://doi.org/10.1016/j.compind.2020.103187, https://linkinghub.elsevier.com/retrieve/pii/S0166361519307584
16. Lilis, Y., Savidis, A.: A Survey of Metaprogramming Languages. ACM Comput. Surv. (2020). https://doi.org/10.1145/3354584
17. Liu, Y., Liu, Y., Chen, T.Y., Zhou, Z.Q.: A Testing Tool for Machine Learning Applications. In: Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, pp. 386–387. Association for Computing Ma- chinery, New York, NY, USA (Jun 2020), https://doi.org/10.1145/3387940.3392694
18. Ljubljana, University of, B.L.: Orange3, https://orangedatamining.com/
19. Manning, C. D.,Raghavan,P., Schütze,H.:Introduction to InformationRetrieval(Jul 2008). https://doi. org/10.1017/CBO9780511809071, ISBN: 9780511809071 Publisher: Cam bridge University Press
20. Panetto, H., Lezoche, M., Hernandez Hormazabal, J.E., del Mar Eva Ale- many Diaz, M., Kacprzyk, J.: Special issue on Agri-Food 4.0 and digitalization in agriculture supply chains - New directions, challenges and applications. Computers in Industry **116**, 103188 (Apr 2020). https://doi.org/10.1016/j.compind.2020.103188
21. Rouane-Hacene, M., Huchard, M., Napoli, A., Valtchev, P.: Relational Concept Analysis: Mining Concept Lattices From Multi-Relational Data. Annals of Math- ematics and Artificial Intelligence **67** (Jan 2013). https://doi.org/10.1007/ s10472-012-9329-3
22. Sayad, S.: Real Time Data Mining (Jan 2017)
23. Tamrakar, E.S.: Formal concept analysis: mathematical foundations (Jan 1997), https://www.academia.edu/3362029/Formal_concept_analysis_mathematical_foundations
24. Wajnberg, M.: Analyse relationnelle de concepts : une méthode polyvalente pour l'extraction de connaissance. Theses, Université du Québec à Montréal ; Université de Lorraine (Nov 2020), https://hal.archives-ouvertes.fr/tel-03042085, issue: 2020LORR0136
25. Wajnberg, M., Valtchev, P., Lezoche, M., Panetto, H., Massé, A.: Mining ProcessFactor Causality Links with Multi-relational Associations. p. 266. Marina Del Rey, CA, United

States (Sep 2019). https://doi.org/10.1145/3360901.3364446
26. Yang, C.S.D., Pollock, L.L.: Towards a structural load testing tool. In: Proceedings of the 1996 ACM SIGSOFT international symposium on Software testing and analysis. pp. 201–208. ISSTA '96, Association for Computing Machinery, New York, NY, USA (May 1996). https://doi.org/10.1145/229000.226318
27. Zaki, M., Hsiao, C.J.: Efficient algorithms for mining closed itemsets and their lattice structure. Knowledge and Data Engineering, IEEE Transactions on **17**, 462–478 (May 2005). https://doi.org/10.1109/TKDE.2005.60
28. Zhang, Y., Fu, W., Nie, C.: monadWS: a monad-based testing tool for web services. In: Proceedings of the 6th International Workshop on Automation of Software Test. pp. 111–112. AST '11, Association for Computing Machinery, New York, NY, USA(May 2011). https://doi.org/10.1145/1982595.1982622

A Strategy for Building Shared Understanding in Requirements Engineering Activities

Vanessa Agredo Delgado[1,2][0000-0003-0870-6895], Pablo H. Ruiz [1,2][0000-0003-2098-2614], Luis E. Garzón[3], and César A. Collazos[2][0000-0002-7099-8131]

[1] Corporación Universitaria Comfacauca - Unicomfacauca, Popayán, Colombia
{vagredo, pruiz}@unicomfacauca.edu.co
[2] Universidad del Cauca, Popayán, Colombia
ccollazo@unicauca.edu.co
[3] SENA Centro de Comercio y Servicios
lgarzonv@sena.edu.co

Abstract. The requirements allow the development team to clearly understand the needs that the customer intends to be solved by the system, in this sense, understanding the context, capturing, negotiating, specifying, verifying, validating, and prioritizing the requirements may seem a relatively simple task, but there is a need to have a correct communication, and throughout this process, many changes and reprocesses occur due to misinterpretation or lack of information, in addition to considering that in the teams that perform these activities participate people from different disciplines, business units, cultures, with different levels of experience and therefore, each one will have different ways of perceiving the tasks, the key problems, which give meaning to the requirements according to their situation and knowledge, without having a joint base of homogeneous understanding within the team. Therefore, this work proposes a strategy for the construction of a shared understanding in the activities of requirements engineering, where its completeness, usefulness, and ease of use were validated, through an experiment executed as part of the development process of a software tool for the management of information and data processing of an agricultural and livestock association in Cauca. Using the conceptual, methodological, and validation cycle of the multi-cycle action research methodology, it was concluded that the strategy is complete and useful, but it is not easy to use, because its definition contains several elements that are difficult to handle, and it lacks adequate support to support and facilitate its application.

Keywords. Shared understanding, Requirements engineering, Requirements, Strategy.

1 Introduction

Software Engineering provides a set of methods and techniques for the creation of quality and reliable software. It covers all phases of the software development cycle, requirements, analysis and design, implementation, testing, and deployment. Much of the success of software development is due to a correct requirements management,

where the needs that the software system must satisfy must be found and identified. The requirements management is considered part of the first phase of software development in which the problem that the software product will solve is abstracted and understood, it is, essentially, a human activity where stakeholders are also identified and established relations with the development team [1]. The requirements allow the members of the development team to understand the needs that the client intends to be solved by the system, in this sense, understanding the context, capturing, negotiating, specifying, verifying, validating, and prioritizing the requirements may seem like a relatively simple task, but the need for successful communication is very high, and throughout this process there are many changes and re-processes due to misinterpretation, or lack of information, in addition to considering that the teams that carry out these activities They involve people from different disciplines, business units, cultures, with different levels of experience and therefore, each one will have different ways of perceiving the tasks, key problems, which makes sense of the requirements according to their situation and knowledge, without counting on a base joint homogeneous understanding within the team. Shared understanding refers to the degree to which team members agree on the steps of a work process, the meaning of those steps, the order, the relationship of activities, and their communication [2]. Considering this, the knowledge creation process implies participation, collaboration, and the achievement of a shared understanding [3]. Carrying out correct management of the communication and understanding of the requirements process is one of the main elements for its success, because there may be different interests with different expectations of understanding and fulfillment that can produce misinformation that, by not being treated correctly, it can generate gaps in communication and understanding between stakeholders and the development team [4]. Therefore, this paper proposes a strategy for the construction of the shared understanding in the requirements engineering activities, which was validated its completeness, usefulness, and ease of use through an experiment executed as part of the development of a software tool for information management and data processing of an agricultural and livestock association in Cauca - Colombia.

The paper is structured as follows: Section 2 describes the methodology, the conceptual, methodological and validation cycle of the methodology used, and section 3 describes the conclusions.

2 Methodology

This research was developed following the multi-cycle action-research methodology with bifurcation [5], for which cycles were followed: conceptual cycle, methodological cycle, and evaluation cycle.

2.1 Conceptual cycle

This cycle consisted of conducting a review of the related works that could support the definition of the proposal, its subsequent construction, and the correct application

to validate it, which is why the most significant related works for this project are shown below:

In [6] a theory of coordination and communication is proposed in software organizations based on shared understanding, where it is highlighted that coordination and communication are essential and problematic elements. The role, value, and use of shared understanding in software engineering are investigated in [7], showing a practices compilation, as well as a roadmap to improve knowledge and practice in this area. For his part in [8] it addresses the influence of team distribution on the success of the project with a shared understanding approach. The theory of shared mental models is used for the construction and maintenance of shared understanding. In [9] communication is analyzed, through shared understanding, of the underlying concepts or relationships of a multidisciplinary team in the development of a mobile application. Similarly in [10] communication and the development of shared understanding based on the language are emphasized. The semantic alignment process is investigated by which stakeholders achieve a shared understanding in the development of software system requirements. On the other hand, in [11] analysis is made of how culture affects the shared understanding of requirements engineers, as well as in the organization and progress of software projects. In [12] the role of cognitive elements is investigated to improve the clarity of the user's story, for which a set of writing elements from different domains is proposed in order to mitigate ambiguity and improve the shared understanding. In [13] a process is proposed to achieve a shared understanding based on the construction of meaning through knowledge of the group and the constructive resolution of conflicts in requirements gathering workshops. For its part, [14] it provides a conceptualization of shared understanding as a sequence of state transitions at the group level based on the specialization construct of the team's mental model that participates in requirements. In [15] the importance of shared understanding is considered in the context of e-science projects, in addition, qualitative case studies are developed to generate recommendations to improve shared understanding in electronics science application requirements. Finally, in [16], a case study is carried out in three small organizations to understand and identify the factors that contribute to the lack of shared understanding in eliciting the non-functional requirements and what relation it has to its reworking.

2.2 Methodological cycle

The methodological cycle refers to the process of creating the strategy. In this sense, to meet this objective, the information previously obtained was analyzed, allowing the creation of a version of the strategy, which contains activities, tasks and steps that will allow executing the requirements engineering activities in a collaborative way, seeking to achieve a shared understanding during the whole process. For our context, a strategy refers to: a set of actions that are aimed at establishing a guide [5] to execute a requirements engineering activities that meets the expectations of the client and the development team of the product to be built.

To build an adequate strategy, it is first necessary to be clear about the requirements engineering activities, this in order for it to be solved and completely guided by the strategy defined here. Accordingly, requirements engineering refers to the process of collecting, analyzing, and verifying the needs of the client or user for a system, delivering a correct and complete software requirements specification [17]. This process consists of the following activities [18]:

- To understand the context in which the system to be developed will be executed
- To capture the necessary information according to the stakeholders and sources considered
- To negotiate with stakeholders, the solutions to the problems identified
- To specify each of the requirements and needs identified in a defined notation
- To verify that the requirements are complete, unambiguous, verifiable, and correctly detailed
- To validate requirements by presenting them to stakeholders in order to ensure that needs and expectations have been properly captured and expressed
- To prioritize requirements according to the value that stakeholders place on the vision of the system, the urgency, time constraints, complexity or preferences

On the other hand, shared understanding refers to creating a new joint perspective that arises from the initially individual contributions of the participants, and from the exchange of knowledge, flow of communication and holding of debates that allow coordinating actions in order to achieve the objective of the collaborative activities they carry out [19] [20]. In this sense, the shared understanding for requirements engineering is an important determinant for performance, as well as a challenge in these heterogeneous groups [21], this is due to, the fact that those involved in each task may be using the same words for different concepts or different words for the same concepts without realizing it, making the final requirements not the most appropriate and each interpreting them differently [22]. Differences in the meaning assigned to key concepts or information can interfere with the productivity of collaborative work if they are not clarified up front [23], [24], [25].

Considering the above, to define and incorporate collaboration into said strategy, the collaborative engineering design approach was followed [26], which addresses the challenge of designing and implementing collaborative work practices for recurring high-value tasks and transferring them to professionals to execute it themselves without the ongoing support of a collaborative professional expert [22]. In this sense, a strategy called "*Brainstorming for shared understanding*" was defined, in order to guide the entire requirements engineering process, which is made up of activities, tasks and steps. Specifically, to comply with a strategy that builds shared understanding and therefore collaboration, each activity must have the following tasks in terms of its structure (See **Fig. *1***).

Fig. 1. Tasks structure of each activity

The individual actions refer to the moment in which each participant must obtain individual results to fulfill the activity objective, based on the needs for the construction of tacit knowledge (that which is acquired through experience itself) and which must subsequently become explicit at the time of materializing it in a result [27].

The resolution of doubts refers to the moment in which each participant solves those questions they have about the subject being analyzed

Share refers to the moment in which each participant inserts a meaning, tuning in to the other groupmates in the group, who listen actively and try to capture the explanation or results given, using them to give meaning to the situation in question [28].

Debate refers to the moment in which a mutual construction of meaning is carried out, treating the differences of interpretation between the participants of the group through discussions with arguments and clarifications [29].

Group actions, is the moment in which the interpretation of meanings or actions carried out with the support and collaboration of all the participants of the group materialize to fulfill the activity objective

With the previous structure, each of the activities of the strategy proposed here was defined, considering, that this structure served as the basis to achieve the objectives of each activity as shown in the following figure (See **Fig. *2***).

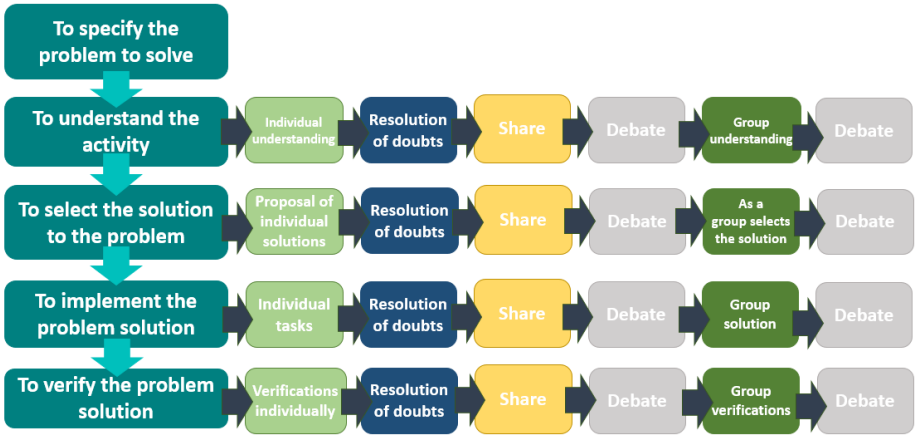


Fig. 2. Activities and tasks of the strategy

The objective of each activity of the strategy is shown below, with their respective inputs and outputs. In addition to what is shown here, it is important to determine that the execution steps, estimated time and their respective formats were defined for each of the tasks to guide each of the steps to be executed.

To specify the problem to solve: The objective of this activity is for the participants to know and contextualize themselves about the problem to be solved, the collaborative activity objective, and what they will have to execute throughout the brainstorming.

Inputs**:** The problem of collaborative activity, the collaborative activity objective, brief description of brainstorming

Outputs**:** None

To understand the activity: The objective of this activity is that each participant individually understands the problem to be solved and the collaborative activity objective and that in the same way, after the debates, an equal understanding is reached for all and where everyone agrees, which is can finally materialize into a deliverable made with the help of everyone, about the problem understood and the objective of the collaborative activity.

Inputs**:** Format to define doubts, format for defining individual ideas, format for discussion, format for the definition of group understanding, format for group understanding discussion

Outputs**:** Individual ideas, group defined idea

To select the solution to the problem: The objective of this activity is to select the solution that is going to be implemented by the group to solve the problem of collaborative activity, initially giving ideas of individual solutions, later, with the socialized ideas, they are categorized and according to these categories, the chosen solution is chosen and formalized with the contribution of all participants.

Inputs**:** Format for individual solution ideas, format to define doubts, format for discussion, format to define chosen solution, format for discussion of the chosen solution

Outputs**:** Individual solution ideas, ideas categorization, solution chosen to implement

To implement the problem solution: The objective of this activity is to define, and subsequently, execute the individual tasks that will allow solving the problem. Each participant will share their individual results, and after this, the complete solution of the problem will be formalized, together with the contributions of all the participants.

Inputs**:** Format for defining individual tasks, format for socializing results of executing individual results, format to define doubts, format for discussion, format to formalize the solution implemented in group, format for discussion of the implemented solution

Outputs**:** Defined individual tasks, individual tasks executed, formalization of the implemented solution

To verify the problem solution: The objective of this activity is to verify that the solution implemented by the group does solve the problem posed in the collaborative activity. For this, each of the participants defines possible scenarios where the solution of the problem is implemented, later socialization is made with the group, and from this, the scenarios in which there is a correct solution to the problem are defined in a group, to determine if it was executed correctly and complies with the request.

Inputs**:** Format for defining individual scenarios, format to define doubts, format for discussion, format to formalize scenarios where the group-defined solution is implemented, format for discussion of the verification carried out

Outputs**:** Individual scenarios, scenarios with the implementation of the solution defined in the group

With the previous strategy defined and considering the requirements engineering activities mentioned above, below, the correspondence of how the entire strategy will be carried out is shown in order to finally obtain a set of requirements that will meet the needs of the end users of the system and a shared understanding between these users and the development team (**Fig. 3**).

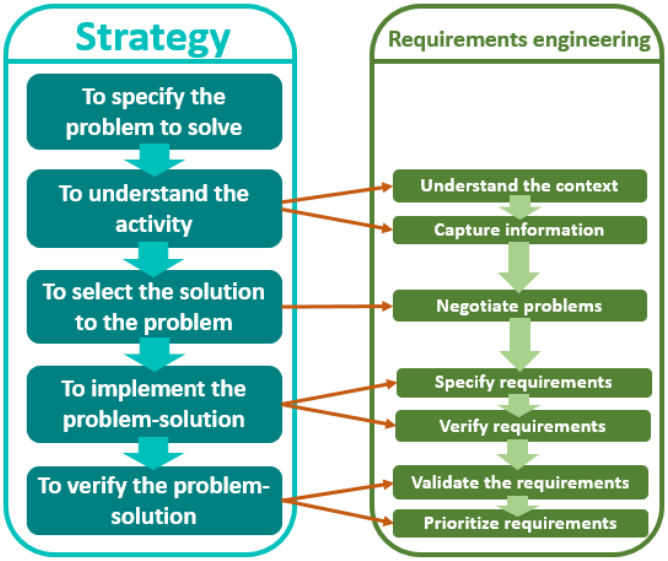


Fig. 3. Correspondence between strategy activities and requirements engineering activities

2.3 Validation cycle

This cycle made it possible to achieve the objective of inquiring about the completeness, usefulness and ease of use of the defined strategy, through its application in the developing process a software tool, for the management of

information and data processing of an association livestock, it is important to clarify that the strategy was subjected to several revisions of its structure and its definition, carried out by a member of IDIS research group of the University del Cauca and a member of MIND research group of Unicomfacauca. In addition, a review session was held with an expert in group work and collaborative engineering and an expert in requirements engineering, who reviewed the strategy to indicate whether it had the necessary elements to satisfy these two areas. Several corrections were made to the strategy before it was implemented in practice. The experiment is summarized in the following sections.

Experiment Context

The entire strategy was applied in a real environment, where it was necessary to obtain the requirements for a software tool development for ASPROLGAN (Asociación de Productores Lácteos y Agro ganaderos del Municipio de Popayán), which is located in Popayán city, Cauca department - Colombia. It is a non-profit association made up of 94 associates who in turn influence 470 people who belong to their family nuclei; All these peasant and indigenous families have found in livestock a form of family sustenance. The association as an organization must support its administrative management processes such as: planning, organization, direction and control, but these processes are not currently being executed in the best way, because there is not enough, standardized, available information and accessible, where now it is managed manually. Therefore, the association's need is mainly to improve information management and data processing through the use of a software tool for the dairy sectors that belong to the ASPROLGAN association

For the implementation of the strategy, 8 members of the development team participated (including the project manager, 2 quality engineer, 3 developers, 2 analysts and 1 software architect) and 5 members of the association.

For the context of this work, the problem to be solved consisted in doing the requirements engineering process to finally obtain a set of requirements that a software tool should have that is in charge of the information management and data treatment processes in the dairy sectors of the indigenous and peasant community of the San Juan and San Ignacio villages belonging to ASPROLGAN. To solve the problem, a group was formed, where both the development team and the 2 members of the association had to follow the strategy outlined here and thus obtain the necessary requirements.

Experiment planning

The experiment objective was to inquire about the completeness, usefulness and ease of use of the proposed strategy for the shared understanding construction in the requirements engineering activities execution. In this sense, the research question was defined as: How complete, useful and easy to use is the strategy “Brainstorming for shared understanding” proposed here? This study had an analysis unit, which was the real context, where the requirements engineering activities was carried out for a software tool construction for information management and data treatment in the

dairy sectors of the indigenous community and peasant from the San Juan and San Ignacio villages belonging to ASPROLGAN, using the proposed strategy.
Hypothesis. Considering the research question, it is intended to evaluate the following hypotheses:

- The strategy "Brainstorming for shared understanding" is complete[2] with respect to having the necessary elements for the shared understanding construction and the requirements engineering activities execution
- The strategy "Brainstorming for shared understanding" is useful[3] for the shared understanding construction and the requirements engineering activities execution
- The "Brainstorming for shared understanding" strategy is easy to use[4] for shared understanding construction and the requirements engineering activities execution

In order to refine the previous hypotheses, the following specific hypotheses with their respective variables were raised (See **Table *1***):

Table 4. Experiment hypothesis.

	Hypothesis	Variables
Completeness	H.1.1 Users who apply the proposal perceive that the activities, tasks and steps are sufficient for the shared understanding construction	It represents the completeness degree perceived by each person when applying the strategy. It is a perceptual judgment of the completeness of the proposal when building shared understanding
	H.1.2 Users who apply the proposal perceive that the activities, tasks and steps are sufficient for the requirements engineering activities execution	It represents the completeness degree perceived by each person when applying the strategy. It is a perceptual judgment of the completeness of the proposal when executing the requirements engineering activities
Utility	H.2.1 Users who apply the proposal perceive that the strategy is useful for the shared understanding construction	It represents the utility degree perceived by each person when applying the strategy. It is a perceptual judgment of the utility of the proposal when building shared understanding
	H.2.2 Users who apply the proposal perceive that the strategy is useful for the requirements engineering activities execution	It represents the utility degree perceived by each person when applying the strategy. It is a perceptual judgment of the utility of the proposal when executing the requirements engineering activities
Ease of use	H.3.1 Users who apply the proposal perceive that the strategy is easy to use for the shared understanding construction	It represents the perceived degree ease to use with which a person can apply the strategy. It is a perceptual judgment of the effort required to apply the proposal when build shared understanding

[2] Completeness in this context refers to the fact that the strategy contains the necessary elements, steps and support.
[3] Utility in this context refers to the fact that the strategy is organized and consistent in its definition to achieve what is necessary.
[4] Ease of use in this context means that the strategy contains instructions, guidelines, supporting elements that are understood and can be used without additional support.

H.3.2 Users who apply the proposal perceive that the strategy is easy to use for the requirements engineering activities execution	It represents the perceived degree ease to use with which a person can apply the strategy. It is a perceptual judgment of the effort required to apply the proposal to execute the requirements engineering activities

Table 2 Summarizes the activities designed for the experiment development, specifying its expected duration and the support instruments that would be used for its development.

Table 5. Experimentation activities summary

Activity	Planned duration	Support instruments
Activity 1: Group organization and information delivery	15 minutes	Documentation with the general context and ASPROLGAN needs
Activity 2: Strategy implementation	15 hours	Input and output formats for strategy activities
Activity 3: Questionnaire fill out	10 minutes	Survey

Execution of the experiment

The following table shows and details how each of the activities of the experiment was executed (**Table *3***):

Table 6. Time invested in each activity

Activity	Time invested
Activity 1	10 minutes
Activity 2	5 sessions of 4 hours each
Activity 3	10 minutes

Activity 1: This activity aimed to socialize and contextualize in a general way what the experiment was. An oral presentation was made in order to inform the participants of how the experiment would be carried out, about the activities that would be developed, in addition to making known, clarifying some concepts used in it, and to socializing in general terms which ones were the ASPROLGAN needs.

Activity 2: The objective of this activity was to execute each of the activities, tasks and steps proposed by the strategy, for each of the strategy activities, independent sessions were held as follows:

- To specify the problem to solve: It was announced that for this experiment the problem was to obtain a set of necessary requirements that would satisfy the association needs for the management of its information and data processing through a software tool.

- To understand the activity: An informal contextualization meeting was initially held between the development team and the association members, in order to publicize their context and the specific needs that the software tool should cover and clear up doubts on the part of the development team. After that, each participant defined their understanding of the context and needs in a format, which were shared with the other members of the development team. This information was then categorized, doubts were solved, debated and finally a specification was made in a format, where everyone participated to define the needs identified in the context, which was also debated to reach a consensus of what was stipulated.
- To select the solution of the problem: Each member of the development team filled out a form that specified how the software tool intended to solve each of the identified needs. Then, these formats were socialized, doubts were solved, categorized, those solutions were chosen, which according to the perception of the development team were the best for the needs, they were discussed and finally, according to these chosen solutions, they were completed and improved with everyone's contribution and with the necessary discussions. These solutions were shown to the association members who gave their points of view, solved some doubts and problems encountered, and in this way the formats presented were corrected.
- To implement the problem solution: The established solutions were divided and assigned to members of the development team, who individually defined the epic stories, user stories, SRS (Software Requirements Specification), and solution prototypes, which corresponded to them. With this done, doubts are resolved, each member of the team socialized what they did, debates are generated with the disagreements found, finally contributions are made between all to improve and correct the deliverables and form a complete solution that is part of the requirements specification, with this generates the necessary discussions. With this ready specification, a verification is made with people external to the development team to determine if the requirements are complete, unambiguous, verifiable and expressed with an appropriate level of detail, according to the results they are corrected and improved.
- To verify the problem solution: A complete prototype is delivered to each of the development team members and to the association members, where each one defines possible scenarios that can happen within the association for be solved with the presented prototype, doubts are solved, the information is socialized obtained, the debates are generated, with the support of all, the errors detected are corrected, the prototype is improved according to the defined scenarios and debates are generated. With a more stable version and where everyone agrees, all requirements are prioritized according to the needs of the association and the time that will be taken to develop the application

Activity 3: In this last activity, the objective was that the participants answered a survey, which made it possible to evaluate the completeness, usefulness and ease of use of the applied strategy.

Results and analysis

The qualitative analysis was carried out from the surveys completed by the development team members and the association members who participated in the strategy application. The responses to the survey were based on the Linkert scale, which is a form of measurement that allows evaluating attitudes and knowing the agreement degree on a set of statements. The measurement scale of the survey was defined as follows: value 1 for the totally disagree option, value 2 for the disagree option, value 3 for the neutral option (neither agree nor disagree), value 4 for the agree option and 5 for the totally agree option. From the hypotheses initially drawn, the following null hypotheses were raised:

- $H.1.1_0$, $\pi_1 <= 60\%$, where π_1 is the perception percentage that evaluates that the activities, tasks and steps of the strategy are sufficient for the shared understanding construction
- $H.1.2_0$, $\pi_2 <= 60\%$, where π_2 is the perception percentage that evaluates that the activities, tasks and steps are sufficient for the requirements engineering activities execution
- $H.2.1_0$, $\pi_3 <= 60\%$, where π_3 is the perception percentage that evaluates the strategy usefulness for the shared understanding construction
- $H.2.2_0$, $\pi_4 <= 60\%$, where π_4 is the perception percentage that evaluates the strategy usefulness for the requirements engineering activities execution
- $H.3.1_0$, $\pi_5 <= 60\%$, where π_5 is the perception percentage that evaluates the strategy ease of use for the shared understanding construction
- $H.3.1_0$, $\pi_6 <= 60\%$, where π_6 is the perception percentage that evaluates the strategy ease of use for the requirements engineering activities execution

From the null hypotheses the following alternative hypotheses were obtained:

- H.1.1, $\pi_1 > 60\%$, where π_1 is the perception percentage that evaluates that the activities, tasks and steps of the strategy are sufficient for the shared understanding construction
- H.1.2, $\pi_2 > 60\%$, where π_2 is the perception percentage that evaluates that the activities, tasks and steps are sufficient for the requirements engineering activities execution
- H.2.1, $\pi_3 > 60\%$, where π_3 is the perception percentage that evaluates the strategy usefulness for the shared understanding construction
- H.2.2, $\pi_4 > 60\%$, where π_4 is the perception percentage that evaluates the strategy usefulness for the requirements engineering activities execution
- H.3.1, $\pi_5 > 60\%$, where π_5 is the perception percentage that evaluates the strategy ease of use for the shared understanding construction
- H.3.1, $\pi_6 > 60\%$, where π_6 is the perception percentage that evaluates the strategy ease of use for the requirements engineering activities execution

From the results obtained in the surveys, it was obtained that:

- For the analysis of activities, tasks and steps of the strategy to determine if they are sufficient for the shared understanding construction, the participants' perception percentage is 68.53%, which determined that H.1.1 can be accepted, it can be said that the strategy is complete for the shared understanding construction
- For the analysis of activities, tasks and steps the strategy to determine if they are sufficient for the shared understanding construction, the participants' perception percentage is 72.3%, which determined that H.1.2 can be accepted, it can be said that the strategy is complete for the requirements engineering activities execution
- For the analysis of the strategy usefulness, the participants' perception percentage is 75.4%, which determined that H.2.1 can be accepted, it can be said that the strategy is useful for the shared understanding construction
- For the analysis of the strategy usefulness, the participants' perception percentage is 67.6%, which determined that H.2.2 can be accepted, it can be said that the strategy is useful for the engineering requirements activities execution
- For the analysis of strategy ease of use, the participants' perception percentage is 55.2%, which determined that H.3.1 can be rejected, it can be said that the strategy is not easy to use for the shared understanding construction
- For the analysis of strategy ease of use, the participants' perception percentage is 46.5%, which determined that H.3.2 can be rejected, it can be said that the strategy is not easy to use for the requirements engineering activities execution

With the specific hypotheses accepted, it can be inferred that the main hypotheses are accepted, determining that: the "Brainstorming for shared understanding" strategy is complete and useful in the sense that it has the necessary elements for the construction of shared understanding and the execution of requirements engineering activities. However, the strategy is not easy to use for the construction of shared understanding and the execution of requirements engineering activities.

3 Conclusions and future work

This paper proposes a strategy for the shared understanding construction in requirements engineering tasks following the conceptual, methodological, and validation cycle of the multi-cycle action-research methodology. The set of activities that are part of the strategy is shown and defined, as well as the correspondence and compatibility with the activities used in requirements engineering. The strategy was validated through an experiment carried out in the development process context of a software tool for information management and data processing of the ASPROLGAN

livestock association. According to the validation carried out in this context, with the hypotheses of H.1.1 and H.1.2 it can be concluded that the participants perceived that the strategy has sufficient elements for the construction of shared understanding and the requirements engineering activities execution. Similarly, in the validation of hypotheses H.2.1 and H.2.2 regarding utility, it can be concluded that the strategy is useful because it was perceived to be organized and consistent in its definition. Regarding the ease of use of the strategy, validated in hypotheses H.3.1 and H.3.2, it can be concluded that the instructions, guidelines, support elements that it contains need additional support so that their understanding is more suitable, in such a way that this affected its ease of use. As future work, it is expected to make improvements to the strategy corresponding to its ease of use, to later be applied in other software development projects that allow its definition to mature.

References

1. M. M. Uolevis Nikual, "La elicitación de requisitos en el contexto de un proyecto software," *Ingenierías USBMed,* vol. 2, no. 2, pp. 25-29, 2011.
2. E. A. Christiane Bittner and J. M. Leimeister, "Creating shared understanding in heterogeneous work groups: Why it matters and how to achieve it," *Journal of Management Information Systems,* vol. 31, no. 1, pp. 111-144, 2014.
3. D. S. Stein, C. E. Wanstreet and H. R. Glazer, "Knowledge building online: The promise and the process," in *Encyclopedia of Information Communication Technologies and Adult Education Integration*, IGI Global, 2011, pp. 985-998.
4. L. Westfall, "Las fallas en la ingeniería de requisitos," *Ingenierías USBMed,* vol. 2, no. 2, pp. 40-47, 2011.
5. V. Lencinas, F. Nataloni, S. Hannover Saavedra, T. G. Villanueva, S. Lacolla, M. López and I. G. Zabczuk, "Una oportunidad para generar conocimiento desde la practica profesional de bibliotecarios y archiveros," *Revista Prefacio ,* vol. 1, no. 1, pp. 10-23, 2017.
6. J. Aranda, *A Theory of Shared Understanding,* Toronto, 2010.
7. M. Glinz and S. Fricker, "On Shared Understanding in Software Engineering," *Lecture Notes in Informatics (LNI) ,* p. 213, 2013.
8. H. Markus, C. Rosenkranz and R. Holten, "The Role of Shared Understanding in Distributed Scrum Development: An Empirical Analysis," *Engineering, Computer Science,* 2016.
9. A. Hoffmann, E. Bittner and J. Leimeister, "The Emergence of Mutual and Shared Understanding in the System Development Process," *Lecture Notes in Computer Science ,* 2013.
10. M. Charaf, C. Rosenkranz and R. Holten, "The emergence of shared understanding: applying functional pragmatics to study the requirements development process," *Information Systems Journal (ISJ) ,* 2012.
11. Y. Hsieh, *Culture and Shared Understanding in Distributed Requirements Engineering,* IEEE International Conference on Global Software Engineering (ICGSE'06), 2006.
12. A. R. Amna, "The Use of Context-Free Grammar and Gamification to Reduce User Story

Ambiguity and Raise Ambiguity Awareness: A Proposal," in *CEUR Workshop Proceedings*, Belgium, 2020.

13. E. A. Christiane Bittner and J. M. Leimeister, "Why Shared Understanding Matters - Engineering a Collaboration Process for Shared Understanding to Improve Collaboration Effectiveness in Heterogeneous Teams," in *46th Hawaii International Conference on System Sciences*, 2013.
14. J. Buchan, "An Empirical Cognitive Model of the Development of Shared Understanding of Requirements," *Requirements Engineering. Communications in Computer and Information Science,* pp. 165-179, 2014.
15. P. Darch, A. Carus and M. Jirotka, "Shared understanding of end-users' requirements in e-Science projects," *5th IEEE International Conference on E-Science Workshops,* 2009.
16. C. Werner, Z. S. Li, N. Ernst and D. Damian, "The Lack of Shared Understanding of Non-Functional Requirements in Continuous Software Engineering: Accidental or Essential?," *28th IEEE International Requirements Engineering Conference,* 2020.
17. M. Arias Chaves, "La ingeniería de requerimientos y su importancia en el desarrollo de proyectos de software," *Revista InterSedes,* vol. VI, no. 10, pp. 1-13, 2005.
18. I. Sommerville, Software engineering 9th Edition, Addison Wesley Publishing Company, 2011.
19. C. E. Wanstreet and D. S. Stein, "Gender and Collaborative Knowledge Building in an Online Community of Inquiry," in *Encyclopedia of information communication technologies and adult education integration*, IGI Global, 2011, pp. 707-722.
20. P. R. Smart, D. Mott, K. Sycara, D. Braines, M. Strub and N. R. Shadbolt, "Shared understanding within military coalitions: A definition and review of research challenges," United Kingdom, 2009.
21. E. A. Christiane Bittner and J. M. Leimeister, "Why Shared Understanding Matters--Engineering a Collaboration Process for Shared Understanding to Improve Collaboration Effectiveness in Heterogeneous Teams," *System Sciences (HICSS),* vol. 46th Hawaii International Conference , pp. 106-114, 2013, January.
22. G.-J. de Vreede, R. O. Briggs and A. P. Massey, "Collaboration engineering: foundations and opportunities: editorial to the special issue on the journal of the association of information systems.," *Journal of the Association for Information Systems,* vol. 10, no. 3, p. 7, 2009.
23. M. Kleinsmann, J. Bujis and R. Valkenburg, "Understanding the complexity of knowledge integration in collaborative new product development teams: A case study.," *Journal of Engineering and Technology Management,* vol. 27, no. 1-2, pp. 20-32, 2010.
24. M. Kleinsmann and R. Valkenburg, "Barriers and enablers for creating shared understanding in co-design projects," *Design studies,* vol. 29, no. 4, pp. 369-386, 2008.
25. S. Mohammed, L. Ferzandi and K. Hamilton, "Metaphor no more: A 15-year review of the team mental model construct.," *Journal of Management,* vol. 36, no. 4, pp. 876-910, 2010.
26. G. L. Kolfschoten and G.-J. De Vreede, "The collaboration engineering approach for designing collaboration processes," in *International Conference on Collaboration and Technology*, Heidelberg, 2007.
27. G. Stahl, "Group cognition in computer-assisted collaborative learning.," *Journal of Computer Assisted Learning,* vol. 21, no. 2, pp. 79-90, 2005.
28. N. M. Webb and A. S. Palincsar, Group processes in the classroom, Prentice Hall International, 1996.

29. P. Van den Bossche, W. Gijselaers, M. Segers, G. Woltjer and P. Kirschner, "Team learning: building shared mental models," *Instructional Science,* vol. 39, no. 3, p. 283–301, 2011.

A Survey in Using Ontologies and Rules Reasoning in Access Control System

Pierre-Yves Gicquel[0000−0002−3639−6137], Jeremy Bouche-Pillon[0000−0001−6923−9915], Pascale Zarate[0000−0002−5188−1616], Nathalie Aussenac-Gilles[0000−0003−3653−3223], and Yannick Chevalier[0000−0002−8617−4209]

Institut de Recherche en Informatique de Toulouse (IRIT), Université de Toulouse, CNRS, Toulouse INP, UT3, UT1, UT2, F-31062 Toulouse, France

Abstract. In today's heavily cloud distributed-service period, access control systems are primary components to guarantee security and con- fidentiality of resource repositories. However, access control systems most widely used were deigned before this period of generalized service based infrastructure. This creates important difficulties in the maintenance of such systems. New approach in access control system, based on a formal and logical approach of security rules have been proposed that makes use of the semantic web technologies. In this paper we propose a survey of these semantic approaches together with a comparison of their respective strength depending on the considered use case.

Keywords. Ontology, Semantic-web, Access control, Rule based system.

1 Introduction

Access control refers to the regulation of access to shared resources depending on the entity requesting access, the characteristic of the resource and the privacy preferences of the resource/data owners [7]. Although access control has been an early topic of interest, starting with file permission systems, the generalization of cloud services and IoT raised several new challenges [8]. Most access-control systems that are *de facto* standards were designed before this generalization and are technically costly to maintain and scale on their current form. One fruitful approach for overtaking these issues is the use of semantic-web technologies to both formally describe and enforce access-control policies [17].

Although ontology-based access control policies are a subject of interest in many works, few systematic surveys were proposed on this topic. To the authors' best knowledge, the most recent one dates back to 2014 [14]. In this paper, we aim at giving a more recent overview taking into account emerging trends in ontology-based access control. We propose an original set of feature-oriented comparison criteria between different methods, which is designed to assist researchers in the choice of a particular method over another. Finally, we identified several

limitations that are consubstantial to all ontology-based control access systems and we discuss an extension of such systems using concepts from Multi-criteria Decision Analysis.
In the remainder of this paper, we first specify the used criteria for analysing the selected papers and we provide an overview of the system presented in these papers. In the second section, we compare these systems based on the proposed criteria. The third section examines limitations shared by all these approaches and discusses an architecture aiming at overcoming these limitations.

2 Surveyed Systems

To conduct this survey, approaches of the access control problem were selected based on their use of languages and formalism from semantic web to either implement or to design access control models. The approaches presented in these papers can be divided into two categories.
On the one hand, there are approaches that use an already existing access control models and try to implement them using semantic web methods and languages. On the other hand, there are approaches that searched to develop new access control models using semantic web techniques and languages.

2.1Approaches Using Already Existing Access Control Models

ROWLBAC. "ROWLBAC" is an implementation of the "Role-Based Access Control" (RBAC) model that uses OWL. The 4 key notions of RBAC are "roles", "actions", "subjects" and "objects". The RBAC policies give rights to certain "subjects" to do certain "actions" to certain "objects" depending on their "roles", like illustrated in the figure 1. The authorizations defined in RBAC are only binary, an action is either authorized or prohibited.

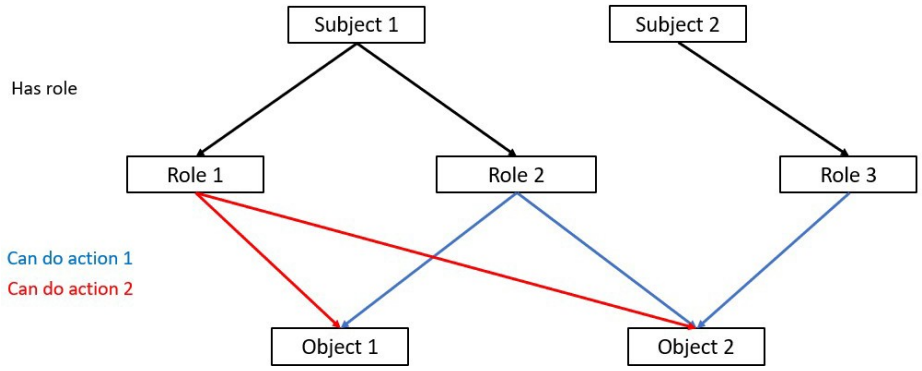


Fig. 1. Example of RBAC control

A reference in the definition of the RBAC model is the standardization works

carried out by the NIST [15, 7]. Different levels of variations of the initial RBAC model are described in [15], each one adding new constraints and capabilities to the previous one:
−*Flat RBAC*: The most basic version of RBAC, there are no relations between roles.
−*Hierarchical RBAC*: The roles are organised into a hierarchy, with a inheritance mechanism for the authorizations given to roles.
−*Constrained RBAC*: This variation introduces "Separation of Duty" (SoD) constraints. These constraints are used to specify if an action needs an authorization from more than one person to be permitted.
−*Symmetric RBAC*: Finally, this variation allows for regular revocations or reassignments of permissions in order to avoid role inflation.
The model used in [16] makes a different distinction between "possible roles" and "active roles" than in the NIST standardization [7]. Indeed, contrary to the NIST model in which a subject could only have one active role at a time, ROWLBAC allows a subject to have several active roles at a time. This modification also has a consequence on the "Separation of Duty" constraints. In ROWLBAC, there are "static SoDs" that specify the "possible roles" that a single subject can not have at the same time, and "dynamic SoDs" that indicate the roles that a single subject can not have "active" at the same time.
ROWLBAC suggests two ways of implementing RBAC principles in OWL. In the first one, each role is modeled by a class, subclass of the general class "Role" and in the second one each role is an instance of this general class [16].
Finally, the access rights are written directly in the ontology, as shown in listing 1.1. Some properties of RBAC that could not be expressed in OWL are enforced using description logic. For example, the ROWLBAC implementation uses the N3Logic language [16].
Listing 1.1: Access rule example with ROWLBAC (from [16]) Permitted Vote Action a r d f s : Class ; r d f s : sub Class Of rbac : Permitted Action ; owl : e q u i v a l e n t C l as s [a owl : Class ; owl : i n te r s e c t i o n O f (Vote [a owl : R e s t r i c t i o n ; owl : all Values From ex : A c t i v e Ci t i z e n ; owl : on Property rbac : s u b j e c t])].

ABAC. The "Attribute-Based Access Control" (ABAC) model is similar to RBAC. But instead of just the notion of "roles", any kind of attribute that allows to describe the access context can be modeled. It may be identities /roles, devices, actions, types of data, location, time, ... For instance, ABAC allows to model the whole context of data access in a smart home environment [6] (cf. figure 2).

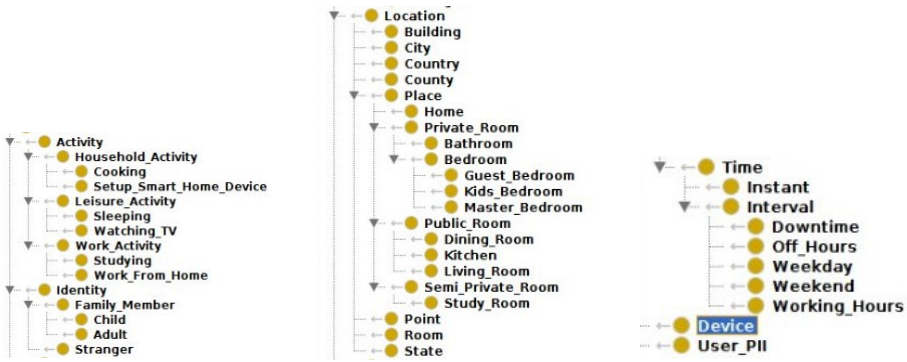


Fig. 2. Example of ABAC attributes (from [6])

One implementation of ABAC using OWL was mentioned at the end of [16], still using N3Logic to define rules. An other implementation proposed in [6] uses SWRL to write the access rules. The listing 1.2 gives an example of a rule for ABAC.

Listing 1.2: Access rule example in SWRL for ABAC model (from [6])

```
sme : CrashReports ( ?requestedData )                    ∧sme :
familyMemberInRoom ( ?aMember )                          ∧sme :
ageOf ( ?aMember , ?someAge )                          ∧swrlb : less
Than ( ?someAge , "18")                                  ∧
=> accessDenied ( ?requestedData )
```

OrBAC. "OrBAC" stands for "Organization-Based Access Control" model. It can be seen as a deepening of RBAC concepts which allows to abstract more than just the "subject" of a permission. Thus, the OrBAC model allows to describe the structure of an organization through three abstract concepts: "roles", "views" and "activities", abstracting respectively "subjects", "objects" and "actions" [9]. Here the idea of "roles" is the same as in RBAC, "views" regroup objects having a common property and "activities" model actions on "views". These abstractions are illustrated in figure 3.

Multi-OrBAC. "Multi-OrBAC" is an extension of OrBAC that allows to consider a complex network composed of many organizations. For example, a context it can be applied to would be to ensure the security of medical data processing in

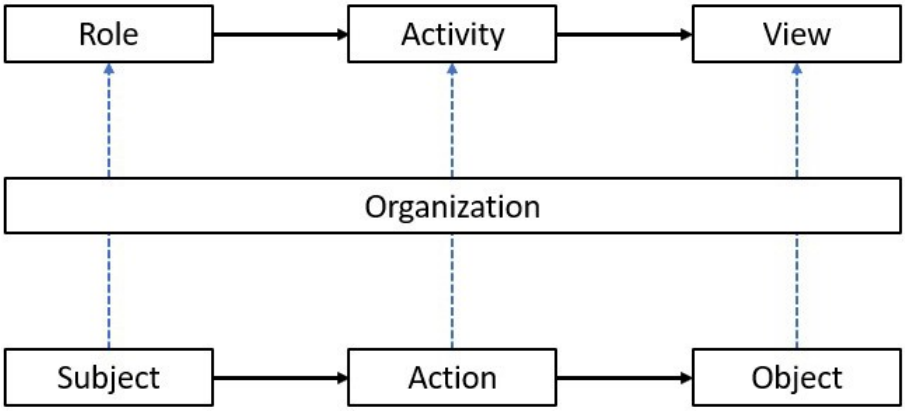


Fig. 3. OrBAC abstraction

the hospital environment [10]. Compared to the original OrBAC, Multi-OrBAC adds a level of abstraction thanks to the notion of "organization", and with it, the necessity to link the notions of "roles", "views", "actions" and "context" to the organisation they are related to. It also allows to define a hierarchy in the considered organizations as well as to describe inheritance for permissions depending on this hierarchy. Finally, the rules in Multi-OrBAC follow a similar principle than in ABAC. An example is given in listing 1.3.

Listing 1.3: Access rule example for Multi-OrBAC model (from [1])

```
Permission(Physician-in-\Org_A, Reading-in-Org_B, MedicalRecord,dis
aster) ∧ Play(Bob, Physician-in-Org_A) ∧
Correspond_to(f1.xml, MedicalRecord-in-Org_B) ∧
Belong_to(Read-xml(), Reading-in-Org_B) ∧
Is-true(disaster-in-Org_B)
-> Is_permitted(Bob, Read-Xml-File(), f1.xml)
```

2.2 Approaches Defining New Access Control Models

OBAC. The "Ontology-Based Access Control" (OBAC) model aims at providing access control mechanisms for FAIR datasets. Although FAIR principles do not necessarily require semantic web technologies to describe metadata, the use of RDF and SPARQL endpoints has become standard *de facto*. Access-control policies can be required to manage FAIR data in sensitive context, such as criminal case reports, as mentioned in [4]. These authors propose to tackle this problem by making use of existing metadata linked to the dataset under consideration. OBAC also uses the notion of roles but, unlike ROWLBAC, roles do not directly appear in rules. Instead, each role is associated to a graph projection (with SPARQL Construct, see example in listing 1.4) in order to restrain access to a subset of the original graph depending on each user's credential.

Listing 1.4: Example of SPARQL Construct query used in OBAC (from [4])

```
?item dwo:hasTitle ?object .IF
?item rdf:type dwo:Item .
?item dwo:hasDestination dwd:USA .
?item dwo:hasOrigin dwd:USA .
?item dwo:hasTopic ?topic .
?topic rdf:type ?topicType .
?topicType rdfs:subClassOf+ dwo:Drugs .
```

KAoS. Although KAos is not *stricto sensu* an access control system, we decided to include it in this paper given its major influence. KAos was initially aiming at providing an agent-based framework making use of semantics in service description to support collaboration between different systems. Multi-agent systems providing an agnostic for message transmission before the generalization of REST protocol. Although authors present KAos, in its first version, [3] as an agent-based framework, following works on KAos have left this option behind. One aspect worth noticing about KAos first version is that it was not semantically grounded despite the authors' claims. The semantics was limited to sets of action-verbs that agents were able to execute.
Nevertheless, through multiple and regular iterations (see for example [18] and [19]), KAos has become agnostic with respect to message brokers and has focused on defining an OWL representation of policies. These policies go beyond access control, for example they can be used to describe mandatory actions applying whenever a given policy fails to apply. KAos is still able to express access-control policies but its scope is larger (cf. figure 4).

3 Feature Based Comparison of Selected Systems

3.1Comparison Criteria

We have selected comparison criteria aiming to choose one of the approaches to implement an ontology-based access control using adapted to the specific needs of end users. These criteria are therefore not orthogonal but express some trade-off between different requirements.
The first criterion is the **expressiveness** of the access control rules defined by an approach. This expressiveness correlates with the granularity and the specificity of the allowed access control rules. Granularity refers to the maximum precision in selecting resources which are under control and specificity refers to the precision in selecting types of users for which the rules will apply.
The second criterion is the **ability to generalize** the approaches to different application domains. The application domain indeed strongly impacts the semantic representation of resources and users. Therefore it is crucial for a fruitful implementation to consider whether and how easily an approach can be applied to a new application domain.

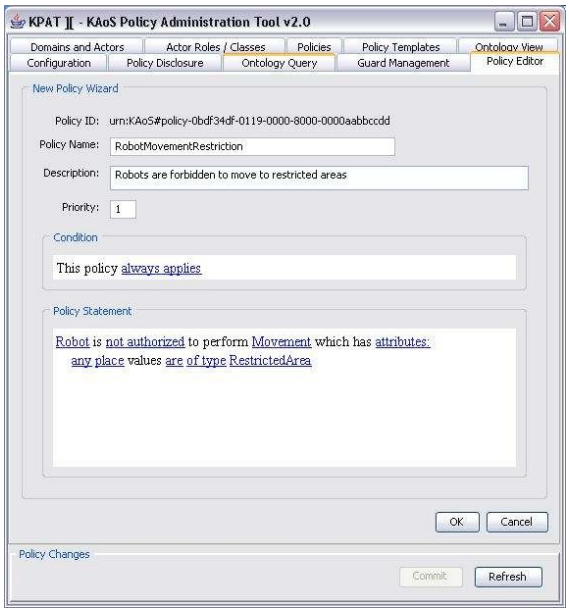


Fig. 4. Example of KAoS policy (from [19])

The third and last criterion is the **intelligibility** of the access control rule formalism. In order to maintain the system and to facilitate user acceptance, it is desirable that the formalism be understandable even with little or no prior knowledge of the semantic web. The conceptual choices have a direct impact on this possibility. Such a criterion also provides information on the auditability of the system, i.e. the possibility to check more or less automatically if the system meets certain requirements.

3.2 Comparison of approaches

Expressiveness of the rules The ROWLBAC approach, based on RBAC is expressive in terms of specificity, allowing to define arbitrary complex roles hierarchy. These roles are then associated to actions that can be either permitted or prohibited. The approach, however, provides very few ways to model actions in an expressive way, for example to organize actions hierarchically. Moreover, the approach does not allow to represent obligation, which is an important concept distinct from authorization and interdiction.

The OWL implementation of ABAC in [6] reflects the greater expressiveness of the ABAC access-control protocol over RBAC [5]. In ABAC, the authorization policies are defined by rules which depend on the attributes of the users performing the request, the considered resource and the environment in which the request is made. The notion of context of a request is introduced by taking the environment into account. [6] propose an elegant and expressive model of

physical contexts by using a transitive object property which allows to represent the physical context with different degrees of granularity. For instance, a user located in the bathroom can be considered as being either in the house or in a private room : house $\overset{\text{partOf}}{\leftarrow}$ private room $\overset{\text{partOf}}{\leftarrow}$ bathroom.
It is difficult to characterize the expressiveness of OBAC since it relies exclusively on the expressiveness of the considered FAIR dataset. Nevertheless it should be noted that the use of SPARQL Construct offers an innovative and potentially powerful way to address the issue of access control.
Multi-OrBAC is an extension of the OrBAC approach that also corrects some modeling errors that went unnoticed in OrBAC [2]. Multi-OrBAC aims at providing a well-founded semantic framework for access control across several organizations. It is together with KAOS [19] the most expressive approach of all. Not only are the concepts of roles and roles hierarchy in organization defined (as in ROWLBAC), but Multi-OrBAC also provides ways to represent resources and activities at an organization level. This does not preclude several organizations to share similar resources and activities. On the contrary, it allows to establish a correspondence between these similar activities. Furthermore, Multi-OrBAC offers the three classic deontic modalities: prohibition, obligation and permission together with a fourth (non deontic) modality: recommendation. It is thus possible to design very expressive rule sets fitting the complexity of multi-organizational access management.
Finally, KAos [**?**] is undoubtedly the most expressive approach, mainly because the scope of this approach goes beyond access-control. KAoS reaches the multi-OrBAC level of expressiveness and extends it in several ways. The most innovative advances are first the ability to define priority in rule application whenever several rules are triggered simultaneously. This allows to precisely control in which order the rules will be applied in complex situations. Second, KAoS provides the ability to take historical context of rule execution to define meta-rules that will use that context to modify the rule set and the rule execution priority.

Generalisability to other application areas The RBAC model implemented in ROWLBAC has been recognized by a NIST study as being adapted to the majority of business needs [7]. Besides, because of its simplicity, it can easily be generalized to other contexts. For example, in [16], it is used to specify authorized actions of people in a society according to their status.
In the ABAC model, the notion of "attribute" is a very broad concept, without any restriction, which makes it possible to model devices, users, actions, as well as dates and places as attributes. For instance, it can be used to model the different rooms of a house, the IoT devices in it, the types of data that can be accessed, the possible roles of the people accessing data and even differentiate time slots [6]. So, like for ROWLBAC, the ABAC model allows to describe the context of almost any situation. Because of its higher expressiveness, it can even be used in more situations than ROWLBAC.
Because OBAC is not based on a well-defined model, it is quite difficult to evaluate the ability to generalize this approach [4]. In the end, the use of OBAC

in a system depends on the format of the used data. OrBAC and Multi-OrBAC, although very useful to easily and efficiently model the structure of a complex organization, cannot be generalized as much as the previous works [2]. The three central concepts of OrBAC, namely roles, views and activities, restrict its use to situations where these notions are relevant.
The KAos approach aims at generality by design, which makes it applica- ble to many new situations. However, it suffers from two particularly strong limitations. The first limitation relates to the choice of compiling rules in an intermediate, non-standard format, to improve performances. This choice makes it difficult to guarantee that any rule expressed in OWL could be actually compiled and the authors provides no theoretical proof giving such guarantee. The second limitation is the fact that some of KAoS features are not expressible in OWL and require to rely on the correctness of the implementation. This is for example the case for assigning priority to rules. This limitations make it difficult to reuse KAoS approach, instead, one would have to use the whole KAoS framework without much possibility of modifying it (the cost of modifying a source framework of this complexity being prohibitive).

Intelligibility for the end user The formal language used in ROWLBAC to express access rules is N3Logic [16], a language that allows the expression of human-understandable rules. Rules in ROWLBAC can enforce some desired properties that could not be expressed in the model, for instance "separation of duty" rules. Unfortunately, these kinds of rules can be complex and hard to understand. Two ways to model roles are presented in [16]. The first one allows the use of Description Logic reasoning to infer the hierarchy between the different roles, but the second one does not. As a consequence, rules in N3Logic must be written to enforce the inheritance of roles. These rules are short and simple. Finally, the last rules are defined to enforce the access control itself. Because of the simplicity of the RBAC model that ROWLBAC implements, these rules are short and they explicitly show the logic behind the authorization of an action.
Because of its greater expressiveness, ABAC [6] is less intelligible than ROWLBAC. Indeed, the contexts that can be described in ABAC are more complex and take into account way more aspects than just roles. And having more com- plex contexts means that the access rules associated to these contexts are likely to be more complex too.
With OBAC, the rules are defined using relations described directly in the data. The examples given in [4] are quite explicit. In order to understand the rules, the user only needs to have an idea of how the data are organized. In the end, with OBAC, the intelligibility of the rules depends on the intelligibility of the data themselves.
The OrBAC model is constrained by its key concepts: roles, views and activities [2]. However, unlike ROWLBAC which is only constrained by the notion of role, understanding the concepts behind the ideas of OrBAC is not trivial and requires some knowledge of the model, this complexity of which makes the access rules harder to understand. Multi-OrBAC even increases this complexity by adding the notion of organization, and with it, the ideas of "Role in an Organization", "View in an Organization", "Activity in an

Organization" and "Context in an organization" [2]. All these elements must be specified in the access rules. In the end, the rules in Multi-OrBAC, although allowing a precise control of data access, are hard to understand without prior knowledge of the model and the concepts behind it.
KAoS framework, with a user interface for rule expression in a syntax close to natural language, certainly is the most intelligible one for end users.

Overview The analysis previously carried out is summarized in table 1.

Table 1. Summary of the analysis

Name	Expressiveness	Generalisability	Intelligibility
RBAC	-	+	-
ABAC	++	++	+
OrBAC	+	-	–
Multi-OrBAC	++	-	-
OBAC	=	=	=
KAoS	++	–	+

4 Discussion and future work

The generalization of open-data is undoubtedly an improvement. This may explain why the issue of controlling access to accessible data-sets has only received little interest in the scientific community. Nevertheless, with the rise of e-governments and the increasing need for collaboration between Law Enforcement Agencies to fight crime at an international level (see [12] for an example on tax evasion) by using large amounts of data, the need for differentiated and semantic access control over data is likely to increase.
Using semantic technologies to ensure access control has undeniable advantages. First of all, the theoretical foundation of the semantic web are very solid and the technology itself has been in use for decades, which offers strong guarantee of reliability for a domain as sensitive as access control. Furthermore, semantic web and ontologies are strongly related to knowledge representation and do not use black box algorithms, so each decision of authorization or prohibition to access a resource by an entity is explicable.
However, these advantages come with several drawbacks that restrain the practical use of semantic technologies for access control. The most important limitation is the cost, both in time and resources, of tailoring an approach to

a particular use-case. Ontological engineering is known to require a substantial amount of conceptualization time before any implementation. Meanwhile, organizations such as national Law Enforcement Agencies are subject to frequent regulatory changes. In a pure semantic approach, each regulatory change would require non-negligible amount of time to update the system properly.
These considerations have led us to define a new approach based on a distinction between slowly evolving primary source of access control rules (i.e organic laws, international treaties) and fast evolving secondary source of access control rules (i.e organizational level regulation). We aim to develop a new approach that will use semantic web technologies for the modeling of the primary sources of access control rules, which applies to each organization and will use a reinforcement learning decision support system for the secondary sources of access control. The latter approach was experienced within the context of supporting end users in authorization management for the Android platform [13].
The Decision Support System, based on a Multicriteria Decision Analysis approach, will therefore be used to handle fast change in organizational policies relative to access control. Its scope, that is, the set of resources for which it will provide authorization, prohibition or permission, will be a subset of the resources managed by the semantic access control system. Indeed, the organizational policies are lower in the hierarchy of norms [11] and must therefore comply with the higher norms, for example a national police regulation must comply with organic laws of the nation the police belongs to.
What we aim is therefore to reuse the works on ontological-access control to model high-level (i.e. organic laws) access control regulation. Then to define an approach based on Multicriteria Decision Analysis to provide a finer and easier to modify access control at institution level.

References

1. Abou El Kalam, A., Deswarte, Y.: Multi-orbac: A new access control model for distributed, heterogeneous and collaborative systems. In: Proceedings of the IEEE Symposium on Systems and Information Security (2006)
2. Abou El Kalam, A., Deswarte, Y.: Multi-orbac: A new access control model for distributed, heterogeneous and collaborative systems. In: Proceedings of the IEEE Symposium on Systems and Information Security (2006)
3. Bradshaw, J., Dutfield, S., Benoit, P., Woolley, J.: Kaos: Toward an industrial-strength open agent architecture. Software Agents (01 1997)
4. Brewster,C., Nouwt, B., Raaijmakers, S., Verhoosel, J.:Ontology- based Access Control for FAIR Data. Data Intelligence **2**(1- 2), 66–77 (jan2020). https://doi.org/10.1162/DINT_A_00029, https://direct.mit.edu/dint/article/2/1-2/66/9993/Ontology-based-Access- Control-for-FAIR-Data
5. Das, S., Mitra, B., Atluri, V., Vaidya, J., Sural, S.: Policy Engineering in RBAC and ABAC, pp. 24–54. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-04834-1_2
6. Dutta, S., Chukkapalli, S.S.L., Sulgekar, M., Krithivasan, S., Das, P.K., Joshi, A.:

Context sensitive access control in smart home environments. In: 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). pp. 35–41 (2020). https://doi.org/10.1109/BigDataSecurity-HPSC-IDS49724.2020.00018
7. Ferraiolo, D., Richard, D.: Role-based access controls. In: Proceedings of the 15th NIST-NSA National Computer Security Conference, Baltimore, Maryland (1992)
8. Imran-Daud, M.: Ontology-based Access Control in Open Scenarios: Applications to Social Networks and the Cloud. Ph.D. thesis, Universitat Rovira i Virgili. (2016)
9. Kalam, A., Baida, R., Balbiani, P., Benferhat, S., Cuppens, F., Deswarte, Y., Miege, A., Saurel, C., Trouessin, G.: Organization based access con- trol. In: Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks. pp. 120–131 (2003). https://doi.org/10.1109/POLICY.2003.1206966
10. Kalam, A., Deswarte, Y.: Multi-orbac: un modèle de contrôle d'accès pour les systèmes multi-organisationnels (06 2006)
11. Kelsen, H.: Pure Theory of Law. University of California Press (1967), https://books.google.fr/books?id=8N5S6BjX3RAC
12. Khlif, H., Amara, I.: Political connections, corruption and tax evasion: a cross- country investigation. Journal of Financial Crime (2019)
13. Oglaza, A., Zaraté, P., Laborde, R.: KAPUER: A Decision Support Sys- tem for Privacy Policies Specification. Annals of Data Science **1**(3), 369– 391 (Dec 2014). https://doi.org/10.1007/s40745-014-0027-3, https://hal.archives-ouvertes.fr/hal-01387752
14. Papagiannakopoulou, E.I., Koukovini, M.N., Lioudakis, G.V., Dellas, N.L., Kaklamani, D.I., Venieris, L.S.: Leveraging semantic web technologies for access control. Emerging Trends in ICT Security pp. 493–506 (1 2014). https://doi.org/10.1016/B978-0-12-411474-6.00030-X
15. Sandhu, R., Ferraiolo, D., Kuhn, R.: The nist model for role-based access control. In: Proceedings of the fifth ACM workshop on Role-based access control - RBAC '00. pp. 47–63. ACM Press (2000). https://doi.org/10.1145/344287.344301
16. for Computing Machinery. Special Interest Group on Security, A.A.: SACMAT '08: proceedings of the 13th ACM Symposium on Access Control Models and Technologies: Estes Park, Colorado, USA, June 11-13, 2008. Association for Computing Machinery (2008)
17. Senthilkumar, S., Viswanatham, M.: Survey on data access control techniques in cloud computing. In: Proceedings of ICETETS 2016 (2016).
18. Uszok, A., Bradshaw, J.M., Lott, J., Breedy, M., Bunch, L., Feltovich, P., John- son, M., Jung, H.: New developments in ontology-based policy management: Increasing the practicality and comprehensiveness of kaos. In: 2008 IEEE Work- shop on Policies for Distributed Systems and Networks. pp. 145–152 (2008). https://doi.org/10.1109/POLICY.2008.47
19. Uszok, A., Bradshaw, J.M., Lott, J., Johnson, M., Breedy, M., Vignati, M., Whit- taker, K., Jakubowski, K., Bowcock, J., Apgard, D.: Toward a flexible ontology- based policy approach for network operations using the kaos framework. In: 2011-MILCOM 2011 Military Communications Conference. pp. 1108–1114. IEEE (2011).

Dynamic Spatial Task Generation for Collaborative Location-based Collecting Systems Coverage Objectives

María Dalponte Ayastuy[1,2][0000−0002−1412−5694], Diego Torres[1,2][0000−0001−7533−0133], and Bruno J.Lattanzio [1] [0000−0001−5380−0915]

[1] Depto CyT, Universidad Nacional de Quilmes Roque Saenz Peña 352, Bernal, Buenos Aires,Argentina.
{mdalponte,brunoj.lattanzio}@unq.edu.ar
[2] LIFIA,CICPBA-Facultad de Informática, Universidad Nacional de La Plata50 y 120, La Plata, Buenos Aires, Argentina.
diego.torres@lifia.info.unlp.edu.ar

Abstract. Collaborative location-based collecting systems (CLCS) are a particular case of collaborative systems where a community of users collaboratively collect geo-referenced data. Each CLCS sets its territory coverage objectives, commonly defined as to guarantee that all the affected territory is surveyed with a particular coverage criterium. This paper presents a three-step pipeline to recommend the subareas that require observations dynamically. The first step generates a disjoint and adjacent set of areas -a mesh- covering the sampling territory. The second step sets a priority and coverage objective for each area. Finally, the third step considers the project's objectives and the area coverage situation to recommend the areas that need surveys. The output of this last step is an input for a user-task distribution process where the user's profile is taken into account. Moreover, an example of meshing strategy and task generation is proposed.

Keywords. Collaborative Location-based Collecting Systems, Meshing, Decision-making, Spatial Crowdsourcing.

1 Introduction

Collaborative Location-based Collecting Systems (CLCS) are collaborative systems where the community of users collects data associated with their location normally by using a mobile application [3]. CLCS can be weighed as a supporting technology of some citizen science projects, such as the AppEar project[1], GeoVin [2] or iNaturalist [14]. Citizen science projects encourage and support the contributions of volunteers to the advancement of scientific research. Some location-based activities need to survey scientific data associated with a location and a timestamp, consolidating these contributions as tuples with the structure:

$$< lat{-}long,\ timestamp,\ sampleData >$$

As an example, a sampleData value for AppEar project includes ecological information about rivers, lakes and estuaries; GeoVin sampleData describes the 'barber bug' insect (Triatoma infestans) presence; iNaturalist sampleData is about biodiversity observations around the globe.

Each CLCS sets its territory coverage objectives, commonly defined as to guarantee that all the affected territory is surveyed with particular coverage criteria [3]. For instance, the coverage criteria can be a sample every 100 meters. Setting coverage criteria implies dealing with the problem of organizing the division of observation tasks so that the project's objectives are met. Specifically, it requires (i) the segmentation of the territory into smaller areas and (ii) assigning coverage priorities to each area considering the project objectives. And so, if necessary, repeat (i) and (ii).

Even though the spatial modeling and segmentation problem is widely approached in geographical information and location-aware systems [9, 12, 8, 18,6, 5], and the growing number of studies in spatial tasks assignment in a wide range of fields [10, 11, 15–17, 7] , the subject of relating tasks to spatial segments considering project's objectives is largely absent.

This article presents an approach for decision-making assistance to dynamically recommend the tasks for those subareas that require observations based on the coverage objectives and a geographical area of scope of the CLCS. Such a tool is much needed in CLCS as well in the industry. As the survey tasks are completed, the recommendation is dynamically executed to find those areas that have not been completed. Also, it is possible to trigger a new segmentation if a particular condition on the system's global status is achieved.

The recommendation is made through a three steps pipeline where the first step generates a disjoint and adjacent set of areas -a mesh- that covers the sampling territory. The second step sets a priority and objectives for each area, and it can be done manually or automatically through a rule set. Finally, the third step considers the project's objectives and the area coverage situation to recommend the areas that need to be surveyed. The output of this last step is an input for a user-task distribution process where the user's profile is taken into account. It is essential to notice that this approach generates a set of tasks considering, on the one hand, what is needed to achieve the project objectives and, on the other hand, the status of the project. However, nothing is addressed concerning the assignment of these tasks to the people who have to solve them. This assignment proposes a challenge in the adaptability research area to analyze the tailoring of tasks based on each user's preferences, characteristics, and behavior. Although this is an external process, the result of the task assignment in terms of how many were completed is essential feedback to update the CLCS's global state and be able to repeat the generation of tasks. In addition, this work presents a possible mesh computation strategy suitable for a cold start and a task generation strategy.

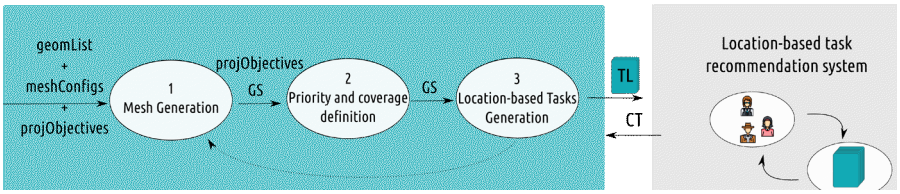
This article is organized as follows. In Section 2 the related work is presented. In Section 3 the pipeline steps are detailed. Finally, Section 4 shares discussions and future work.

2 Related work

Spatial crowdsourcing is the process of crowdsourcing a set of spatial tasks (i.e. tasks related to a location) to a set of users, which requires the users to perform the spatial tasks by physically traveling to those locations[11]. Location-based task assignment needs to assess the available sensing resources to meet the objectives of the CLCS. The criteria for optimization of task assignment include sensing costs, coverage of areas of interest, quality, and redundancy of sampled data. The approach in [15] proposed a coverage-based task assessment that finds the least costly subset of participants to achieve the coverage goal. The work in [16] also proposed a coverage-based task assignment method for assigning viewpoints to a group of moving participants. Similarly, [17] focuses on one class of spatial crowdsourcing, in which the users send their locations to the server. After that, the server assigns the tasks in proximity to the user's location to every user, intending to maximize the overall number of assigned tasks. Notice that the tasks are defined by a geolocated point in all these approaches but not associated with a sampling geolocated area.
Regarding the segmentation into sampling areas, the discipline of geographic information systems (GIS) defines and uses different tessellated models to represent information about the earth [8]. Sometimes these models have the objective to build a hierarchical discretization of a high definition raster image or to support a vector representation of terrain surface, such as the Voronoi regions [9] or the Delaunay triangulation [18]. Articles that address the meshing problem by calculating the Voronoi regions from a set of centroids were found, as is the case of Fleischman et al. [6] and Du and Gunzberg [5]. These approaches can be helpful when, as was mentioned in Section 1, a new segmentation of the territory is needed considering the accumulated sampling activity.

3 Approach

CLCS spatial decision-making needs to optimize users' work by combining the defined objectives for the CLCS with the completed tasks in the territory. As was mentioned, one possible way to present the objectives is in terms of the number of samples per subarea and setting criteria for building a set of areas. Furthermore, these objectives can be complemented by prioritizing specific subareas.
This article proposes a pipeline to support a located-based task generation considering the project's objectives and the system's global status, as it is shown in Figure 1. The first step aims at generating a Mesh, given a list of geolocated geometries and some initial mesh configurations.
As a second step, the coverage and priority for each area must be defined, and it can be done manually or automatically through the application of a ruleset. These two configurations are essential at different times of the process: coverage setting is needed for the task generation step, and the area priority definition is a requirement for the location-based tasks recommendation system (see the gray box in Figure 1).



The third step feeds on the project objectives and the system's global status to generate a list of spatial tasks to be distributed to users. After that, the recommendation of tasks (and their subsequent completion) offers an updated system's global status as feedback for this task generation (see CT -Completed Tasks- parameter in Figure 1).
Moreover, the mesh can be thought of as a dynamically adapted mesh that is recalculated based on the newly completed location-based tasks that represent the feedback to step 3 and can trigger feedback to step 1.
The following subsections introduce a few preliminary definitions and then give details about the steps of the pipeline, with an example strategy in each case.

3.1Preliminary definitions

Two methods are frequently used to represent geographic phenomena in ways that can be encoded in spatial databases, called raster and vector methods. Both can be used to encode continuous fields and discrete objects, but there is a strong association between raster and continuous fields and between vector and discrete objects in practice. One of the most common forms of raster data comes from remote-sensing satellites, which capture information as high-definition images. In a vector representation, all lines are represented as points connected by a straight line, and areas are captured as a series of points or vertices connected by straight lines, called polygons. Lines are represented in the same way, and the term polyline (or multiline) has been coined to describe a curved line represented by a series of straight segments connecting vertices [12].

Definition 1 (Point). *It is a tuple (latitude, longitude) of geographic coordinates. Negative values in latitude refer to points in the southern hemisphere, and negative values in longitude refer to points located west of the Greenwich meridian.*
For instance, a point is pair with latitude: -24,5073190, and longitude: - 68,3958578.

Definition 2 (Multiline). *A multiline object is a sequence of more than two points*.

Definition 3 (Polygon). *A polygon is a particular case of a multiline element, where the first point is also the last one, making a closed geometry.*

3.2 Mesh Generation

As presented in the introduction, CLCS needs to relate each obtained sample with an area of the mesh. Depending on each project's particular characteristics, the area mesh must meet certain conditions on shape and granularity. However, in all cases, the mesh is a set of adjacent and disjoint cells organized following the shape of a given geometry.

Table 1. Mesh generation interface.

Input Parameters			Output
`geomList`	Set of geometries representing the target territory	`gs`	initial system global status $GS = \{< a, rt, ct, w >\}$ where w is undefined
`meshConfig`	List of key/value tuples of configurations for mesh generation		
`projObjectives`	List of key/value tuples of project objectives definition		

For this reason, in this first step, it is necessary to delimit the territory and be ableto configure the desired characteristics of the mesh. Table 1 details the input and output value types for the Mesh Generation step. The geolocated territory is definedby the geomList input parameter, while the characteristics are defined by the meshConfig input parameter. The geomList parameter can contain a single element that represents the total area of scope of the project or a set of geometries thatspatially defines the territory. However, the details of this parameter depend on the particular meshing strategy.

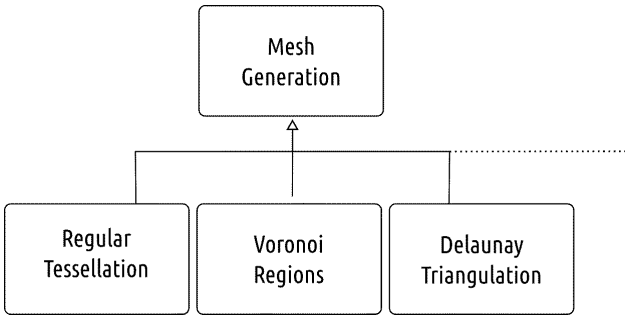


Fig. 2. Mesh Generation strategies

The output of the Mesh Generation step is the initial system global status as is defined below:

Definition 4 (System global status). *The system global status is a list of tuples:*
GS = {< a, rt, ct, w >}

where a is a sampling area, rt is an integer number representing the required number of sampling tasks, ct is an integer number representing number of completed

sampling tasks in area a, and w is the assigned priority weight to area a.

As is shown in Figure 2, in this work, several strategies are proposed, and they can be further extended. The Regular Mesh Generator (detailed in the following sub- section) is based only on a geometry (or a list of geometries) that defines the territory and the map of meshing parameters. This strategy is useful to tackle the cold-start meshing. Other strategies approach the mesh generation based on the territory spatial definition and the existence of other intermediate points located in the territory. Examples of these are the Voronoi regions [9], or Delaunay triangulations [18],among others.
Moreover, the granularity level of meshes can be thought of as dynamic and can be supported as an automatic process triggered in step 3. As an example, this can be used to divide areas that fulfill certain conditions, updating the system's status in the areas involved. Similarly, the areas can be merged to achieve larger areas.

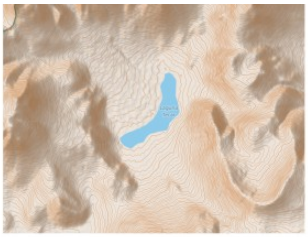
Regular Mesh Generation (cold-start) Some scenarios require the sampling activity to be done in the proximity of a geographic element (a Point, multiline or polygon) without other points as a reference. This situation is associated with a cold-start mesh, and the geomList parameter is populated with these elements geographic, aiming at building a mesh following the shape of these geometries. For example, consider AppEar project, which needs to survey the ecological situation at the shores of rivers, lakes, and estuaries. In this case, geomList is made up of spatial geometries that represent the rivers and lakes.
This particular meshing strategy where the generated areas are organized in a regular shape grid following the shape of a given geometry is called Regular Mesh. Inother words, the upper limit of the mosaic is the geometry's limit, the lines that separate the cells' rows are parallel to that limit, and the number of rows and the dimension of its cells are set through the configuration map.
To detail the example scenario, suppose that sampling tasks need to be completed on the shore of a lake, that a ring parallel delimits the shore to the shoreline, spaced 100 meters apart, and that at least one sample is needed every 50 meters. This requirement would need an initial mesh with one row (or ring) of 50 meters width cells, like the one depicted in Figure 3 (b), and the meshConfig parameter is:

$$meshConfig = \{< cellWidth, 50 >, < nrows, 1 >, < gridHeigth, 50 >\}$$

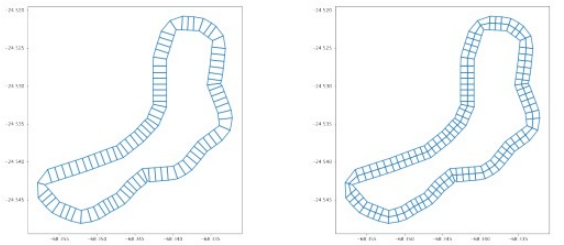
Another situation could need also a second sample of 70 meters from the shore, and in this case a different mesh would be needed, with two rings or rows around the lake. This is graphically explained in Figure 3 (c). In this case the meshConfig parameter is:

$$meshConfig = \{< cellWidth, 50 >, < nrows, 2 >, < gridHeigth, 100 >\}$$

Similarly, the different sampling requirements can be applied to the geometry of a river, as is depicted in Figure 4.

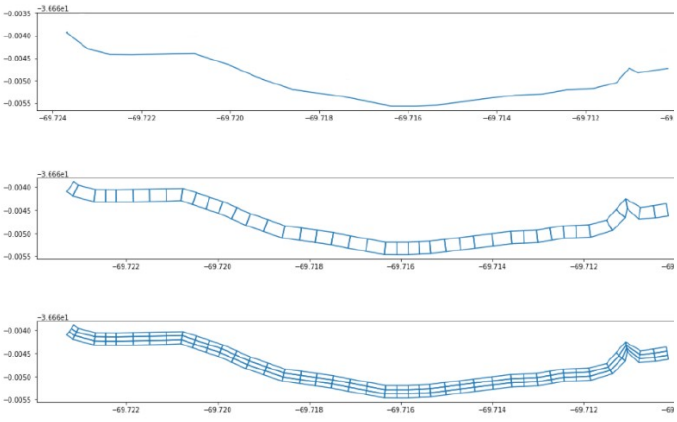


(a) Lake polygon



(b) Adaptive Mesh with one row (c) Adaptive Mesh with two ro

Fig. 3. Lake polygon and Regular Mesh examples



(a) Original river geometry

(b) Mesh with one row

(c) Mesh with three rows

Fig. 4. Adaptive mesh for a river

The regular mesh generation algorithm is presented in Listing 1.1 and builds a mesh for a given spatial geometry, a cell width, a total grid height, and a row number. Notice that this geometry can be a multiline (definition 2) or a polygon (definition 3), but in a general way, geometries are made up of sequences of segments, where each

segment is a pair of connected points.
To generate the polygon's grid, this algorithm builds a grid for each segment, and after that, it connects the grids of consecutive segments. In each segment union, one of two possible situations is faced: either the segments form a convex angle (that is, there is an uncovered space between the grids to be connected), or they form a concave angle (that is, these grids overlap). Therefore, an additional process must be carried out that performs extrapolation or interpolation,respectively.

Listing 1.1. GridConnector algorithm

```
generate_regular_mesh(geom, cellWidth, gridHeight, nrows):
  polygons = []
  For i in geom.index:
    segment = LineString(geom[i:i+1])
    pSet = gridify_segment(segment, cellWidth, gridHeigth,
    nrows)
    polygons = fix_append(polygons,pSet, nrows)
  if ((geom.type == 'MultiPolygon') | (geom.type == 'Polygon'
  )):
    polygons = fix_last_and_first_point(polygons, rows)
  res = gpd.GeoDataFrame({"geometry" : polygons})
  res['cid'] = res.index
  return res
```

Although this mesh generation is handy when the mesh must follow the shape of a geolocated geometry, it can be applied as a cold start strategy in other domains.

3.3 Area Priority and Coverage definition

A common trait in CLCS is to present the project's objectives in terms of coverage in a set of areas, as an integer value associated to each one. In this work, the coverage need is modeled with the *rt* value in the system global status (defined in 4). Furthermore, special areas can also be highlighted using polygons or points of interest (POIs), where another criterion is applied for setting the required number of tasks *rt*.

Table 2. Area priority definition interface

Input Parameters		Output	
`GS`	System global status detailed by area	`GS`	System global status
`projObjectives`	List of key/value tuples of project objectives definition		

As was introduced, this step can be done manually or automatically employing a rule-based system and needs to be carried out after the subareas are defined (step 1 is executed). In any case, this is useful for generating tasks if the strategy applied for this considers their priorities. The first time this step is executed, the input system's global state has no defined weights, but after regeneration of the mesh, it may be necessary to update some of those weights.

Table 3. Example project objectives definition

key	value
Default samples per area	20
Default area weight	5
Special area weight	7
Points of interest (POIs)	$\{< -24,5073190, -68,3958578 >\}$

Considering a rule-based system, a possible projObjectives map is described in Table 3. This map determines that all areas must have a required sampling task of 20 samples (rt = 20), and the same weight value (w = 5), except for those areas that contain any of the points of interest defined in the POIs property, in which case w = 7.

3.4Task Generation

In this step, the set of sampling tasks that are necessary to achieve the project objective is built. Specifically, is a set of tuples

$$TL = \{< a, n >\}$$

where a is a sampling area, n is an integer number representing the pending sample work.

Different strategies can be applied, considering that task completion is not directly related to the task list as users do not always complete the assigned task. For example, this step could estimate the number of redundant tasks based on the feedback through which completed tasks are reported.

Table 4. Task generation interface

Input Parameters		Output	
`GS`	System global status	`TL`	Sampling task list

In the different approaches to solving step 2, the first time this step is executed, it is based only on the system's global state. However, in the following iterations, feedback from step 3 is considered: how the community of users completed the required tasks. This feedback is modeled in a list of tuples

$$CT = < a, m >$$

where a is the surveyed area and m the completed task count that informs the finished

sampling tasks in that area.
In the next section, a particular strategy for task generation is presented.

Offset task generation strategy The offset task generation approach is a particular task generation strategy that focuses on the pipeline feedback from an external task recommendation system. The value m is not necessarily less than or equal to the value n that was generated for that area in the previous cycle as an output of this step. Since the users community might perform fewer sampling tasks than those requested for that area, more tasks may be carried out since there are geographical locations that can attract more visitors.
The difference (offset) between m and n is calculated to measure the response to the lastly generated task list. If δ=m-n is positive, it is an over-sampled area, but if δ=m-n is negative, then an under-sampled area is present. In the first case, this area is already covered and is separated from the output. In the second case, redundancy is applied to the calculation of the offset, which is a scale factor (f) to take into account that the area in question is rarely visited. In addition, it is helpful to establish a threshold (t) that allows establishing which areas need more visibility.
Therefore, the task list update is done with the algorithm described in Listing 1.2.

```
updateArea(tl,areaId,n,m):
offset = m-n
if(offset>=0):
  tl.remove(areaId)
else:
  abs = absolute(offset)
  if (abs<t):
    tl.update(areaId,abs)
  else
    tl.update(areaId,abs*factor)
```

Listing 1.2. Task list update

Suppose the situation where for a certain area a 1 15 samples were requested but only 12 were solved, then δ is -3. On the other hand, if another area a 2 had 15 requested samples but only 3 were solved, then $\delta = -12$. Also, if $t = 10$ and $f = 2$, meaning that when treshold is exceeded, double sample quantity is requested. Therefore in the first example the new value for area a 1 is $n' = 3$, and for a 2 is $n' = 12 \times 2 = 24$.
Like the mesh generation step, several strategies could be defined for this step, and the pipeline could be configured according to the domain and objectives of each project.

3.5 User Location-based Task Recommendation System

An external system could consume the sampling task list generated in the former step to assign the tasks to particular users.
There are several approaches to distributing tasks taking into account other aspects such as user experience [4]. Although this step is introduced in the main description of the pipeline process, its details are out of the scope of this article. However, according to the nature of geolocated activities, some research lines are introduced in

the following.
The use of health statutes, user behavior, the weather, and other context variables could be considered when recommending tasks to users. There are approaches related to physical exercise and the idea of avoiding users getting boring doing it [10]. In the context of this article, the sampling task could be assigned to people who complement their physical activities within the objective of the original projects: people do exercise and cooperate in collecting data of interest in an area.
In addition, several approaches centered the user preferences in order to recommend points of interest based on external resources like a knowledge base of restaurants, shopping malls, or any other social places[13]. However, the point of interest that feeds those approaches could be combined with the sampling task list, which can be considered as located areas of interest.
Finally, using user traveling behaviors is another approach to focus the distributionof sampling tasks. This means selecting those tasks that better adapt to the spatial and temporal behavior of users.

4 Conclusion and Future work

This work presented a pipeline approach for decision-making in generating location-based sampling tasks. This proposal considered the project's coverage objectives, area priority configuration, and global system status.
The generated task list may be a requirement for a recommendation system that assigns tasks to the users who participate in the project based on user profiles that can consider the person's characteristics, preferences, or historical behavior. Also, that system can recommend game items in an adaptive gamification approach.
The meshing approach developed in this work can be considered a cold start for an adaptation strategy based on the community's behavior. An initial set of areas can then be adjusted by calculating Voronoi regions taking as centroids the samples or from other hot spots that represent the busiest areas. This complemented approach allows having a dynamic set of sampling areas to propose a better distribution of the samples.
As further work, the inclusion of the proposed approach with the adaptive user-preferences-centered task distribution will be analyzed. Although there is a wide range of approaches, the first steps will be focused on using gamification strategies in the recommendation. For example, to generate specific gamification elements based on the generated task list to consider the project and user goals and preferences.

References

1. Cochero, J.: Appear: A citizen science mobile app to map the habitat quality of continental waterbodies. Ecologia Austral 28, 467–479 (08 2018)
2. Cochero, J., Pattori, L., Balsalobre, A., Ceccarelli, S., Marti, G.: A convolutional neural network to recognize chagas disease vectors using mobile phone images. Ecological Informatics 68, 101587 (2022)
3. Dalponte Ayastuy, M., Torres, D.: Relevance of non-activity representation in traveling

user behavior profiling for adaptive gamification. In: Proceedings of the XXI International Conference on Human Computer Interaction. Interacción '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3471391.3471431
4. Dalponte Ayastuy, M., Torres, D., Fernández, A.: A model of adaptive gamification in collaborative location-based collecting systems. In: Degen, H., Ntoa, S. (eds.) Artificial Intelligence in HCI. pp. 201–216. Springer International Publishing, Cham (2022)
5. Du, Q., Gunzburger, M.: Grid generation and optimization based on centroidal voronoi tessellations. Applied Mathematics and Computation 133(2), 591–607 (2002). https://doi.org/https://doi.org/10.1016/S0096-3003(01)00260-0, https://www.sciencedirect.com/science/article/pii/S0096300301002600
6. Fleischmann, M., Feliciotti, A., Romice, O., Porta, S.: Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. Computers, Environment and Urban Systems 80, 101441 (2020). https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2019.101441, https://www.sciencedirect.com/science/article/pii/S0198971519302856
7. Franzoni, C., Sauermann, H.: Crowd science: The organization of scientific research in open collaborative projects. Research Policy 43(1), 120 (2014). https://doi.org/https://doi.org/10.1016/j.respol.2013.07.005, https://www.sciencedirect.com/science/article/pii/S0048733313001212
8. Gold, C.: Tessellations in gis: Part i—putting it all together. Geo-spatial Information Science 19(1), 9–25 (2016). https://doi.org/10.1080/10095020.2016.1146440
9. Gold, C.M., Remmele, P.R., Roos, T.: Voronoi methods in GIS, pp. 21–35. Springer Berlin Heidelberg, Berlin, Heidelberg (1997). https://doi.org/10.1007/3-540-638180_2
10. He, Q., Agu, E., Strong, D., Tulu, B.: RecFit: a context-aware system for recommending physical activities. In: Proceedings of the 1st Workshop on Mobile Medical Applications. pp. 34–39. ACM, Memphis Tennessee (Nov 2014). https://dl.acm.org/doi/10.1145/2676431.2676439
11. Kazemi, L., Shahabi, C.: Geocrowd: Enabling query answering with spatial crowdsourcing. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems. p. 189–198. SIGSPATIAL'12, Association forComputing Machinery, New York, NY, USA (2012). https://doi.org/10.1145/2424321.2424346
12. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: Geographic information science and systems. John Wiley & Sons (2015).
13. Massai, L., Nesi, P., Pantaleo, G.: PAVAL: A location-aware virtual personal assistant for retrieving geolocated points of interest and location-based services. Engineering Applications of Artificial Intelligence 77, 70–85 (2019) https://doi.org/https://doi.org/10.1016/j.engappai.2018.09.013, https://www.sciencedirect.com/science/article/pii/S0952197618301994
14. Nugent, J.: inaturalist, citizen science for 21st-century naturalists. Science Scope 41(7), 12–13 (2018)
15. Reddy, S., Estrin, D., Srivastava, M.: Recruitment framework for participatory sensing data collections. In: Floréen, P., Krüger, A., Spasojevic, M. (eds.) Pervasive Computing. pp. 138–155. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
16. Shirani-Mehr, H., Banaei-Kashani, F., Shahabi, C.: Efficient viewpoint assignment for urban texture documentation. In: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 62–71 (2009)
17. To, H., Shahabi, C., Kazemi, L.: A server-assigned spatial crowdsourcing framework. ACM Trans. Spatial Algorithms Syst. 1(1) (jul 2015). https://doi.org/10.1145/2729713
18. Tsai, V.J.D.: Delaunay triangulations in tin creation: an overview and a linear time algorithm. International Journal of Geographical Information Systems 7(6), 501–524 (1993). https://doi.org/10.1080/02693799308901979

Transformations on Knowledge Representation between OWL and RDF Knowledge Graphs: A Study Case

Mariano Ferreirone[1,2,3], Mario Lezoche[1], Diego Torres[2,3], and Hervé Panetto[1]

[1] University of Lorraine, CNRS, CRAN, F54000 Nancy, France
mariano-julian.ferreirone@univ-lorraine.fr, mario.lezoche@univ-lorraine.fr, herve.panetto@univ-lorraine.fr
[2] LIFIA, CICPBA-Facultad de Inform´atica, UNLP, Argentina
mferreirone@lifia.info.unlp.edu.ar, dtorres@lifia.info.unlp.edu.ar
[3] Depto. CyT, UNQ, Bernal, Argentina
dtorres@unq.edu.ar, mferreirone@unq.edu.ar

Abstract. Is well known that the semantic web is having a tremendous impact on many aspects of the world and that it's a wave that is far away from going down. Ontology and Knowledge graphs are two meth- ods of knowledge representation that are part of the basis of this wave, and both have their pros and cons. A big part of the agricultural devel- opment focuses on these models, mainly interested in the possibility of exploiting implicit knowledge. In this work, there is an analysis over the relation between a rigid knowledge representation model as OWL, and asimple and more flexible one like RDF. This is based on the attempt oftransforming an OWL knowledge graph into an RDF knowledge graph, taking into account the interesting possibility of combining knowledge graphs that were created with different levels semantic expressiveness. The work also presents a case of study on the chess domain.

Keywords. Ontology, Knowledge graph, OWL, RDF, Knowledge representation.

1 Introduction

During the last years, Knowledge Graphs (KG) have been demonstrated that are well known alternatives for knowledge discovering. There are well known cases in the industry that uses KG in order to organize their data and then discover underlying knowledge that are not directly present in their data sources [8]. Since the agriculture development is aware of this knowledge representation

model, an important part of the community efforts are dedicated to creating these representations. Taking into account that there are many different systems that are being constructed in this industry, it's also important to work in the interoperability between them, in order to state the goal of having an integrated and consistent knowledge representation model.

A KG is a formal representation of knowledge in the form of a labeled directed graph, where the nodes represent concepts or an actual entity from the real world, meanwhile the edges represent different relations between these nodes [5]. A standard data model to represent KG is the Resource Description Framework (RDF), [10] which use triples of the form (subject, predicate, object). Additionally, the RDF model is complemented by query languages in which SPARQL isthe most prominent.

KGs could be represented with different models which co-exists. The simplestone is the aforementioned RDF, which only supports really simple semantics. It's based on XML, so it inherits the XML datatype definitions. Another model is RDFS (Resource Description Framework Schema) which is an extension of RDF, and it introduces simple constraints and semantics, as class and property subtypes, property range and domain restrictions. Finally, OWL (Web Ontology Language) that introduces several ontological characteristics on top of RDFS[1]. Taking into account the aforementioned, the authors found interesting to investigate how these different KGs, with different levels of semantic expressiveness, could be combined, in order to increase the interoperability between them.

During the last years, the amount of RDF knowledge graphs has been increasing and the most big and popular knowledge representations of this type are created with this formalization language. As an example, Wikidata[2] is constructed on RDF and it has several billions of triples [4].

The RDF knowledge graphs could have a schema behind in order to add consistency to the model and the data, and this is the case when they are ontology-based. In the case that the KG is purely developed in RDF, the lack of a formal schema does not ensure the consistency of the data, and there's a lack of semantic expressiveness in comparison with an ontology. Another strong point to highlight is that the knowledge graphs which use a shared ontology are more inter-operable since the ontology structure is unambiguous and it has an accepted and common meaning in the community [5]. On the other side, RDF/Sknowledge graphs are more flexible with the addition of new information, and since they don't have to do a strict review of the structure, is more efficient from a computational point of view. On the other hand, OWL KG has more expressiveness in their semantics, but they are not as fast computationally as the Knowledge graphs [7]. These are reasons that make attractive the idea of working in a bidirectional connection or transformation of these two knowledge representation models.

Some of the advantages of this approach are detailed as follows: Having an ontology with all the instances from the equivalent knowledge graph is useful for

[1] Wikidata, OWL, https://en.wikipedia.org/wiki/WebOntologyLanguage

[2] Wikidata, https://www.wikidata.org/wiki/Wikidata:MainPage

visualizing the hierarchy and the structure in a clear way, and it also ensures the consistency of the model. Furthermore, having into account that RDF Knowledge graphs are likely to have a good efficiency, but they lack on the consistency since they don't have a complex semantic structure behind, there's a gap that awakes the scientific research interest, in order to add this expressiveness that, for example, OWL has, creating a good combination of both characteristics. Moreover, the translation would help to the interoperability between systems that have their knowledge representation models, giving the possibility of choos- ing which model is the preferred one, to create the integration following it. For example, if the system A has its model in OWL, and the system B has it in RDF, the goal of this development is to give the possibility of selecting if the integrated model will be in OWL or in RDF, only transforming the corresponding one.

The approach that the authors of this article have developed, in order to start a solution path for the aforementioned points of improvement, is to create a set of rules or steps with the objective of transforming an OWL Knowledge graph into an RDF knowledge graph, with the intention of analyzing what is lost in the middle and define further steps.

There are some existent interesting works which are related to converting an RDF and OWL into different formats. An example of this is the converter developed as part of the tool named CoGui[3] by the GraphiK team at LIRMM[3]. In this case, the transformation is done from RDF to conceptual graphs, and the resulting OWL file is exported to different languages. When referring to RDF, currently, OWL rules, constraints and type disjunctions are ignored. Another interesting work is the converter tool of the University of Manchester[4]. This converter doesn't have the possibility of converting an OWL structure to a simple RDF syntax. Another interesting concept to take into account is the ontology alignment or ontology matching, which is based on generating a set of correspondences between concepts, properties or instances of different structured KGs, with the objective of unifying them into a new one [2].

The objective of this work is to develop a simple algorithm, aiming to introduce the initial analysis of applying transformation rules which transform an OWL KG into an RDF KG, primarily focused on the semantic loss that takes place in the middle of the process. The approach is conducted over a study case of a chess ontology. Learned lessons and challenges are reported.

The article is structured as follows: in section 2 there's a general description about the transformation process that the authors have designed and applied. In section 3 we describe the creation of an OWL ontology over the chess domain, and the application of the transformation to the corresponding file. Finally, in section4 the learned lessons are described, and the next steps and further challenges are mentioned.

[3] CoGui Homepage, https://www.lirmm.fr/cogui/3/index.html. Last accessed5/5/2022

[4] OWLSyntaxConverter,UniversityofManchester http://mowl-power.cs.man.ac.uk:8080/converter/. Last accessed 5/5/2022

2 Transformation

The transformation developed in this work consists of the creation of a program, with the objective of converting an OWL KG into different types of RDF triplesrepresentations, which means, RDF Knowledge graphs.
As a first step, the user can make the decision of working with the original OWL KG input, or if it's desired to infer the file in order to include also the implicit axioms in the forward steps, using a reasoner engine and creating a new version of the file. On the second step, the program reads the chosen input KG and obtains all the elements from it, including classes, sub-classes, properties, etc., and assigns each of them to a graph structure that was previously created by the authors. This graph structure has the following classes: Graph, which contains the name of the graph, all the classes and all the individuals; Class, that contains information related to a class, as his name and iri, all the subclasses and the individuals of the class; Individual, which contains the name, iri and parent class of the individual. The figure 1 describes the graph structure. Once this graph is fed, the data and the knowledge are ready for being processed.

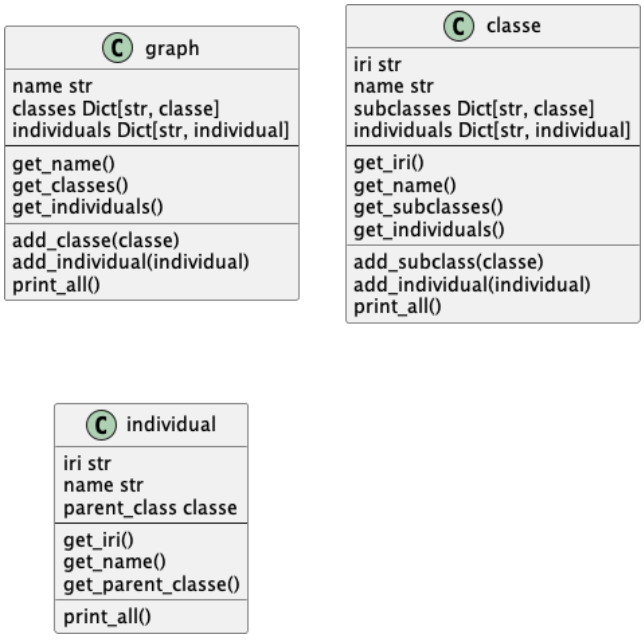


Fig. 1. Graph structure

As the next step, the generation of an intermediate output, which is going to be structured as simple triple stores in the form (subject, predicate, object) takes place. Then, the triple stores are translated to RDF/XML syntax, including

also those triples that are formed by OWL elements and don't correspond to the RDF semantic level. This is done using a template, which is described in the figure 2, and creating a customized RDF triple, where some triples could be compounded, concatenating information in the predicate or in the object. Finally, the latter generated output will pass through a cleaner with the objective of creating an RDF/XML output in the corresponding RDF semantic level.

```
<rdf:Description rdf:about="${subject}">
    <ex:${predicate}>
        <rdf:Description rdf:about="${object}"/>
    </ex:${predicate}>
</rdf:Description>
```

Fig. 2. Example of Subclass triples obtained

The implementation of the program was done using the programming lan- guage Python. Furthermore, a part of the manipulation of the OWL file was developed re-utilizing functions of the Python library called OWLREADY2[5], which also offers the possibility to execute SPARQL queries over the ontology. Regarding the inference, the reasoner Hermit[6] was the chosen one. For visualizing the obtained RDF triples in a graphical representation, the Python package Graphviz[7] was used. However, when the file is too large, the picture is hard to read, and the graph is often stretched. The application Neo4J with the plugin Neosemantics was also utilized to load the output triples and represent them in a graph. The transformation process is represented in the figure 3. It's important to mention that in this stage of the project, the work is not going to be focused on the computational efficiency. Instead, the efforts are concentrated in the functional sense.

3 Case of study: application on the chess domain

3.1 The ontology

In an attempt of analyzing the existent knowledge formalization done on the chess domain, the work done by Adila Krisnadhi and Pascal Hitzler on their published chapter "Modeling With Ontology Design Patterns: Chess Games As a Worked Example" [6] was found very interesting. However, the design of the aforementioned ontology is more oriented to the representation of a chess competition.

Regarding this case study, the first step was to create an ontology with some

[5] OWLREADY2, documentation, https://owlready2.readthedocs.io/en/v0.37/
[6] Hermit, http://www.hermit-reasoner.com/
[7] Graphviz, https://graphviz.org/

individuals (an OWL KG), based on the chess domain. The idea of the designis to represent the game, with all the important factors, and also represent a match that has been played as a list of movements, or in other words, as the evolution of the pieces on the board. As a supplementary support, the rules of the game could be found in the internet[8].

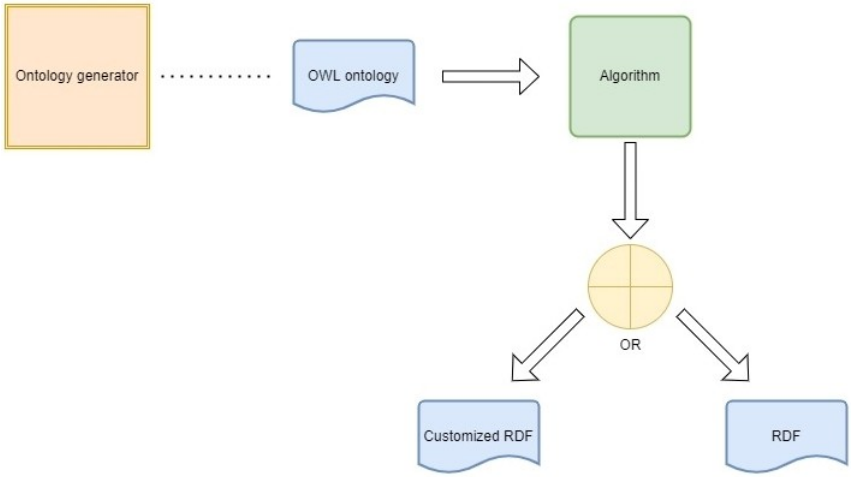


Fig. 3. Diagram of the transformation process

A description of the structure of the created ontology is as follows: The ontol-ogy has five main classes which are: ”Board”, ”Match”, ”Pieces”, ”Players” and ”Rules”. The ”Board” class contains the sub-classes ”Cell”, ”file” and ”rank”, with the objective of representing each position where a piece can be placed. The first subclass contains sixty-four individuals in order to instantiate this, and each one of these individuals is related through two object properties to one individual of the subclass ”file” and one of the subclass ”rank”. They are also related to a string value, so for example, the cell c3 is related to the string value ”c3”. This was done to workaround the following issue: the reasoner engine is not capable to understand semantically the name of a subclass or an individual. The ”Match” class is designed to represent a specific list of movements that are attached to one specific game played by two entities. An individual of the match is going to be related to an object property with two different players, which one will be identified with the white pieces, meanwhile the other one is going to be related with the black pieces (both sides of the match). The class ”Pieces” contains all the different pieces that are part of the game, like the Bishop, the

[8] https://en.wikipedia.org/wiki/Chess

Tower or the King as sub-classes. Each one of this, has individuals to represent the specific pieces that are on the board. For example, the subclass ”knight” has two individuals per color. ”Players” contains two sub-classes called ”AI” and ”Person”, and they represent the entities that will play the match. ”Rules” class contains some specific allowed movements (for example ”en passant”, the castle, the promotion) and also the conditions that could get the game to the end in the subclass ”win conditions”.
The class hierarchy of the ontology is graphically represented in Protege inthe figure 4 and in the corresponding VOWL diagram in figure 5

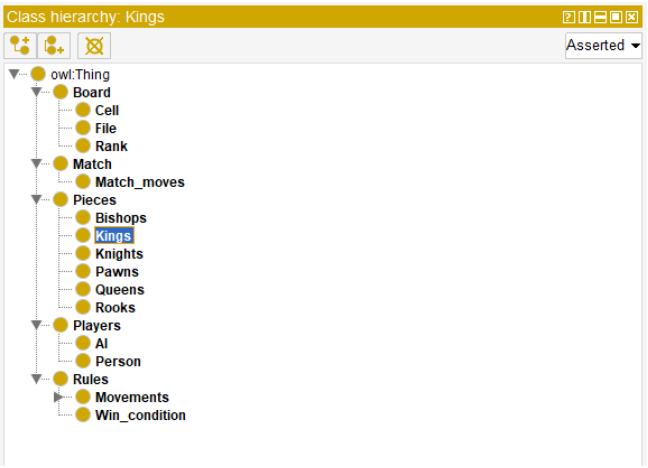


Fig. 4. Protégé structure

The main relations which help to represent a played game are the following:
- Match *Where black player is* Player
- Match *Where white player is* Player
- Match moves *To cell* cell
- Match moves *Moving Piece* Pieces
- Match moves *Next move* Match moves
- Pieces *Strat cell is* Cell

The object property hierarchy of the ontology is described in figure 6.
In order to represent a match, the design takes into account the following statements: A game is played by two players. One plays the white pieces and the other plays the black pieces. A match is a set of movements alternatively of white pieces and black pieces from a starting situation to the end of the match.A movement, in our chess ontology, is basically a piece which is moving from a specific cell to another specific cell of the board.

This ontology was created utilizing the free and open source software Protégé[9], which is widely used to create ontologies and it has a lot of interesting utilities. The mentioned ontology has been stored in OWL format, since it's one of the most common languages for this purpose, and it has a high level of schema formalization, so it represents the lack of flexibility that it's needed for the objective of this work. This file format can be open as a text file or with different specific software like Protégé. A little part of the structure of the language is shown in figure 7 and in figure 8.

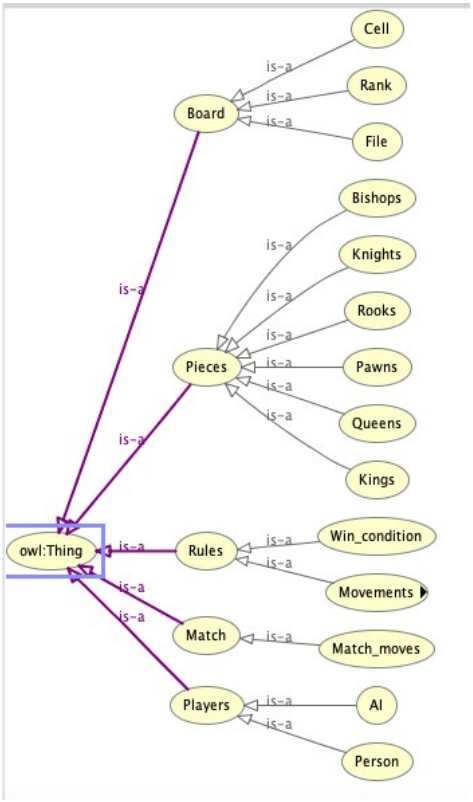


Fig. 5. VOWL diagram of the class hierarchy

In order to visualize the ontology in the tool Protégé, the plugins OWLViz[10] and VOWL[11] were utilized. A VOWL representation of the ontology can be seen in the figure 9.

[9] Protégé, Homepage https://protege.stanford.edu/ Last accessed 5/6/2022
[10] OWLViz, https://protegewiki.stanford.edu/wiki/OWLViz
[11] VOWL, http://vowl.visualdataweb.org/

3.2From OWL to RDF

The second step of this work consists in the application of the transformation described in section 2, with the objective of converting the OWL KG file into RDF triples that represent an RDF KG. The created chess KG mentioned in the above section was exported from Protégé as a .OWL file and different tests were executed in order to evaluate the different possible behaviours of the process. The transformation was done with the original ontology file and with the inferenced one, and several outputs were generated formatted as RDF triple stores, customized RDF and RDF/XML.

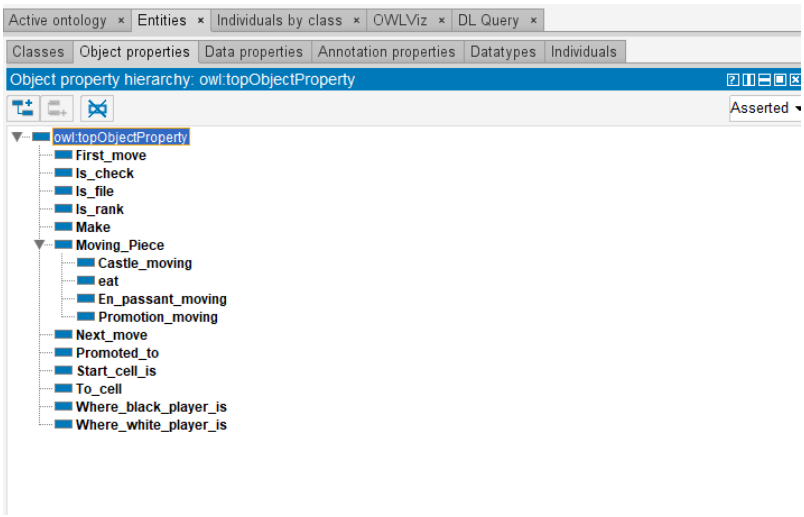


Fig. 6. Object properties in Protégé

```
    <!-- http://www.semanticweb.org/killi/ontologies/2022/0/Chess_Ontology#Bishops -->

    <owl:Class rdf:about="http://www.semanticweb.org/killi/ontologies/2022/0/Chess_Ontology#Bishops">
        <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/killi/ontologies/2022/0/Chess_Ontology#Pieces"/>
        <owl:disjointWith rdf:resource="http://www.semanticweb.org/killi/ontologies/2022/0/Chess_Ontology#Kings"/>
        <owl:disjointWith rdf:resource="http://www.semanticweb.org/killi/ontologies/2022/0/Chess_Ontology#Knights"/>
        <owl:disjointWith rdf:resource="http://www.semanticweb.org/killi/ontologies/2022/0/Chess_Ontology#Pawns"/>
        <owl:disjointWith rdf:resource="http://www.semanticweb.org/killi/ontologies/2022/0/Chess_Ontology#Queens"/>
        <owl:disjointWith rdf:resource="http://www.semanticweb.org/killi/ontologies/2022/0/Chess_Ontology#Rooks"/>
        <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">This is a piece. There are two bishops of
each side (black and white). See also Bishop&amp;apos;s movement.</rdfs:comment>
    </owl:Class>
```

Fig. 7. Bishop class in OWL file

```
<!-- http://www.semanticweb.org/killi/ontologies/2022/0/chess-ontology-2#Moving_Piece -->

<owl:ObjectProperty rdf:about="http://www.semanticweb.org/killi/ontologies/2022/0/chess-ontology-2#Moving_Piece">
    <rdfs:subPropertyOf rdf:resource="http://www.w3.org/2002/07/owl#topObjectProperty"/>
    <rdfs:domain rdf:resource="http://www.semanticweb.org/killi/ontologies/2022/0/chess-ontology-2#Match_moves"/>
    <rdfs:range rdf:resource="http://www.semanticweb.org/killi/ontologies/2022/0/Chess_Ontology#Pieces"/>
    <rdfs:comment>Hadrien : It is not a functional property because the caste movement &quot;switch and move&quot;
both the King and the Rook</rdfs:comment>
</owl:ObjectProperty>
```

Fig. 8. Moving Piece property in OWL file

In order to obtain the knowledge and being able to process it, the OWL structure was transformed into the python graph structure. This load was then evaluated through a comparison between the number of elements that were present in Protégé and the number of elements that were loaded to the python graph

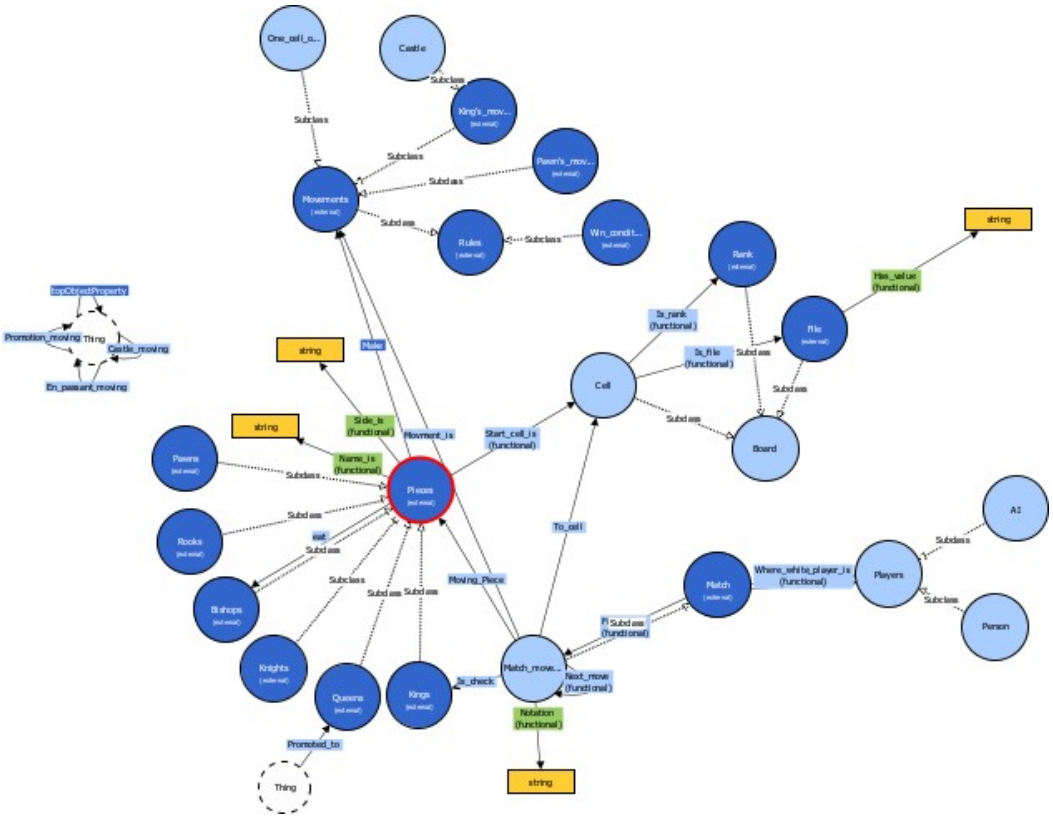


Fig. 9. VOWL ontology representation

structure. In this case, this test had successful results. The figure 10 shows an example where the number of individuals is compared.
The first output is generated in triples of the form (subject, predicate, object), which represent the relation between two nodes in the knowledge graph. Futhermore, for the sake of the analysis, a customized RDF structure output has been generated, in which the triples are compound. This means that everything from the OWL input is present in the output, concatenating information in the predicate or in the object. As an example, the input OWL statement *pawn promoted to "Pieces not(Kings) not(Pawns)"* was translated to an "RDF" triple with a compound object *"not(Kings) not (Pawns)"*. After cleaning the customized RDF output, the triples are also translated to an RDF structure,

Fig. 10. Comparison of individual between Prot´eg´e and program

which is described in the figure 2. The obtained triples in this occasion have a lack of expressiveness, since there's an obvious difference between it and OWL. This makes that the constraints, types of relationships and all the characteristics that are specific from OWL, are not represented in RDF. Examples of the RDF triples obtained are shown in the figures 11 and 12.

```
======================================================================
==================  SUBCLASS TRIPLES ==================
======================================================================
('Chess_Ontology.AI', 'subClassOf', 'Chess_Ontology.Players')
('Chess_Ontology.Castle', 'subClassOf', "Chess_Ontology.King's_movement")
('Chess_Ontology.Cell', 'subClassOf', 'Chess_Ontology.Board')
('Chess_Ontology.Double_forward', 'subClassOf', "Chess_Ontology.Pawn's_movement")
('Chess_Ontology.Eat', 'subClassOf', "Chess_Ontology.Pawn's_movement")
('Chess_Ontology.En_passant', 'subClassOf', "Chess_Ontology.Pawn's_movement")
('Chess_Ontology.Match_moves', 'subClassOf', 'Chess_Ontology.Match')
('Chess_Ontology.One_cell_one_piece', 'subClassOf', 'Chess_Ontology.Movements')
('Chess_Ontology.One_forward', 'subClassOf', "Chess_Ontology.Pawn's_movement")
('Chess_Ontology.One_square_move', 'subClassOf', "Chess_Ontology.King's_movement")
('Chess_Ontology.Person', 'subClassOf', 'Chess_Ontology.Players')
('Chess_Ontology.Promotion', 'subClassOf', "Chess_Ontology.Pawn's_movement")
("Chess_Ontology.Queen's_movement", 'subClassOf', 'Chess_Ontology.Movements')
('Chess_Ontology.Bishops', 'subClassOf', 'Chess_Ontology.Pieces')
('Chess_Ontology.Kings', 'subClassOf', 'Chess_Ontology.Pieces')
('Chess_Ontology.Knights', 'subClassOf', 'Chess_Ontology.Pieces')
('Chess_Ontology.Movements', 'subClassOf', 'Chess_Ontology.Rules')
('Chess_Ontology.Pawns', 'subClassOf', 'Chess_Ontology.Pieces')
('Chess_Ontology.Queens', 'subClassOf', 'Chess_Ontology.Pieces')
('Chess_Ontology.Rooks', 'subClassOf', 'Chess_Ontology.Pieces')
('Chess_Ontology.Win_condition', 'subClassOf', 'Chess_Ontology.Rules')
("Chess_Ontology.Bishop's_movement", 'subClassOf', 'Chess_Ontology.Movements')
("Chess_Ontology.King's_movement", 'subClassOf', 'Chess_Ontology.Movements')
("Chess_Ontology.Knight's_movement", 'subClassOf', 'Chess_Ontology.Movements')
("Chess_Ontology.Pawn's_movement", 'subClassOf', 'Chess_Ontology.Movements')
("Chess_Ontology.Rook's_movement", 'subClassOf', 'Chess_Ontology.Movements')
```

Fig. 11. Example of Subclass triples obtained

```
======================================================================
==================  INDIVIDUALS TRIPLES ==================
======================================================================
('Chess_Ontology.B1', 'individualOf', 'Chess_Ontology.Bishops')
('Chess_Ontology.c1', 'individualOf', 'Chess_Ontology.Cell')
('Chess_Ontology.B2', 'individualOf', 'Chess_Ontology.Bishops')
('Chess_Ontology.f1', 'individualOf', 'Chess_Ontology.Cell')
('Chess_Ontology.B3', 'individualOf', 'Chess_Ontology.Bishops')
('Chess_Ontology.c8', 'individualOf', 'Chess_Ontology.Cell')
('Chess_Ontology.B4', 'individualOf', 'Chess_Ontology.Bishops')
('Chess_Ontology.f8', 'individualOf', 'Chess_Ontology.Cell')
('Chess_Ontology.K1', 'individualOf', 'Chess_Ontology.Kings')
('Chess_Ontology.e1', 'individualOf', 'Chess_Ontology.Cell')
('Chess_Ontology.K2', 'individualOf', 'Chess_Ontology.Kings')
('Chess_Ontology.e8', 'individualOf', 'Chess_Ontology.Cell')
('Chess_Ontology.N1', 'individualOf', 'Chess_Ontology.Knights')
('Chess_Ontology.b1', 'individualOf', 'Chess_Ontology.Cell')
('Chess_Ontology.N2', 'individualOf', 'Chess_Ontology.Knights')
```

Fig. 12. Example of Individual triples obtained

In order to show the lost semantics that this transformation has, the OWL statement *pawn promoted to "Pieces not(Kings) not(Pawns)"* can be highlighted as an example again. In this case, the object or range is only one element for the OWL file, but it cannot be transformed into only one element in the RDF structure. Moreover, it would be necessary to create several triples to represent only one OWL relation with constraints. An example of a graphical representation of the generated triples is in the figure 13.

4 Learning lessons and further steps

Through this exercise, the authors have analyzed the lost of semantics that takes place in a simple transformation from an OWL KG into an RDF KG, which is related to the difference over the expressiveness between them. RDF is a lighter and more flexible representation, and it's not designed to express the level of semantics that OWL does. The experiment showed that the following characteristics weren't able to be represented in the output of the transformation: type of relationships (symmetric, transitive, etc.), cardinality constraints, exact values, complex classes and more. Furthermore, the transformation had the necessity of looking for generic keywords in order to not depend on the specific syntax of a specific file. Regarding the visualization of the output, it was not possible to generate a clear one, since the triples were many and the visualization was too diffused.

Currently, the authors have in mind two possible options to treat this expressiveness difference. The first one is to create a representation of the OWL ontology without constraints, but maintaining all the logic that is behind these rules or definitions. A possible path to do this is investigating and testing the

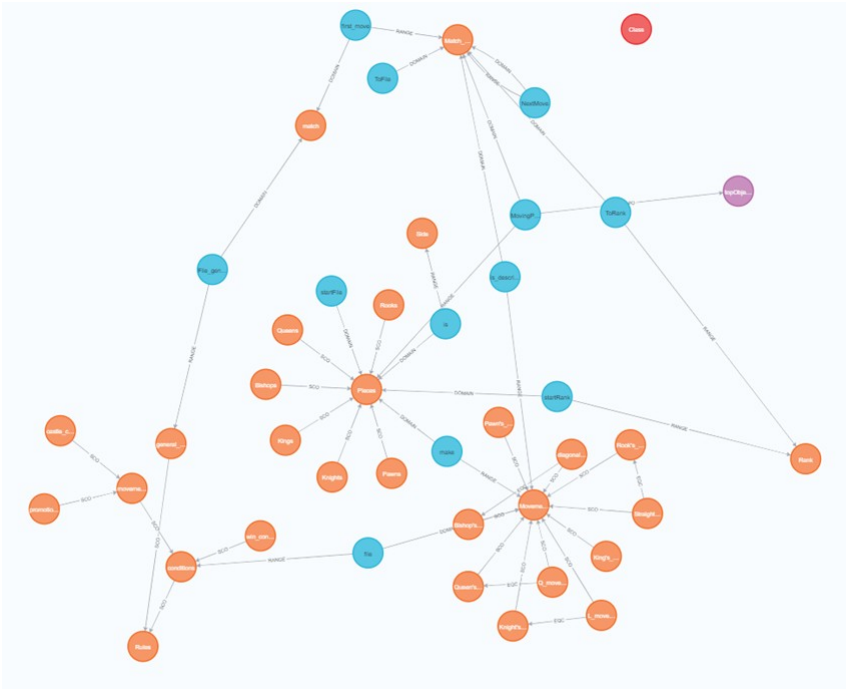


Fig. 13. Output triples in Neo4j

output generated after the inferences over the OWL structure and deleting the constraints. The second option is to keep all the constraints and finding a way to express them. This could represent the creation of new nodes, and additional triples would represent the complex axioms in RDF. With the customized RDF output, a primitive and first attempt of maintaining all the characteristics in the RDF KG was done, but this is only with analysis purposes since it's not practical as its. However, this output will be also taken into account in the further steps.

Regarding the next steps, one possible path would be to add a new step in the program with the objective of interpreting this customized RDF structure, giving to the compound triples a representation that fits with the RDF syntax. Another possible path to follow, in order to add these characteristics to an RDF graph, could be to add a layer of representation, to express this structure in an efficient manner. The authors will research existent work related to adding formalization to RDF KG. It's interesting for the authors to study Shapes Constraint Language [1], the characteristics of the property graphs [9], ontology alignment [2] and the state of the art of its applications into the knowledge graphs. The objective would be to integrate some of this existent concepts, or develop a new one inspired by them, and introduce it in this transformation process. Furthermore, it's interesting for the authors to investigate the transformation in

the opposite direction, which means from an RDF KG to an OWL KG.

5 Acknowledgments

The application of the idea to the case of study has been done with the help of Hamza ABDOULHOUSSEN, Killian CRESSANT and Hadrien ROCHU, whoare currently doing their master in computer science at the Université of Lor- raine (Nancy). Their contribution was done inside the framework of the ”Projet Interdisciplinaire de recherche” (PIDR), which is a research project conducted as part of the TELECOM Nancy curriculum.

References

1. Shapes constraint language (SHACL). Tech. rep., W3C (Jul 2017), https://www.w3.org/TR/shacl/
2. Ardjani, F., Bouchiha, D., Malki, M.: Ontology-alignment techniques: Survey and analysis. International Journal of Modern Education and Computer Science **7**, 67–78 (11 2015). https://doi.org/10.5815/ijmecs.2015.11.08
3. Baget, J.F., Chein, M., Croitoru, M., Fortin, J., Genest, D., Gutierrez, A., Lecl`ere, M., Mugnier, M.L., Salvat, E.: RDF to Conceptual Graphs Translations. In: CS-TIW: Conceptual Structures Tool Interoperability Workshop. vol. LNAI, p. 17. Springer, Moscow, Russia (Jul 2009), https://hal-lirmm.ccsd.cnrs.fr/lirmm-00410621
4. Desouki, A., R¨oder, M., Ngonga Ngomo, A.C.: Ranking on very large knowledge graphs. pp. 163–171 (09 2019). https://doi.org/10.1145/3342220.3343660
5. Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G.d., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., Ngomo, A.C.N., Polleres, A., Rashid, S.M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., Zimmermann, A.: Knowledge Graphs. Synthesis Lectures on Data, Semantics, and Knowledge **12**(2), 1–257 (Nov 2021). https://doi.org/10.2200/S01125ED1V01Y202109DSK022, https://www.morganclaypool.com/doi/abs/10.2200/S01125ED1V01Y202109DSK022, publisher: Morgan & Claypool Publishers
6. Krisnadhi, A., Hitzler, P.: Modeling With Ontology Design Patterns: Chess Games As a Worked Example, vol. 25, chap. 1, p. 3–21. IOS Press (2016). https://doi.org/10.3233/978-1-61499-676-7-3
7. Lera, I., Juiz, C., Puigjaner, R.: Performance-related ontologies and semantic web applications for on-line performance assessment of in- telligent systems. Science of Computer Programming **61**(1), 27– 37 (2006). https://doi.org/https://doi.org/10.1016/j.scico.2005.11.003, https://www.sciencedirect.com/science/article/pii/S016764230600013X, special Issue on Quality system and software architectures
8. Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., Taylor, J.: Industry-scale knowledge graphs: lessons and challenges. Communications of the ACM **62**(8), 36– 43 (Jul 2019). https://doi.org/10.1145/3331166
9. Tomaszuk, D., Angles, R., Thakkar, H.: Pgo: Describing property graphs in rdf. IEEE Access **8**, 118355–118369 (2020)
10. Wood, D., Lanthaler, M., Cyganiak, R.: Rdf 1.1 concepts and abstract syntax. W3C Recommendation, W3C (2014)

Coastline Generator: A Tool for Generating Topographic Tessellations Around Polygons and Lines

Bruno Lattanzio[1][0000−0001−5380−0915], María Dalponte Ayastuy[1,2][0000−0002−1412−5694] and Diego Torres[1,2][0000−0001−7533−0133]

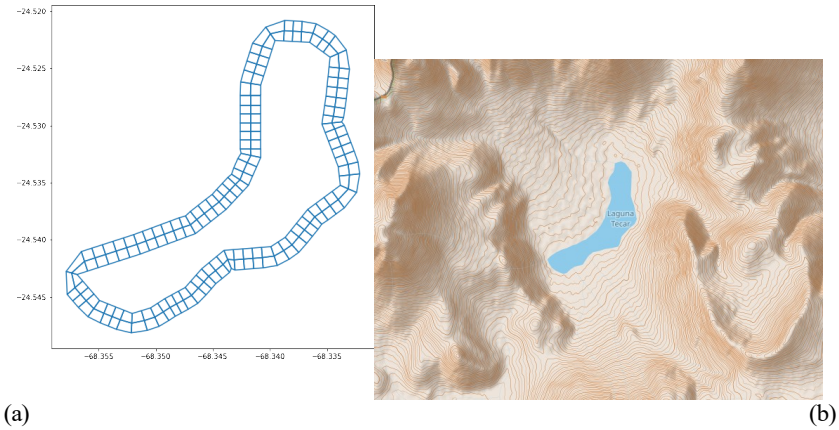
[1] Depto CyT, Universidad Nacional de Quilmes Roque Saenz Peña 352, Bernal, Buenos Aires, Argentina.
{brunoj.lattanzio,mdalponte}@unq.edu.ar
[2] LIFIA,CICPBA-Facultad de Informática, Universidad Nacional de La Plata 50 y 120, La Plata, Buenos Aires, Argentina.
diego.torres@lifia.info.unlp.edu.ar

Abstract. Many citizen science projects that carry out survey tasks based on location require that the territory to be studied be fragmented into smaller areas with the objectives of, on the one hand, keeping a record of the level of coverage of the regions, and on the other hand, to present spatially bounded objectives to the volunteers of the project. In some cases the sampling areas are related to a terrain feature, such as when surveying the shores of rivers and lakes. Therefore, the afore-mentioned segmentation must respect the topographical shape of the geographical object to be studied (river or lake).In this work, this type of tessellation is defined as topographic tessellation (TT). Aiming at building the TT, indicating the distance it should have from the shore and the specific measurements of each smaller area is needed. This article presents a framework for the automatic generation of topographic tessellations, which are sets of disjoint and adjacent polygons that form a mosaic following the shape of a georeferenced geometry, and builds a new geographical layer. This tool is useful for spatial-task asignment decision-making.

1 Introduction

Collaboration between professional scientists and volunteers committed to research is not new. Actually, the recognition of the scientific task as a paid profession is a fact whose antiquity does not exceed two centuries and previously it was carried out by amateur scientists. However, these citizens who do science have not disappeared and their observation skills are very useful in astronomical, natural sciences or archaeology projects, among others. Today formal scientists work in teams with citizens scientists on projects that have been specifically designed to give amateurs a role for the benefit of the project's objectives[7].

Citizen science encompasses a range of methodologies that encourage and support the contributions of the public to the advancement of scientific and



(a) (b)

Fig. 1. (a) Lake polygon and (b) topographic tessellation

engineering research and monitoring in ways that may include co-identifying research questions; co-designing/ conducting investigations; co-designing/ building/ testing low-cost sensors; co-collecting and analysing data; co-developing data applications; and collaboratively solving complex problems [9]

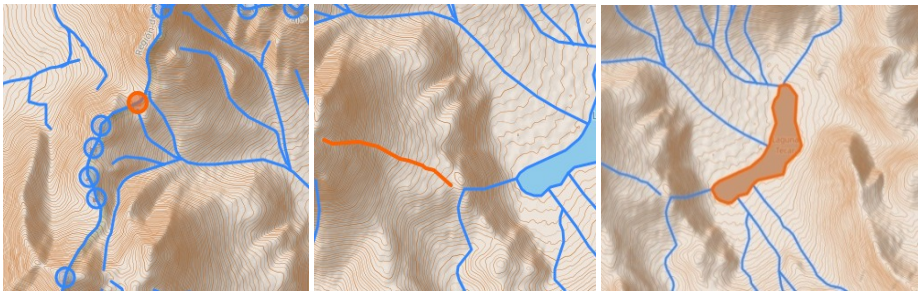
Usually, each CLCS sets its territory coverage objectives, such as, for example, guaranteeing that all the affected territory is surveyed [1]. This leads to the discussion about what level of granularity or precision is needed. A possible approach to specify the needed precision, as a first step requires to divide the territory into smaller areas, which allows indicating how many samples are needed in each of them. Aiming at this segmentation, specific cartography is needed to delimit the project's sampling areas and, in particular to the AppEar project, this set of areas are located around the shores of rivers, lakes, lagoons and estuaries.

In order to pursue its objectives, each project needs to dynamically define the tasks that it must present to volunteers based on the current situation of the project, and to assist in this sense (decision making) it is necessary to have a model of the space to be surveyed with a certain level of granularity where the activity of people can be registered.

Therefore it is necessary to generate a disjoint and adjacent set of areas -a tessellation- that are located following the shape of a geometry. Moreover, this areas must be organized in a regular shape grid where the number of rows and the dimension of its cells. In this work, this kind of tessellation is defined as Topographic Tessellation (see image 1 b). In the case of a river, the geometry that represents it can be a line or a polygon -depending on river's width-, and in the case of a lake or lagoon, the geometry is a polygon (see image 1 a).

The discipline of geographic information systems (GIS) defines and uses different tessellated models to represent information about the earth[4], sometimes to build a hierarchical discretization of a high definition raster image, or to support a vector representation of terrain surface, such as the Voronoi regions [5] or the Delaunay triangulation [8].
No articles directly related to the problem presented in this work were found, but there are works that address similar problems by calculating the Voronoi regions from a set of centroids, as is the case of Fleischman et al. [3] and Du and Gunzberg [2]. These approaches are not directly applicable in this domain because it requires a set of centroids and a bounding area to be defined. Without denying that this problem could be thought of in those terms, the definition of the centroids would represent additional and artificial steps.
Therefore, this paper presents and details a framework capable of generating a *Topographic Tessellation (TT)* following the shape of a line or a polygon. Also, the presented approach of cartographic generation allows to customize the grid's dimensions (cells width and row count) to be adaptable to the project's needs.
This article is organized as follows. In section 2 the background and domain- specific concepts are defined. In section 3 the system architecture and topographic tessellation algorithm is detailed. Finally the section 4 shares discussions and future work.

2 Background: Topographic tessellation



(a) Point element (b) Multiline element (c) Polygon element

Fig. 2. GIS elements examples

In this work, since cartography of the Argentine territory is used, provided by the National Geographic Institute[1] (NGI), it is appropriate to use the WGS84 reference system, which is the same used by the NGI. The hydrographic data that were used were obtained from the GIS layer of the NGI, in the section of *hydrocography and oceanography*, subindex of *continental waters*. These data are available in GeoJSON,

[1] https://www.ign.gob.ar/NuestrasActividades/InformacionGeoespacial/CapasSIG

an open standard format designed to represent geographical elements together with its non-spatial attributes, based on JSON. Some of these non-spatial attributes are described in Table 1.

Field	Type	Description
Entidad	Integer	Object ID
Objeto	String	Polygon subtype: Channel, waterflow, reservoir, perennial water mirror, rural reservoir, wetland, intermittent water mirror. Line type: Aqueduct, Canal, Perennial water currents, intermittent water currents, waterfall waterfall jump, reservoir wall, ditch.
FNA	String	Complete name of the geographical object
FDC	String	Id of the name and type of source used to capture the information. May include date and other additional data

Table 1. Input dataset metadata

Two methods are used to represent geographic phenomena in ways that can be encoded in spatial databases, called raster and vector methods. Although both can be used to encode both continuous fields and discrete objects, in practice there is a strong association between raster and fields, and between vector and discrete objects. One of the most common forms of raster data comes from remote-sensing satellites, which capture information as high definition images. In a vector representation, all lines are represented as points connected by a straight line, and areas are captured as a series of points or vertices connected by straight lines, called polygons. Lines are represented in the same way, and the term polyline (or multiline) has been coined to describe a curved line represented by a series of straight segments connecting vertices [6].

Definition 1 (Point). *Geographic coordinate defined by the tuple (latitude, longitude) of gegraphic coordinates. Negative values in latitude refer to points in the southern hemisphere, and negative values in longitude refer to points located west of the Greenwich meridian.*

As an example, see Figure 2 (a) where a point with latitude: -24,5073190, and longitude: -68,3958578-57,9380456 is shown.

Definition 2 (Multiline). *A multiline object is a sequence of more than two points.*

In Figure 2 (b) a multiline example is shown, made up with twenty nine points.

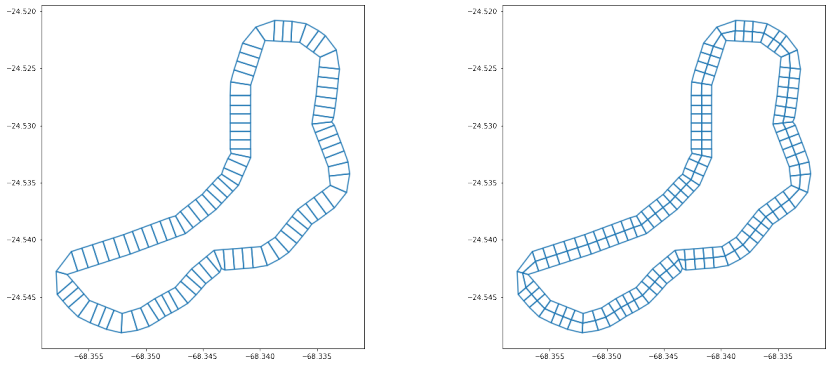
Definition 3 (Polygon). *A polygon is a multiline, where the first point is also the last one, making a closed geometry.*

The last example in Figure 2 is a polygon with 63 points.

Definition 4 (Topographical Tessellation). *Set of adjacent and disjoint cells that are organized following the shape of a given geometry. In other words, the upper limit of the mosaic is the geometry's limit, and the lines that separate the cells' rows are parallel to that limit.*

Two examples of Topographical Tessellations are shown in Figure 3.

3 Approach



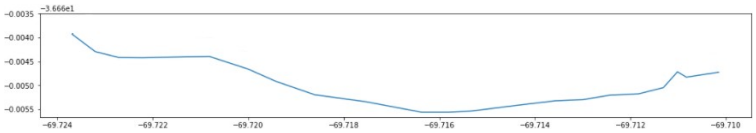
(a) TT with one row (b) TT with two rows

Fig. 3. Topographical tessellations for a lake

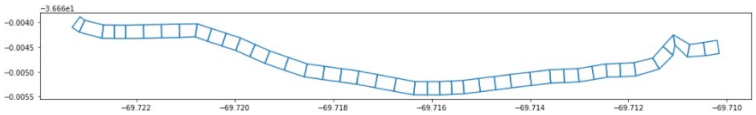
As presented in the introduction, the task of surveying the state of the coasts of rivers and lakes must be associated with a cell of the topographic tessellation (TT). As an example, consider a project that needs one sample every 50 meters around the lake. This requirement would need a TT with one row of 50 meters width cells, like the one depicted in Figure 3 (a). Another situation could need also a second sample of 70 meters from the shore, and in this case a different TT would be needed, with two rings or rows around the lake. This is graphically explained in Figure 3 (b). Also, the different sampling requirements can be applied to the geometry of a river, as is depicted in Figure 4.

3.1 General Architecture

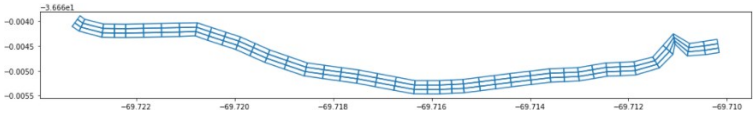
This framework has three main components (see Figure 5). The *CoastlineGenerator* object receives an input dataset in geoJson format and returns a new dataset with the topographic mosaic (TT) of each of the geometries of the original dataset, complemented with the necessary metadata so that it can be exported in geojson format (see Figure 6). The *GridConnector* object receives



(a)Original river geometry



(b) TT with one row



(c)TT with three rows

Fig. 4. Topographical tessellations for a river

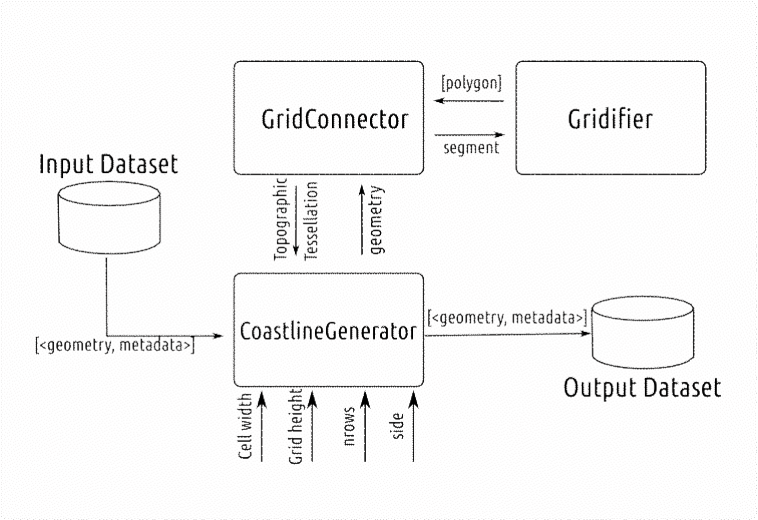


Fig. 5. Coastline generator architecture

	geometry	cid	pos	gid	source_object	source_gna
0	POLYGON ((-69.72500 -36.65485, -69.72519 -36.6...	0	left	264890	Acueducto	Acueducto
1	POLYGON ((-69.72519 -36.65486, -69.72538 -36.6...	1	left	264890	Acueducto	Acueducto
2	POLYGON ((-69.72538 -36.65487, -69.72557 -36.6...	2	left	264890	Acueducto	Acueducto
3	POLYGON ((-69.72557 -36.65488, -69.72572 -36.6...	3	left	264890	Acueducto	Acueducto
4	POLYGON ((-69.72572 -36.65492, -69.72586 -36.6...	4	left	264890	Acueducto	Acueducto
..	...	...	...	...	...	...
21	POLYGON ((-69.72803 -36.65463, -69.72826 -36.6...	21	right	264890	Acueducto	Acueducto
22	POLYGON ((-69.72824 -36.65480, -69.72857 -36.6...	22	right	264890	Acueducto	Acueducto
23	POLYGON ((-69.72826 -36.65472, -69.72858 -36.6...	23	right	264890	Acueducto	Acueducto
24	POLYGON ((-69.72857 -36.65488, -69.72886 -36.6...	24	right	264890	Acueducto	Acueducto
25	POLYGON ((-69.72858 -36.65479, -69.72886 -36.6...	25	right	264890	Acueducto	Acueducto

Fig. 6. Output dataframe

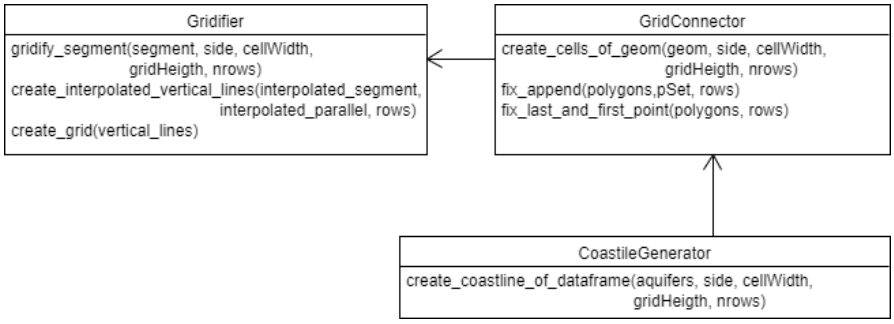


Fig. 7. UML Class diagram

a complex geometry (composed of several segments) that can be multiline or polygon, collaborates with Gridifier to generate the grids of each segment and then joins each grid with the adjacent one. The *Gridifier* object is responsible for generating a grid of cells from a segment, with the parameters that define the grid: side, width of the cells, number of rows, total height of the grid.

3.2 Gridifier

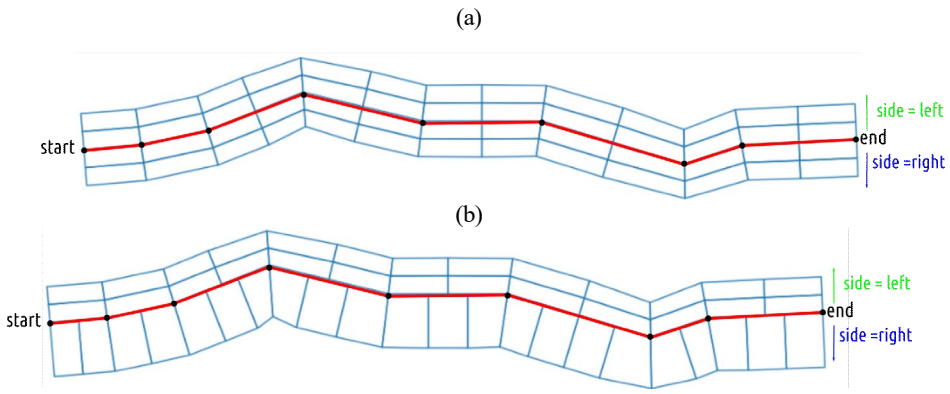


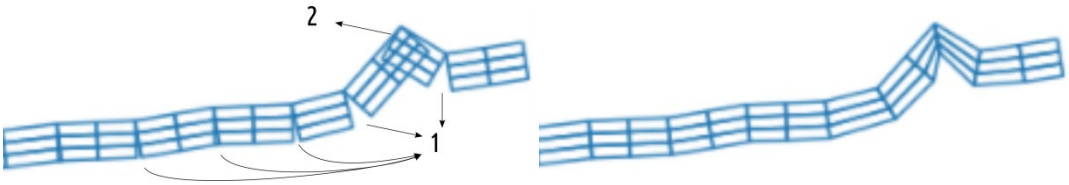
Fig. 8. TT of a river with same parameters (a) and different parametes (b) in each side

As was said, this object generates a grid of cells (non overlaping adyacent cells) from a given segment that is received as a parameter. It also needs the position where the cells are generated (side parameter), the width of each cell (cellWidth parameter), the total transversal distance (gridHeight parameter) to the original segment, and the number of rows (nrows parameter) of cells of the grid.
The side parameter, which possible values are {left, right} indicates the orientation of the grid with respect to the sequence of points in the input segment. The generated grid is going to be limited by the input segment and a parallel line distanced from the first as indicated by the parameter gridHeight. An example is shown in Figure 8 (a), where the red line indicates de input geometry of a river, and travelling from the start point to the end point, left side is the upper grid of the figure, while right side is a the lower one, and both grids are built from the same input parameters. In figure 8 (b) left side grid has 2 rows, a cell width of **40** meters and a grid height of **20** meters, while right side grid has 1 row, a cell width of **20** meters and a grid height of **30** meters.
Generating the grid requires calculating the length of said segment to obtain the number of columns, of the required size, that this segment admits. On the other hand, a parallel line is generated using the method called paralleloffset
provided by the Shapely[Tecnologia] library, located at the distance indicated by the parameter gridHeigth and the side. Both segments are subdivided to find the lateral vertices of each cell, that will define the vertical lines connecting the two parallel segments. If the parameter nrows is greater than one, these verticals are also partitioned to set the rows of the grid. This process generates a list of polygons that correspond to each of the grid cells, and is depicted in the algorithm if Listing 1.1. Following example of Figure 8 (a), each grid is a set of 26 polygons, unlike Figure 8 (b) that left side grid has 26 polygons and right side grid has 22 poligons.

```
gridify (segment , side , cellWidth , grid Heigth , nrows)
    n_columns  = (segment. geometry_length () // cellWidth ) + 1
  parallel  = paralleloffset (segment , side , grid Heigth )
  interpolated_segment  = segment. interpolate (n_columns)
  interpolated_parallel  = parallel. interpolate (n_columns)
  vertical_lines  = create_interpolated_vertical_lines (
    interpolated_segment , interpolated_parallel , nrows)
  polygons  = create_grid (vertical_lines )
  return  polygons
```

Listing 1.1. Topographic Tessellation algorithm

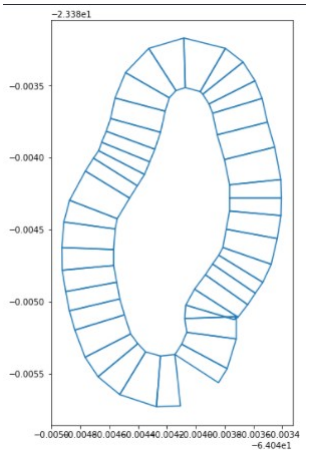
3.3 GridConnector



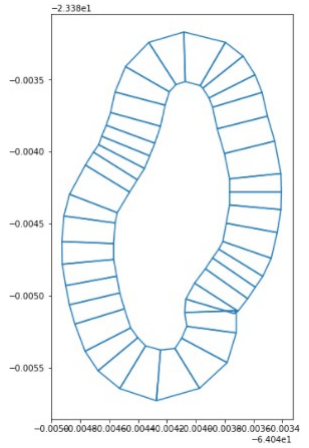
(a) Overlaping (marked as 2) and separation prob- (b) Final TT with interpolated points lems (marked as 1)

Fig. 9. GridConnector fixAppend issue solved

The main function of this object is to connect the grids that were generated for each segment -of a MultiLineString or Polygon- by the *Gridifier* object. With this objective, it takes the input geometry, divides it into segments (see Listing 1.2, line 5), i.e. taking two points at a time. Then, it collaborates with the Gridifier object to generate the grid of each segment, finally connects the vertices of the generated grids obtaining a topographic tessellation. If the original geometry has polygonal type, the cells corresponding to the space between the last and the first point are created.



(a) TT over polygon without fixLastAndFirstPoint



(b) TT over polygon with fixLastAndFirstPoint

Fig. 10. Polygons TT completed

Since the grids generated by the *Gridifier* object are generated from simple segments, they have a rectangular structure. However, each vertex of the polyline generates an angle between the two consecutive segments, and therefore an angle between the respective grids, which leads to problems of overlapping or separation in the path of the TTs. This two problems are presented in an example of a TT from a multiline in Figure 9 (a): with number 1 a separation problem is indicated, and with number 2 an overlapping problem. To solve this, a mechanism was developed to replace conflicting vertices with new vertices resulting from an interpolation process, as shown in 9 (b). This mechanism, encapsulated in the fixAppend method of the *GridConnector* object, also allows the cells corresponding to the closure of a polygons object to be

generated, as can be seen in Figure 10.

```
1  create_cells_of_geom (geom, side, cellWidth, gridHeigth, nrows
       ):
2    polygons = []
3    For i in geom.index:
4        segment = LineString (geom[i:i+1])
5        pSet = Gridifier. gridify_segment (segment, side,
       cellWidth, gridHeigth, nrows)
6        polygons = fix_append (polygons,pSet, nrows)
7    if ((geom.type == 'MultiPolygon') | (geom.type == 'Polygon'
       )):
8        polygons = fix_last_and_first_point (polygons, rows)
9    res = gpd.GeoDataFrame ({" geometry " : polygons })
10   res['cid'] = res. index
11   res['side'] = side
12   return res
```

Listing 1.2. GridConnector algorithm

The dataset generated by GridConnector consists of a row for each new polygon generated (representing a cell of the grid), whose identification is a data composed of a value cid, which is an integer value between 0 and the number of cells of the TT, and the side value. As an example, consider the multiline geometry of Figure 8 (b): the generated dataset has 26 polygons with side=left and cid$\in$ [0..25], and 22 polygons with side=right and cid$\in$ [0..21].

3.4 CoastlineGenerator

The purpose of this object is to generate a Topographical Tessellation for each element of an homogeneous geopraphic dataset, this means that all elements have the same geometry type (multiline, line or polygon). the output dataset contains a set of polygons that represent the cells of all the tessellations. These polygons that correspond to the same input geographic element are related by the **gid** data that identifies its geometry (river or lake) that was used to generate the TT.
Particularly, the CoaslineGenerator iterates over the dataset keeping certain alphanumeric data (metadata) that are of interest in this domain: the **GID**, the **object** and the **GNA** of the geometric object (see Listing 1.3, line 3), and then it generates the TTs based on the **side** parameter. In the case of rivers represented by a multiline, it is important to generate the TT of both sides (side=both).
After creating the TT on one element of the dataset, it adds the metadata to each generated cell. To identify each one, a composed identifier made up by the **GID**, the CID and the side parameter is used (see Listing 1.3, line 13).

```
create_coastline_of_dataframe (aquifer, side, cellWidth,
    gridHeigth, nrows)
for row in aquifer:
    metadata = row.metadata
    geom = row.geometry
    if(side == "both"):
        rside = gridConnector.create_cells_of_geom (geom, "
    right", cellWidth, gridHeigth, nrows)
        lside = gridConnector.create_cells_of_geom (geom, "
    left", cellWidth, gridHeigth, nrows)
        cells = pd.concat([rside,lside])
    else:
        cells = gridConnector.create_cells_of_geom (geom, side
    , cellWidth, gridHeigth, nrows)

    cells[objeto],cells[gna] = metadata.objeto, metadata.gna
    cells[gid] = metadata.gid()
```

Listing 1.3. Topographic Tessellation algorithm

3.5 Tecnology

For the processing of geospatial information, GeoPandas[1] library was used. GeoPandas is an open source project that extends the data types used by Pandas to allow spatial operations on geometric types. In addition, GeoPandas uses shapely as a geometric data manipulation and analysis library.

To specify the reference system, PyProj was used, a Python interface for PROJ. PROJ is generic coordinate transformation software that transforms geospatial coordinates from one coordinate reference system (CRS) to another. This includes map projections and geodetic transformations.

The source code developed for this work is available as a python notebook on the Kaggle platform[2].

4 Discussion and Future work

This work presented an strategy to generate a topographic tessellation from a given multiline or polygon geometry. Despite it was an ad-hoc development for the AppEAR project, this can be usefull for the application in other domains such as urban planning, or any situation where to represent or analyze the area around a geometry is needed.

The approach that was developed in this work can be considered as a cold start for an adaptation strategy based on the behavior of the community. That is, an initial set of

[1] https://geopandas.org/en/stable/

[2] https://www.kaggle.com/brunolattanzio/coastlinecellsgenerator/edit/run/72445421

areas that can then be adjusted by calculating Voronoi regions taking as centroids the samples or from other hot spots that represent the busiest areas. This complemented approach allows having a dynamic set of sampling areas to propose a better distribution of the samples.
The segmentation of space within a polygon -which could represent an island- is scheduled for future work, as is the task of performing a topographic tessellation around a single point.
Unlike the polygons that delimit lakes, where the TTs are generated with the parameter side=right so that they correspond to the outer side of the lake, the polygons of the islands must be interpreted in the opposite way to generate the TTs inside the polygon. In other words, if the coast of a lake is normally generated from the right (external) side, instead in an island it should be generated from the left (internal) side.
On the other hand, as the used cartography used is about rivers and lakes, it doesn't have objects representing islands, meaning that there is no direct and automatic way to identify when an island is present. This means that these cases must be addressed individually, this is, find -by observation- the island and generate its TT coast on the internal side, which would be the left.
In this work, both linear and polygonal geometries are considered, but so far those that are only composed of a point, are not taken into account. In the AppEar project case study these would correspond to, as an example, waterfalls, jumps or dams. The main reason why they have not been taken into account is that, despite being an individual geometric object, they are usually part of another object. That is to say, although the algorithm cannot generate a coast for a waterfall, it can do it for the river of which the waterfall is part of.
However, methods have been thought to address this problem in future works, such as defining a circle around the point and dividing it using a similar mechanism to that of the other geometries.

References

1. Dalponte Ayastuy, M., Torres, D.: Relevance of non-activity representation in traveling user behavior profiling for adaptive gamification. In: Proceedings of the XXI International Conference on Human Computer Interaction. Interaccion '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3471391.3471431
2. Du, Q., Gunzburger, M.: Grid generation and optimization based on centroidal voronoi tessellations. Applied Mathematics and Computation **133**(2), 591–607 (2002). https://doi.org/10.1016/S0096-3003(01)00260-0
3. Fleischmann, M., Feliciotti, A., Romice, O., Porta, S.: Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. Computers, Environment and Urban Systems **80**, 101441 (2020). https://doi.org/10.1016/j.compenvurbsys.2019.101441
4. Gold, C.: Tessellations in gis: Part i—putting it all together. Geo-spatial Informa- tion Science **19**(1), 9–25 (2016). https://doi.org/10.1080/10095020.2016.1146440
5. Gold, C.M., Remmele, P.R., Roos, T.: Voronoi methods in GIS, pp. 21–35. Springer Berlin Heidelberg, Berlin, Heidelberg (1997). https://doi.org/10.1007/3-540-63818-02
6. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: Geographic informa- tion science and systems. John Wiley & Sons (2015)

7. Silvertown, J.: A new dawn for citizen science. Trends in Ecology & Evolution **24**(9), 467–471 (2009). https://doi.org/10.1016/j.tree.2009.03.017
8. Tsai, V.J.D.: Delaunay triangulations in tin creation: an overview and a linear-time algorithm. International Journal of Geographical Information Systems **7**(6), 501–524 (1993). https://doi.org/10.1080/02693799308901979
9. Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., Wagenknecht, K.: The science of citizen science. Springer Nature (2021)

FoodCraft: Design of a Precision Agriculture System with IoT in Indigenous Communities in Rural Areas with Difficult Internet Access in the Department of Cauca-Colombia

Juan Sebastián Orozco Monje[1][0000-0001-7426-5536], José Javier Arce Santacruz[2][0000-0002-5080-2461] and Gineth Magaly Cerón Rios[3][0000-0002-9979-2137]

[1] Universidad del Cauca, Calle 5 #4-70, Colombia
juanmon@unicauca.edu.co
[2] Fundación Universitaria de Popayán, Calle 5 #8-58, Colombia
arceman@gmail.com
[3] Corporación Universitaria Unicomfacauca, Calle 5 #4-70, Colombia
gceron@unicomfacauca.edu.co

Abstract Nowadays, many countries have been implementing precision agriculture systems based on the Internet of Things (IoT) as they are useful tools to manage and monitor crop information for farmers. There is a wide diversity of approaches that help and facilitate producers to manage their agricultural land. IoT systems provide capabilities to adapt to the user's environment, for example by sensing data through wireless sensors or other devices. In agriculture, they can support the early detection of possible crop alterations and prevent production losses, thus improving the quantity and quality of the harvest, considering that each crop requires a personalized intervention. This research aims to propose an IoT system for the management and monitoring of the agricultural lands of the indigenous communities of Cauca, where Internet access is limited. The system adapts to the needs of the user, their agricultural and environmental information, time, location and traditional practices. This article introduces the IoT architecture and models, which were used to implement and test a prototype called Food-Craft. For the design of this system, there is satellite communication between IoT points with difficult access to the Internet. On the other hand, all sensor data is successfully received by the Arduino mega and sent to the database with the ESP8266 so that it can be accessed through the integrated Android app and web-site.

Keywords. Sensors, IoT, agriculture.

1 Introduction

Precision agriculture systems supported by IoT are a type of system that adapts to the current circumstances of agricultural land and the environment, providing accurate real-time information on different variables, time and resources. Data can be obtained from wireless sensors, farmers, handheld measuring devices, or other devices[1]. IoT systems understand the process of taking data in real time based on the agricultural context where they are located, and when the context changes, the agricultural variables can also vary [2] Precision agriculture systems have great potential to support management and crop monitoring, for example, generating early warnings of changes in soil minerals, or temperature, allowing the farmer to make decisions to treat the problem in time or, in the best case, prevent damage to the crop. Examples of precision agricultural systems are monitoring of environmental parameters [1], IoT communication system[3], Smart irrigation system[4] IoT in the agricultural sector[5], they are IoT systems that are responsible for communicating and managing information of agricultural land to reduce the effects of environmental changes, lack of water and poor communication, based on the form of data collection. LoRa-LBO [6] focuses on the evaluation of protocols for low-power wireless sensors such as LoRaWAN, which facilitates the collection of crop information, the correlation of variables and their analysis to generateknowledge.

This article proposes a computational architecture, as well as the design of an IoT prototype for data collection and information visualization in a graphic and easy to understand way, as well as the generation of early warnings against possible alterations in agricultural and environmental variables. Based on the proposed architecture, the communication system based on a test satellite modem and the IoT sensor network is implemented, capable of providing real-time information and knowledge of soil minerals,temperature, amount of water and the location, as well as the activation of alarms. On the other hand, there is a web application supported by: ThingSpeak and firebase for information storage.

2 Methods

IoT-enabled agriculture systems help users to focus more on their tasks while reducing human-computer interaction, because the system automatically integrates agricultural and environmental information from crops to monitor environment variables and the needs of farmers. crops and farmers [7][8] [9][10] [11][12] [13][14] therefore, the architecture of this type of system must be based on continuous data communication in real time, the formal definition of the correlation of context variables, IoT models and information visualization friendly and easily accessible by communities in rural areas.

Based on the generic models of IoT systems proposed and the derived computational architecture, it is possible to implement an adaptable dashboard based on data collected by sensors and data detected from mobile phones or manually. The interface design

processes were based on the UCD (User Centered Design) methodology[15], [16]. The process involves the future users (indigenous communities) of the system throughout the design and development process, and is an iterative method. UCD focuses on a deep understanding of who will use the IoT system, that is, it is based on an explicit understanding of indigenous people and their knowledge of technologies, tasks and environments. It is driven by user-centric evaluation and considers the user experience in its entirety. The first step in the development process is exploring the context, then specifying the design criteria or requirements. Then, the design decisions, technologies, communications among others that can be addressed. Fifth, specification of the prototype and finally evaluation of the pilot through usability instruments, such as SUS (Sys-tem Usability Scale) [17].

3 Architecture

3.1 Computational architecture

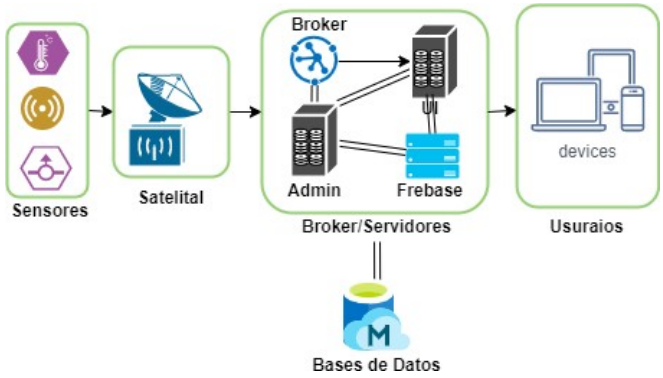


Fig. 1. Architecture of the IoT agriculture system (source: own)

Meters/Sensors: using different detection and communication technologies, etc., NPK is used to measure the nutrients of the land in order to guarantee farmers greater efficiency and quality in the crops [18]. GY-91 sensor (BMP280 + MPU 9250): This is a very complete sensor, as it is capable of measuring pressure, temperature, humidity, altitude, it also has a magnetometer, accelerometer and gyroscope. Although the last three data are not relevant for this project, it is valid to clarify that they have a fairly broad functionality. providing information to generate early warnings [19][20]. GPS(Global Positioning Systems) could be used to give locations of crops that require prompt attention[21].

Meter/broker list: Intermediate and control module, managing the communication between the client and the presentation layers. In the implemented prototype, the

broker supports three (3) types of communications: 1) GPS to monitor specific locations. It requires the availability of that technology in the wireless device and sensor network. 2) NPK and GY-91 to monitor nutrients and other variables. 3) Cloud communication as the main communication channel, available for all devices, counts for wireless communication with ESP8266. Allows data synchronization with the database through the application server layer. It also shows the data obtained and the alerts.

ADMIN/Application Server: Performs the business logic of the system and connects the presentation layer and the database layer. In addition, it acts as anintermediary for other servers and databases.

Google firebase/cloud services: This component stores groups of media content that are hosted on online servers owned by organizations outside theproject. Its function is to collect and organize data that has been sensed withlocation technologies and stored in the context database, taking into accountthe data that is relevant to the crop. This data is sent to the smart server component.

UI/Intelligent Server: The task of this component is to receive data from the context server, process it, make inferences and communicate with the content server to generate alerts and recommend information according to the context. In addition, visualize the data in a usable dashboard.

Database server: This component manages the databases, performs CRUD (Create Read Update Delete) operations, and coordinates the synchronization of the data stored in the offline database (information in the memory of the wireless device) with general databases. Includes the databases represented in Figure 1.

3.2 Satellite Communication

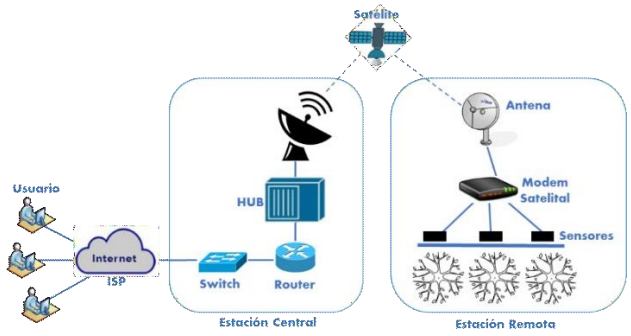


Fig. 2. Satellite network IoT support (source: own)

One of the advantages of satellite communications is to establish a connection in places with difficult access. In **Figure 2**, a satellite network topology is described in a general way. The remote station has an antenna which points to a specific satellite and a modem that converts the RF signal to IP (VSAT System). Similarly, at the central level there is a master antenna and the Satellite Hub.
For this topology, the earth station has a connection to an ISP, allowing it to providean Internet service to remote clients. For example, for an IoT system in the remote station, a sensor system can be implemented which allows different temperature and humidity environments to be controlled and an action to be executed, where the user will have visualization and control at a central level. This topology will be used in places where Internet connectivity is limited and it is required to take only data, it does not have more characteristics derived from the satellite because it is an approval of a simple modem and it must be low cost so that it is easy to use. access by the community and be sustainable over time and generate sustainable agriculture.

4 FoodCraft

This precision agriculture system with IoT in indigenous communities in rural areas with difficult access to the Internet in the department of Cauca-Colombia, stems from a need of indigenous cultures to preserve their survival, knowledge and cultural traditions. These communities have been apathetic to the insertion of new technologies. Onthe other hand, the lack of connectivity in areas with difficult Internet access has limited the use of technological applications in these ancestral territories. Today, with technological advances and the expansion of Internet coverage in many regions, it has opened an opportunity to use ICTs, for the benefit of society, sustainable development and the quality of life of people by conserving their agriculture and without being invasive in their own knowledge and creative processes. The user interface (UI) was designed following the UCD (User Centered Design) methodology. When designing the user interface for Foodcraft, the first step was choosing the ThingSpeak and blynk mobile app. The next step was how to capture the data, the position of graphics, text boxes and messages in the interface.
The resulting interfaces are described below. Foodcraft uses communication modules like ESP8266 and the Max485 to get the best quality data. In this investigation, ithas been connected to the firebase system database for storage.
This research focuses on the collection of data via IoT for the measurement of temperature, pressure, location and soil nutrients, focused on precision agriculture, in such a way that the efficiency of crops with live seeds is improved. Survey the acidity of the soil to determine the state of the land before cultivating to generate a fertilization plan based on phosphorus and potassium, with which it is expected to help the agricultural sector to cultivate plants of better quality, better nutrients and have more precise control of plant development. In such a way that a system capable of measuring physical variables related to the crop such as relative humidity, soil PH is designed, in additionthe system is capable of transmitting in approximate real time the information

obtained by the sensors thanks to the communication modules and the servers of the ThingSpeakpage.
The system has 2 types of GY-91 and NPK sensors mainly and the ESP 8266 wireless connection module.
Pressure, temperature and water level data **GY-91**, the programming of this sensor turns out to be problematic, since several factors must be taken into account. Especially the libraries available for such programming, as they commonly fail. In addition to that it is recommended to use I 2 C for its configuration. This is a very complete sensor, as it is capable of measuring pressure, temperature, humidity, altitude, it also has a magnetometer, accelerometer and gyroscope.

For the measurement of the nutrients N (nitrogen), P (phosphorus) and K (potassium), which allows determining the percentage and amount of additional nutrients that must be added to the soil to increase soil fertility and have better quality crops. The measurement of soil nutrients allows to accurately determine the nutritional deficiency or excess of the components of the land destined for agriculture.
There are several methods to determine the amount of nutrients in the soil, among them the use of spectrometers and optical sensors. However, the accuracy of the measured data ranges from 60-70%. Comparing such data with chemical methods, it is difficult to determine the effectiveness of the precision. Well, the amount of data is scarce.

NPK Sensor: This sensor is highly accurate, portable and capable of taking large amounts of data with considerable speed, as well as being 100% electronic and does not require chemical reagents for its operation. It can be used with almost any microcontroller (ESP32, Raspberry, Arduino, etc.) which is why an RS485 module must be implemented.

MAX485 TTL to RS-485 interface

This module is responsible for using the device's digital communication signaling (NPK sensor) that works in a highly noisy environment and allows data to be transmitted over long distances at a maximum rate of 2.5Mbit /sec This is a multipoint module, which allows you to connect multiple devices to it. The connection is quite simple. Forthis, an Arduino Mega 2560 has been used, which connects to the RS485 module, whichin turn communicates with the NPK sensor.

ESP8266 module: For the development of the project it was necessary to use I²C (inter-integrated circuit), as this allows the use of only two ports, the input and output ports, in addition to the voltage and ground. Using I²C is beneficial for the project, as it greatly simplifies the processes and reduces the complexity of the project.

5 Evaluation

The tests carried out with the sensors used and the results obtained are presented

below. These results can be seen on the Blynk and ThingSpeak screens.

GY-91 sensor tests. Successful data capture tests were performed, temperature, altitude and pressure were taken. To visualize said data, the serial monitor of the ArduinoIDE is used. Thus, the data obtained as a function of time are:

- Temperature: Temperature is one of the data that presents the most variationsat the time of measurement, since it changes a lot depending on the time of day in which it is found (see **Figure *3***). Thus noticing that at noon (12:00) the temperature peak occurs, which has come close to 26°C. While the lowest temperatures have been recorded at approximately 03:00, where tempera- tures of 8 and 7°C, respectively, have been reported.

Fig. 4. Temperature data (source: own)

- Atmospheric pressure: The pressure is given in millibars (mb). It is easy to see that the atmospheric pressure data does not vary much as a function of time. Which shows a relative stability of the sensor, as well as its precision at the time of measuring the data, since these have been compared with pages such as accuweather.com that show real-time data regarding temperature, altitude, atmospheric pressure. Key data to compare and calibrate the sensors (see **Figure *4***).

Fig. 5. Atmospheric pressure data at different instants of time (source: own)

- Water level: see **Figure 5**.

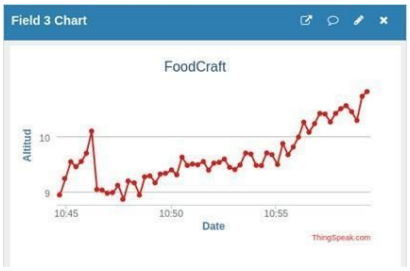


Fig. 6. Water level data different instants of time (source: own)

NPK sensor tests. The results obtained from carrying out the design tests in different soil conditions, varying the humidity, soil and depth, are shown.

For this, some considerations have been taken into account. First, the data must have been taken on the same day to avoid large variations. Second, do not alter the conditions during the measurement (add water or other components that can modify the data). In such a way that an accurate and reliable measurement can be guaranteed.

For this analysis, the first 10 samples delivered by the system have been taken as a reference. Thus, the samples were made in the following way: the first test (see **Figure 6**) was carried out with conventional soil from a park, burying the sensor 6 cm into the ground. The second test (see **Figure 7**) has been carried out by extracting soil from a depth of 15 cm. The third test (see **Figure 8**) was carried out with soil 15 cm deep, but adding plenty of water and salt to alter the components of the soil.

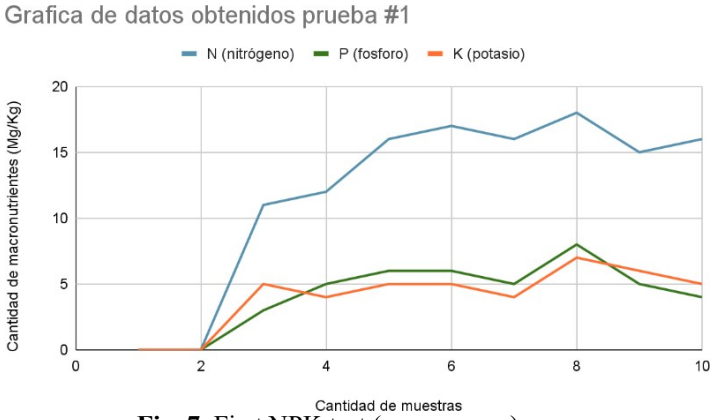


Fig. 7. First NPK test (source: own)

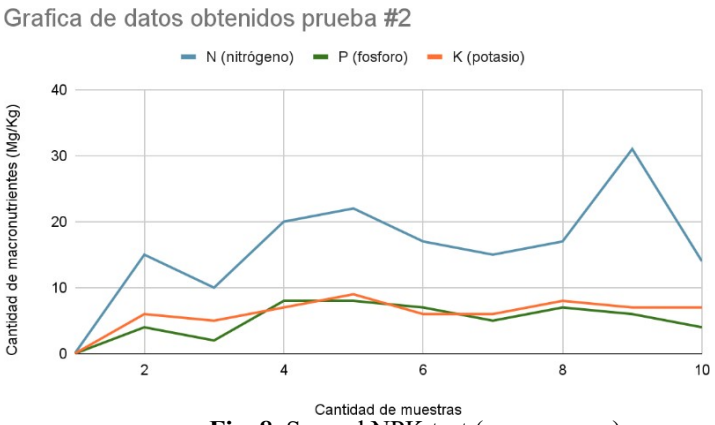


Fig. 8. Second NPK test (source: own)

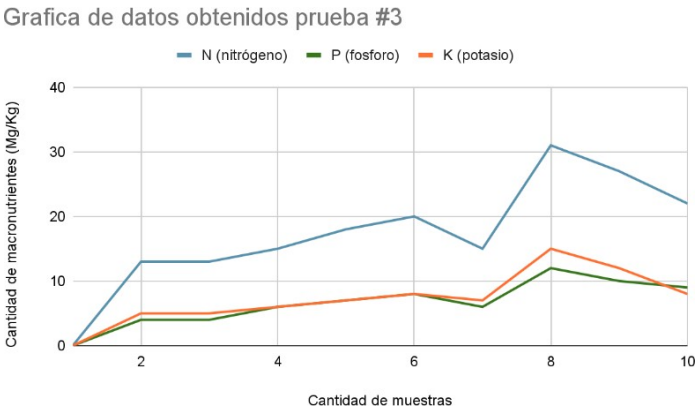


Fig. 9. Third NPK test (source: own)

Based on the data provided by the system, it is possible to see how the first samples deliver low values of each element, this is due to the time the sensor needs to calibrate. In this process, the sensor takes data from scratch, to later deliver high data and finally stabilize its measurement at more precise and reliable values.

- ***Comparison of the results:***

Comparing the average of the values obtained by the three samples (see ***Table 1***). Which shows that the last test shows a better average.

Table 1. Comparative results

Variable	Average		
Nutrients	**Test 1**	**Test 2**	**Test 3**
N(Nitrogen)	12.1	14.5	17.4
P (Phosphorus)	4.2	5.125	6.6
K (Potassium)	4.1	5.875	7.3

The average of the data delivered in the three tests carried out for each of the nutrients of the NPK sensor does not show a variation according to what was expected when changing the soil and altering the sample with external agents. However, it was possible to appreciate that the components are usually relatively constant in a certain range of values.

6 Conclusions

This research allows to improve the implementation of crop management and monitoring systems in rural economies and the quality of plant cultivation in Cauca, it is also necessary to open doors, in other rural areas of Colombia, in the creation and innovation of management and monitoring systems, being easier and more suitable in the process of precision planting with excellent quality.

Some of the advantages of using this sensor over other types of sensors are the speed and abundance of data that are taken, their precision and that the sensor is a low-cost device as well as being ideal for being connected to cloud systems. With IoT technology, where the data can be stored to carry out a soil analysis. What makes it ideal for detecting different types of soil; such as acids, alkalines, coconut bran, compost, among others.

This system allows productivity to be increased, under the correct conditions, thanks to the fact that, by monitoring the crops, risks are reduced, or unnecessary high supplies of water, land or energy, such as those that are still used today, which are very conventional. The production of different types of agricultural food is much more generous with the environment under this method, at the same time it can be said that the level at which these crops are produced has increased.

Automation allows better control and monitoring of this type of system in rural areas, since, with satellite communications, without being present in a tangible way, it is possible to monitor and control this type of system, which makes it more efficient.

Additionally, an IoT system contributes to improving the quality of the final products, since the environmental conditions are managed to continuously have the correct ones,

helping to have a better performance in productivity.

Agriculture can be developed sustainably and can become the way food is produced in the future, and can be had by anyone in the world no matter how distant they are and at low cost to benefit.

Acknowledgments

We thank the Project "Consolidation of Science, Technology and Innovation Capacities to Strengthen Rural Economies in the Ancestral Territories of the Kokonuko People-Cauca" financed with resources from the *Sistema General de Regalias* and the ancestralterritory of the Indigenous Community of Puracé.

References

1. P. Placidi, R. Morbidelli, D. Fortunati, N. Papini, F. Gobbi, and A. Scorzoni, "Monitor-ing Soil and Ambient Parameters in the IoT Precision Agriculture Scenario: An Original Modeling Approach Dedicated to Low-Cost Soil Water Content Sensors," *Sensors 2021, Vol. 21, Page 5110*, vol. 21, no. 15, p. 5110, Jul. 2021, doi: 10.3390/S21155110
2. M. S. Farooq, S. Riaz, A. Abid, T. Umer, and Y. bin Zikria, "Role of IoT Technology in Agriculture: A Systematic Literature Review," *Electronics 2020, Vol. 9, Page 319*, vol. 9, no. 2, p. 319, Feb. 2020, doi: 10.3390/ELECTRONICS9020319
3. A. P. Atmaja, A. E. Hakim, A. P. A. Wibowo, and L. A. Pratama, "Communication Systems of Smart Agriculture Based on Wireless Sensor Networks in IoT," *Journal of Robotics and Control (JRC)*, vol. 2, no. 4, pp. 297–301, Jul. 2021, doi: 10.18196/JRC.2495
4. L. García, L. Parra, J. M. Jimenez, J. Lloret, and P. Lorenz, "IoT-Based Smart Irrigation Systems: An Overview on the Recent Trends on Sensors and IoT Systems for Irrigation in Precision Agriculture," *Sensors 2020, Vol. 20, Page 1042*, vol. 20, no. 4, p. 1042, Feb. 2020, doi: 10.3390/S20041042.
5. E. Cornejo-Velázquez, H. Romero-Trejo, O. A. Acevedo-Sandoval, and A. Toriz-Pala-cios, "Aplicación del Internet de las Cosas en el Sector Agrícola," *Pädi Boletín Cientí-fico de Ciencias Básicas e Ingenierías del ICBI*, vol. 7, no. 13, pp. 62–67, Jul. 2019, doi: 10.29057/ICBI.V7I13.4403
6. M. Swain, D. Zimon, R. Singh, M. F. Hashmi, M. Rashid, and S. Hakak, "LoRa-LBO: An Experimental Analysis of LoRa Link Budget Optimization in Custom Build IoT TestBed for Agriculture 4.0," *Agronomy 2021, Vol. 11, Page 820*, vol. 11, no. 5, p. 820, Apr. 2021, doi: 10.3390/AGRONOMY11050820
7. O. Debauche *et al.*, "Data management and internet of things : A methodological review in smart farming," *Internet of Things*, vol. 14, p. 100378, Jun. 2021, doi: 10.1016/J.IOT.2021.100378
8. "Arquitectura de radio cognitiva para servicios masivos de IoT con acceso dinámico al espectro." https://repository.javeriana.edu.co/handle/10554/57770 (accessed Nov. 15, 2021).
9. I. Ullah and D. H. Kim, "An optimization scheme for water pump control in smart fishfarm with efficient energy consumption," *Processes*, vol. 6, no. 6, Jun. 2018, doi:

10.3390/PR6060065.
10. K. R. Chozo Juárez and I. en Telecomunicaciones, "Metodología de agricultura urbana distribuida empleando internet de las cosas en los proyectos de investigación aplicada a seguridad alimentaria del grupo Gidatic de la UPB," 2021, Accessed: Aug. 22, 2021. [Online]. Available: https://repository.upb.edu.co/handle/20.500.11912/8194
11. E. Schiavon, A. Taramelli, A. Tornato, and F. Pierangeli, "Monitoring environmental and climate goals for European agriculture: User perspectives on the optimization of the Copernicus evolution offer," *Journal of Environmental Management*, vol. 296, p. 113121, Oct. 2021, doi: 10.1016/J.JENVMAN.2021.113121
12. Y. Li *et al.*, "Toward Location-Enabled IoT (LE-IoT): IoT Positioning Techniques, Er- ror Sources, and Error Mitigation," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4035–4062, Mar. 2021, doi: 10.1109/JIOT.2020.3019199
13. Y. Song, F. R. Yu, L. Zhou, X. Yang, and Z. He, "Applications of the Internet of Things(IoT) in Smart Logistics: A Comprehensive Survey," *IEEE Internet of Things Journal*,vol. 8, no. 6, pp. 4250–4274, Mar. 2021, doi: 10.1109/JIOT.2020.3034385
14. A. Bouyedda, B. Barelaud, and L. Gineste, "Design and Realization of an UHF Fre- quency Reconfigurable Antenna for Hybrid Connectivity LPWAN and LEO Satellite Networks," *Sensors 2021, Vol. 21, Page 5466*, vol. 21, no. 16, p. 5466, Aug. 2021, doi: 10.3390/S21165466
15. E. Schiavon, A. Taramelli, A. Tornato, and F. Pierangeli, "Monitoring environmental and climate goals for European agriculture: User perspectives on the optimization of the Copernicus evolution offer," *Journal of Environmental Management*, vol. 296, p. 113121, Oct. 2021, doi: 10.1016/J.JENVMAN.2021.113121
16. B. Ortiz-Crespo *et al.*, "User-centred design of a digital advisory service: enhancing public agricultural extension for sustainable intensification in Tanzania," vol. 19, no. 5–6, pp. 566–582, 2020, doi: 10.1080/14735903.2020.1720474
17. P. Vlachogianni and N. Tselios, "Perceived usability evaluation of educational technol-ogy using the System Usability Scale (SUS): A systematic review," https://doi.org/10.1080/15391523.2020.1867938,2021
18. R. P. Potdar, M. M. Shirolkar, A. J. Verma, P. S. More, and A. Kulkarni, "Determination of soil nutrients (NPK) using optical methods: a mini review", vol. 44, no. 12, pp. 1826–1839, 2021,doi: 10.1080/01904167.2021.1884702
19. G. Soni, "Implementation of LM35 Interfacing of Temperature Sensor with Arduino using LabVIEW 2015," *2022 IEEE Delhi Section Conference (DELCON)*, pp. 1–3, Feb.2022, doi: 10.1109/DELCON54057.2022.9753030
20. B. Deepa, C. Anusha, and P. Chaya Devi, "Smart Agriculture Using IOT," *Advances in Intelligent Systems and Computing*, vol. 1171, pp. 11–19, 2021, doi: 10.1007/978-981-15-5400-1_2
21. J. Arshad *et al.*, "Implementation of a LoRaWAN Based Smart Agriculture Decision Support System for Optimum Crop Yield," *Sustainability 2022, Vol. 14, Page 827*, vol.14, no. 2, p. 827, Jan. 2022, doi: 10.3390/SU14020827

Short papers

Proposal of a System Based on Direct and Indirect Techniques and their Correlation by Chlorophyll Quantification

Kevin Sthiff Vivas Fernández[1], Jhon Álvaro Muñoz Reyes[1], Saúl Eduardo Ruiz Sarzosa[1], Andrés Felipe Solis Pino[1], and Amalfy Anacona Chicangana[2]

[1] Corporación Universitaria Comfacauca - Unicomfacauca, Popayán, Colombia
kevinvivas@unicomfacauca.edu.co,jhonalvaro994@gmail.com, sruiz@unicomfacauca.edu.co, asolis@unicomfacauca.edu.co
[2] Servicio Nacional de Aprendizaje - SENA, Popayán, Colombia
anaconac@sena.edu.co

Abstract. The prediction of normalized vegetation indices in coffee crops using multispectral images obtained by aerial mapping aims to generate a technological strategy using aerial mapping employing drones (RPAS) to predict normalized vegetation index (ENDVI) in coffee crops. During the research process, reference is made to the ENDVI according to the multispectral footprints generated by the different nutrients on the plants in the production stage of the coffee crop, using RPAS for the realization of aerial mapping works in precision agriculture. This reflects the importance of implementing technological tools to improve the planning of agricultural activities, predict damage and decide in situations that affect the development of coffee crops. This study took multispectral images of coffee crops from aerial mapping in the coffee plantations of the Popayan plateau region. It will also analyze the health status of the plants using a chlorophyll meter. From this comparative analysis of the different ENDVI, it is possible to define management alternatives to improve production. However, the images will be captured with unique cameras incorporated in the RPAS, allowing the identification of the variations of the lots and coffee plants in the formative stage of their phenological development, the absorption of nutrients, and the water stress of the crop. Finally, some strategies for integrating expert systems in aerial mapping are proposed.

Keywords. Aerial mapping, Precision agriculture, Spectral index.

1 Introduction

Coffee is one of the most important crops worldwide; it generates an income of over 15 billion dollars and is a source of direct and indirect employment for almost 20 million people. Specifically, in Colombia, coffee is the flagship export product; However, there have been some problems concerning the yield of its production; according to the Federación Nacional de Cafeteros, there has been a decrease in the production of this fruit in recent years, caused by the persistence of unfavorable weather. Besides the above, there are other causes, such as crop renewal, low planting density, and lack of support from state entities regarding technical help [1].

Specifically, in the Cauca department, coffee production is quite rudimentary compared to other geographic zones; this is because, besides the climatic causes mentioned, it can be said that the low productivity is also because of the lack of training, support for the correct implementation of technologies and production and monitoring methodologies based on phenology and evidence of stress in the plants. This makes it possible to identify the farmer's lack of technical competencies and agronomic and extension tools for the management of specialty coffee crops and the scarce technological innovation in their management [2].

In this line, the technologies applied to coffee production could be an excellent alterna-tive to improve fruit production, thus achieving a better capitalization of the production and consequently improving the producers' profit margins. Some technologies that have been applied to coffee production in Cauca are multispectral images, genetic engineering, and biotechnology applied to productivity [3]. Specifically, a study carried out by Cenicafé mentions that there are some activities within the production of the fruit that are carried out manually, which damages the production chain and the speed with which it is carried out; therefore, some alternatives are proposed, such as a portable back vacuum cleaner that is technically and economically viable in Colombian coffee production [4]. Another study shows that soil acidity is the main factor affecting the availability of nutrients, and it should adapt its management according to soil mineralogy. Therefore, in [5], a device is proposed that measures this variable, allowing one to estimate crop production at an optimal scale. Another technology that has positively impacted Caucasian coffee growing is multispectral imaging. In [6], the authors propose a statistical relationship between spectral index and laboratory techniques such as optical spectrophotometry, finding high correlations between these techniques, concluding that it can be an excellent alternative to avoid damaging the coffee when studying chlorophyll.

With the above, it is possible to state that the department in question lacks scientific studies regarding the adoption of technologies in the coffee production processes, limiting coffee growers from promoting their products in the market. In addition, there are no intensified studies on the study of soils, coffee physiology, more productive varieties, pest control, and climatic conditions that allow the definition of alternative solutions to mitigate the conditions expressed [7]. Therefore, this research proposes a quantification system based on indirect and direct techniques for quantifying chlorophyll, which allows the determination of the statistical relationship between these techniques to be later implement in professional systems.

2 Emphasis

Coffee is a complex productive chain that articulates not only the careful processes of transformation and processing of the bean but also the management of the crop, the plant, and the land in which it is grown, affecting the specific characteristics of the quality of the bean and the cup profile, which in the end is the accurate indicator of quality [8]. The Department of Cauca is competitive in the production of high-quality coffees because it has an unbeatable environmental offer that added to the commitment to the activity of over 87 thousand coffee-growing families that cultivate close to 74 thousand hectares of coffee, guarantees an essential contribution to the economic development of the department and the national industry [9]. The internal competitiveness agenda generates a specific productive bet for specialty coffees, which is expected to strengthen the productive system [7]. Therefore, it is proposed to contribute through precision agriculture, specifically aerial cartography images captured through RPAS, the phenological monitoring allowing more accurate planning of the production processes based on the early identification of deficiencies and problems in the crop, which vary according to the microclimatic, agroecological and productivity conditions within the different cultivation lots.
Humanity will have to produce more food than was generated in all past years combined in the next four decades. Otherworld authorities show that food production will have to increase by 70-100% by 2050 to feed the estimated population of 9 billion people by then [10]. A possible solution to this need can be found in precision agriculture, a concept gaining momentum in recent years. Traditional agriculture is based on the premise of homogeneity, in which the processes of land preparation, phytosanitary control, sowing, and harvesting are carried out in the same way throughout the field. Precision agriculture is based on the principle of variability, which recognizes the existence of inequalities in soil properties within the same territory, which require different treat-ment according to their conditions.

3 Materials and methods

The project starts with an experimental design based on pixel quantification of the photographs, looking for the best relationship between illumination and drone flight distance to determine if there is a correlation between the normalized vegetation index obtained from the multispectral images and the samples obtained by direct methods with the chlorophyll meter and spectrophotometer.
The project contemplates the following phases:

- **Phase I:** Implement a docking system for multispectral sensors adaptable to the aerial vehicle**.**
- **Phase II:** Determine standardized vegetation index (ENVI).

- **Phase III:** Evaluate statistical techniques to determine the relationship between the different methods of chlorophyll estimation.

The materials to be used are Phantom 4 drone, CCM 200 chlorophyll meter, GENESYS^{TM} 20 visible spectrophotometer, multispectral camera micasense red-edge-MX, and MX blue. In addition, supplies such as 90% acetone.
For the acquisition of multispectral images, it drew a flight plan up. Afterward, with the chlorophyll meter, the spectral indexes will be taken through the images, then the coffee leaves of the study will be selected, and it will quantify the chlorophyll a and b in the leaves of the coffee plant employing the chlorophyll meter. Finally, the samples (leaves) will be taken to the laboratory for processing using the spectrophotometer.
For optical spectrophotometry, a wave range is used to calculate chlorophyll content between 350-750 nm. For this procedure, the maximum absorbance value of the spectrum obtained must be between 0.5 and 1.5 units. Otherwise, the sample must be further diluted in acetone. Finally, the results are compared with equations (1) and (2) proposed in [11] to determine the chlorophyll content in mg/L..

$$h \quad \overline{3} = 11.93 \;\; 664 - 1.93 \;\; 647 \qquad (1)$$

$$h \quad \overline{} = 20.36 \;\; 647 - 5.50 \;\; 664 \qquad (2)$$

4 Partial Results

After making the CAD design of the prototype (Image 1a) in SolidWorks software based on the dimensions of the camera and drone, the realization of the 3D printing using CAM software, the assembly of the cameras to the drone using the parts obtained (Image 1b) and the tests made in the field, ideal results were obtained concerning to the resistance of the prototypes, guaranteeing the stability of the center of mass of the drone at the time of flight, in addition, the drone's camera is accessible and visible for its respective use.

Image 1. (a) 3D design of support prototype. (b) Assembly of support to the drone.

However, different proposals have been made regarding the algorithm for the determi-nation of spectral index to determine these mathematical equations and perform the radiometric and geometric correction process. The first prototype (Image 2) was deve-loped in Matlab and can determine almost 20 normalized indexes.

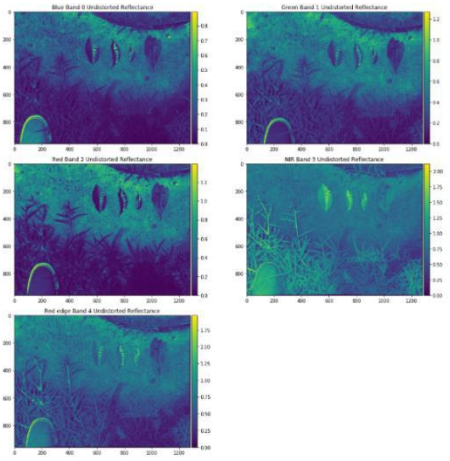


Image 2. Processed spectral images by the proposed algorithm.

5 Conclusions

The proposed docking system for the multispectral cameras is efficient and aerodynamic thanks to the breaks in the sides, considering the strong air currents where the flights will take place. Also, the drone has good maneuverability when tethered with the mounts and the multispectral cameras. Finally, the drone's battery life in flight depletes faster with incorporating supports and sensors. However, it is sufficient for the expected captures per flight.

Regarding the spectrophotometer, initial evidence shows it will obtain linear relationships for indirect sampling using the multispectral cameras attached to the RPAS.

The next activity consists to use of the chlorophyll meter for direct sampling of the chlorophyll value in the coffee tree of the selected area in order to contrast these data with those derived from the normalized vegetation index from the multispectral images obtained by aerial mapping and to estimate the relationship of direct and indirect chlorophyll quantification methods.

Finally, future work includes incorporating expert systems that will allow more efficient determination of spectral indices, thus finding patterns in the measurements of multispectral images.

References

1. G. S. Parra, «Identificación de posibles causas de la baja producción de café en el municipio de Almeida Boyacá.», p. 73.
2. L. F. Salazar, «Capacitación en buenas prácticas agrícolas (BPA) para la producción de café especial para los caficultores de la Vereda El Salado Las Estrellas municipio de Sotará, Departamento del Cauca.», p. 76.
3. C. Dussán-Lubert, H. Duque-Orrego, y J. González-López, «Caracterización tecnológica de caficultores de economía campesina, de los principales municipios cafeteros de Colombia», p. 20.
4. C. E. Oliveros-Tascón y J. R. Sanz-Uribe, «Ingeniería y café en Colombia», *Rev. Ing.*, n.o 33, pp. 99-114, ene. 2011, doi: 10.16924/revinge.33.10.
5. F. E. R. Higuita, «Tendencias modernas en la fertilización eficiente del cafeto para la producción de cafés especiales», *Suelos Ecuat.*, vol. 45, n.o 2, Art. n.o 2, dic. 2015.
6. A. F. Solis, D. A. R. Luna, D. A. C. Ceballos, y C. A. G. López, «Correlación del conte-nido de clorofila foliar de la especie Coffea arabica con índices espectrales en imágenes», *Biotecnol. En El Sect. Agropecu. Agroindustrial*, vol. 19, n.o 2, Art. n.o 2, ene. 2021, doi:10.18684/bsaa.v19.n2.2021.1536.
7. C. J. Pacheco Sosa y A. J. Ceballos Freire, «Factores sociales y económicos que inciden en la adopción tecnológica en el cultivo de café como estrategia para el desarrollo soste-nible local. Caso de la vereda San Pedro. Municipio Piendamó-Cauca», *Univ. Autónoma Manizales*, 2012, Accedido: 16 de mayo de 2022. [En línea]. Disponible en: https://repo-sitorio.autonoma.edu.co/handle/11182/226. Accedido: 16 de mayo de 2022. [En línea]. Disponible en: https://federa-ciondecafeteros.org/static/files/Cauca09.pdf
8. F. B. Valera, O. F. Torrens, y A. Peña, «Evaluación agronómica con técnicas de agricul-tura de precisión en parcelas de café (Coffea arabica) en la cuenca del río Castán. Estado Trujillo agronomic evaluation with technical of agriculture of precision in coffee (Arabic coffea) crop fields in the Castán river basin. Trujillo State», p. 22, 2011.
9. V. E. Oliva, J. L. Blanco, J. C. Nieto, y M. Á. M. Alfaro, «Patrones espaciales de cambio de cobertura y uso del suelo en el área cafetalera de la Sierra Norte de Puebla», p. 16.
10. S. W. Jeffrey y G. F. Humphrey, «New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton», *Biochem. Physiol. Pflanz.*, vol. 167, n.o 2, pp. 191-194, ene. 1975, doi: 10.1016/S0015-3796(17)30778-3.

Hacia la construcción de un proceso de trabajo colaborativo para la elicitación de requisitos basado en entendimiento compartido

Valentina Vergara Realpe[1][0000-0002-1302-7398], Jesús-David Perea[1][0000-0002-4140-7582], Vanessa Agredo Delgado[1,2][0000-0003-0870-6895], Pablo H. Ruiz[1,2][0000-0003-2098-2614] and César A. Collazos[2][0000-0002-7099-8131]

[1]Corporación Universitaria Comfacauca - Unicomfacauca, Popayán, Colombia
{valentinavergara, jesusperea, vagredo, pruiz}@unicomfacauca.edu.co
[2]Universidad del Cauca, Popayán, Colombia
ccollazo@unicauca.edu.co

Resumen. La elicitación de requisitos se considera como una actividad compleja y vital del proceso de desarrollo de software, en la cual se busca abstraer y comprender problemas y necesidades que se pueden resolver con la construcción de un producto software. Así como es conocida su importancia, sigue siendo una de las actividades más débiles del proceso, siendo esencialmente, una actividad humana en la cual se intercambian diferentes perspectivas, experiencias y conocimientos para recopilar y modelar lo que el software debe hacer. Además, en esta se conjugan una mezcla de estrategias, herramientas y colaboraciones entre personas, lo que hace que el entendimiento común y homogéneo de estos problemas y necesidades no se logre fácilmente. En consecuencia, este artículo presenta el avance inicial de la investigación que ha permitido articular y fundamentar una propuesta para la definición de un proceso de trabajo colaborativo, en la elicitación de requisitos basado en entendimiento compartido. Este proceso busca guiary apoyar la construcción del entendimiento compartido para que la elicitación de los requisitos sea más confiable y por ende la construcción de producto softwaretambién.

Palabras clave. Entendimiento compartido, Elicitación de requisitos, Trabajo colaborativo, Proceso.

1 Planteamiento del Problema

La Ingeniería de Software es una disciplina que abarca diversas etapas del ciclo de vida del software como son: la elicitación de requisitos, análisis, diseño, implementación, pruebas y mantenimiento [1]. Los modelos del ciclo de vida del software pretenden abarcar todo el proceso del desarrollo del software, creando para cada etapa elementos de proceso que apoyan y describen la ejecución del proyecto software [2]. A través de

los años se ha podido evidenciar que la elicitación de los requisitos es la etapa más importante en un proyecto de desarrollo de software [3]. Esta etapa se utiliza para traducir las necesidades y deseos imprecisos e incompletos de los usuarios potenciales del software en especificaciones completas, precisas y formales. En esta etapa se busca abstraer y comprender problemas y necesidades que se pueden resolver con la construcción de un producto software. Lograr una correcta elicitación de requisitos es una actividad compleja y vital, para desarrollar software eficaz y reducir los errores en su construcción [3], puesto que establece el punto de inicio para la planeación, estimación de recursos y la elaboración de cronogramas que soportaran toda la gestión del proyecto software. Es conocido que los fracasos de la gran mayoría de los productos de software se basan en una mala elicitación, gestión, análisis y especificación de los requisitos, incrementando los costos, generando re-trabajo y así, excediendo los tiempos de desarrollo [4]. Pero, así como es conocida su importancia, sigue siendo una de las etapas más débiles del proceso, se trata, esencialmente, de una actividad humana en la cual se intercambian diferentes perspectivas, experiencias y conocimientos para recopilar, especificar y modelar lo que el software debe hacer [5]. Además, en esta etapa se conjugan una mezcla de estrategias, herramientas y colaboraciones entre personas, lo que hace que el entendimiento común y homogéneo de estos problemas y necesidades no se logre fácilmente, es decir el entendimiento compartido entre los participantes de esta etapa es difícil de construir [6]. En este sentido, el entendimiento compartido se refiere al grado en que los miembros de un equipo coinciden en los pasos de un proceso de trabajo, el significado de esos pasos, el orden, la relación de las actividades y su comunicación [7]. Realizar una correcta gestión de la comunicación en el proceso de elicitación, es uno de los principales elementos para su éxito, debido a que pueden existir diversos intereses con diferentes expectativas de comprensión e interpretación de los problemas y necesidades del contexto del producto software, que pueden producir desinformación que al no ser tratadas a tiempo, puede generar lagunas evidentes en la comunicación y entendimiento entre los usuarios y el equipo de desarrollo [8]. Además, en la elicitación se debe generar un entorno abierto y de colaboración, donde todo aquel que tenga algo que decir o aportar, pueda hacerlo libremente y sin sentirse coaccionado y sobre todo que todos entiendan lo mismo [9]. La elicitación de requisitos se basa en la interacción y la cola-boración entre personas donde su base es el entendimiento mutuo, por lo tanto, al no existir, es otra de las razones de su fracaso [6].

El trabajo colaborativo requiere que los participantes que ejecutan una actividad compartan el proceso de creación de conocimiento [10], [11] lo que implica brindar espacios de interacción entre las personas para cumplir un objetivo basado en la utilización de diferentes elementos de colaboración (roles, tareas, estrategias, herramientas, métricas, etc.) que permitan incorporar las perspectivas de los participantes y de su trabajo en equipo [12], [13]. La elicitación de requisitos demanda trabajo colaborativo entre todos los participantes del proyecto, para conocer el dominio del problema, descubrir las necesidades reales y llegar a un acuerdo en cuanto a la visión y objetivos del producto a desarrollar, además de hablar el mismo lenguaje de los usuarios finales [14]. Por todo lo anterior la pregunta de investigación para esta propuesta es: *¿Cómo apoyar la construcción del entendimiento compartido*

en la elicitación de requisitos para el desarrollo de productos software? En consecuencia, este artículo presenta el avance inicial de la investigación que ha permitido articular y fundamentar una propuesta para la definición de una estrategia de trabajo colaborativo, en la elicitación de requisitos basada en la construcción de entendimiento compartido.

El resto de este documento se estructura de la siguiente manera: la sección 2 describe la justificación. La sección 3 presenta trabajos relacionados. La sección 4 define los objetivos de la propuesta. La sección 5 describe la metodología, y la sección 6 se describen las conclusiones preliminares del trabajo.

2 Justificación

El entendimiento compartido se refiere al grado en que los miembros de un equipo coinciden en los pasos de un proceso de trabajo, el significado de esos pasos, el orden, la relación de las actividades y su comunicación [7]. Por otro lado, la elicitación de requisitos demanda trabajo colaborativo [14], debido que es una actividad humana en la cual pueden participar diversas personas con pluralidad cultural, diferentes niveles de experiencia, diversidad de conocimientos, lo cual puede plantear desafíos en la comunicación y en la resolución de problemas conjunta, para lo que es necesario construirun entendimiento compartido para lograr los objetivos de la actividad [15]. Por lo tanto, en este trabajo de investigación se pretende definir una estrategia de elicitación de requisitos basada en el trabajo colaborativo y en la construcción de entendimiento compartido, que pretende brindar a la comunidad empresarial un enfoque de apoyo a la elicitación de requisitos la cual es fundamental para el éxito de los proyectos de desarrollo de software, y de esta forma brindarles elementos que puedan beneficiar la calidad de los productos software. Diferentes autores han propuesto técnicas y modelos para obtener requisitos, donde la mayoría describen estrategias en las cuales no se trabaja la construcción del entendimiento compartido [2], por lo tanto, se requiere ofrecer a la comunidad de investigación una estrategia novedosa para la elicitación de requisitos que incorpore elementos de trabajo colaborativo en combinación de estrategias de construcción de entendimiento compartido que beneficien el desarrollo de esta ardua actividad. Finalmente, a la comunidad académica, se pretende brindar un enfoque útil y adecuado al contexto de los laboratorios de ingeniería de software para que pueda serconsiderado como parte de las metodologías de desarrollo de software utilizadas en este campo.

3 Trabajos Relacionados

Entendimiento compartido en ingeniería del software

En [16] se propone una teoría en las organizaciones de software basada en entendimiento compartido, donde se resalta que la coordinación y la comunicación son elementos esenciales y problemáticos. El papel, el valor y el uso del entendimiento compartido en la ingeniería de software se investiga en [17], mostrando una recopilación deprácticas, así como una hoja de ruta para mejorar el conocimiento y la

práctica en estaárea. Por su parte en [18] se aborda la influencia de la distribución del equipo en el éxito del proyecto, utilizando la teoría de modelos mentales compartidos para la construcción y el mantenimiento del entendimiento compartido. En [19] se analiza la comunicación, mediante el entendimiento compartido, de los conceptos o relaciones subyacentes de un equipo multidisciplinar en el desarrollo una aplicación móvil. Específicamente, en la elicitación de requisitos en [20] se investiga el papel de los elementos cognitivos para mejorar la claridad de las historias de usuario, proponiendo un conjunto de elementos de escritura desde diferentes dominios con el fin de mitigar la ambigüedad y mejorar el entendimiento compartido. En [21] se hace un análisis sobre cómo la cul- tura afecta el entendimiento compartido en los ingenieros de requisitos y el equipo de desarrollo así como en la organización y el progreso del proyecto.

Trabajo colaborativo en ingeniería del software

En [22] se aplica la ingeniería de colaboración para derivar un módulo de proceso (thinkLet "MindMerger") para apoyar sistemáticamente a grupos de trabajo heterogéneos en la construcción entendimiento compartido en procesos industriales. En [23] se describe una infraestructura de soporte a la ingeniería de software colaborativa, donde se consideran roles de la evaluación heurística, análisis de registros y visualización para cuantificar los beneficios de esta infraestructura. Por su parte en [24] se propone un modelo para la enseñanza de la ingeniería de software que busca experimentar los principales problemas del trabajo colaborativo en el desarrollo de productos software. Similarmente en [25] se muestran objetivos, métodos y herramientas que se utilizan en una clase de ingeniería de software global enfatizando en el trabajo colaborativo, además, esfuerzos para automatizar datos de medición participativa y prácticas utilizadas en equipos globales de desarrollo de software. En [10] se muestra SmartWiki, una herramienta colaborativa para ayudar enfrentar los desafíos de la ingeniería de requisitosdistribuida, ayudando a mejorar la calidad de los requisitos y a perfeccionar las anotaciones para que sus especificaciones sean coherentes. Finalmente, en [26] se muestra una revisión sistemática de la literatura que hace un análisis de la ingeniería de software transcultural, identificando impactos potenciales de los factores culturales sobre los enfoques colaborativos y de comportamiento, además, identifica cómo las dimensiones culturales pueden afectar el trabajo colaborativo en la ingeniería de software.

De acuerdo con los trabajos relacionados, existen diferentes acercamientos en la literatura respecto al trabajo colaborativo e independientemente del entendimiento compartido, pero ninguno, define un proceso completo que involucre el trabajo colaborativo y la construcción del entendimiento compartido en la elicitación de requisitos desde que se hace su planeación, posteriormente su ejecución y validación, brindando el qué y cómo hacerlo, por lo tanto, esta pretende ser la mayor contribución de este trabajo de investigación.

4 Objetivos

Objetivo General
Definir un proceso[1] basado en trabajo colaborativo para la elicitación de requisitos en el desarrollo de productos software mediante la construcción de entendimiento compartido.

Objetivos Específicos
- Caracterizar e identificar de la literatura los diferentes elementos disponibles para el trabajo colaborativo en la elicitación de requisitos
- Identificar los elementos de entendimiento compartido que sean necesarios para elicitar requisitos en el desarrollo de productos software
- Construir un proceso de trabajo colaborativo para la elicitación de requisitos utilizando elementos de entendimiento compartido
- Validar la viabilidad del proceso propuesto mediante su aplicación en un contexto académico o industrial

5 Metodología

Este trabajo se llevará a cabo siguiendo la metodología de investigación-acción multi-ciclo con bifurcación que define tres ciclos: conceptual, metodológico y de evaluación [27]. Los ciclos y sus respectivas actividades se describen brevemente a continuación: *Ciclo Conceptual:* Planificación y ejecución de la identificación de características y elementos de los diferentes enfoques que servirán para la creación del proceso. Las actividades que están relacionadas con el primer y segundo objetivo son: a) Planeación del mapeo sistemático, se definen los objetivos, preguntas, criterios de inclusión y exclusión y las posibles bases de datos fuentes de información, b) Las cadenas de búsqueda son definidas y se ejecutan en las bases de datos de donde se extraerán los contenidos y datos relevantes, c) La información recolectada se analiza respecto a los diferentes elementos del trabajo colaborativo que apoyen la elicitación de requisitos, d) Identificar de acuerdo a la literatura los elementos del entendimiento compartido que puedan apoyar la elicitación de requisitos, e) Clasificar y/o definir los elementos del entendimiento compartido para ser integrados en la elicitación de requisitos.

Ciclo Metodológico: Se define la estructura base de los elementos que formarán parte del proceso propuesto. Las actividades relacionadas con el tercer objetivo son: a) Definir la estructura base de elementos que harán parte del proceso propuesto, b) Incorporar incrementalmente cada uno de los componentes que serán parte del proceso.

Ciclo de Evaluación: Se elegirá un mecanismo de evaluación mediante el cual se

[1] Proceso para este proyecto se considera como una secuencia de pasos organizados de forma lógica que se enfocan en lograr algún resultado específico.

validará el proceso propuesto. Las actividades relacionadas con el cuarto objetivo son: a) Escoger un mecanismo de evaluación con el cual se validará el proceso propuesto, b) Planear, diseñar y ejecutar el mecanismo de validación, c) Refinar el proceso teniendo en cuenta el análisis de los resultados obtenidos en su validación.

6 Conclusiones

Este artículo presenta el avance inicial de la investigación que ha permitido articular y fundamentar una propuesta para definir un proceso basado en trabajo colaborativo para la elicitación de requisitos mediante la construcción de entendimiento compartido, desde que se planea hasta que se valida, definiendo el qué y cómo hacerlo, considerando que según los trabajos relacionados, existen diferentes acercamientos en la literatura respecto al trabajo colaborativo y entendimiento compartido, de forma aislada, pero ninguno, define una estrategia que involucre ambos elementos y que brinde un paso a paso, por lo tanto, esta pretende ser la mayor contribución de este trabajo de investigación, buscando mediante la coordinación y colaboración entre personas que comparten conocimientos, técnicas y herramientas se pueda lograr entendimiento compartido en la elicitación de requisitos. Además, se identifica que la elicitación de requisitos es unaactividad compleja que requiere la participación de diversas personas con pluralidad cultural, con ideas y conocimientos diferentes, donde es necesario comprender e interpretar adecuadamente las opiniones heterogéneas que surgen en la elicitación, estableciendo un entendimiento homogéneo.

Referencias

1. L. Garcés y L. M. Egas, «Evolución de las Metodologías de desarrollo de la Ingeniería de software en el proceso la Ingeniería de Sistemas Software,» *Revista Científica y Tecnológica UPSE,* vol. 1, nº 3, 2013.
2. I. Sommerville, Ingeniería del software, Madrid: Pearson educación, 2006.
3. A. Chakraborty, M. K. Baowaly, A. Arefin y A. N. Bahar, «The Role of Requirement Engineering in Software Development Life Cycle,» *Journal of Emerging Trends in Computing and Information Sciences,* vol. 3, nº 5, pp. 723-729, 2012.
4. L. R. Wong Portillo, «Contribuciones en el proceso de elicitación de requisitos: factores, actividades y cualidades,» Universidad Nacional Mayor de San Marcos, Lima, 2019.
5. M. Arias Chaves, «La ingeniería de requerimientos y su importancia en el desarrollo de proyectos desoftware,» *InterSedes: Revista de las Sedes Regionales,* vol. 6, nº 10, pp. 1-13, 2005.
6. M. Corvera Charaf, C. Rosenkranz y R. Holten, «The emergence of shared understanding: applying functional pragmatics to study the requirements development process,» *Information Systems Journal,* vol. 23, nº 2, pp. 115-135, 2013.
7. E. A. Christiane Bittner y J. M. Leimeister, «Creating Shared Understanding in Heterogeneous Work Groups: Why It Matters and How to Achieve It,» *Journal of*

Management Information Systems, vol. 31, nº 1, pp. 111-144, 2014.
8. L. Westfall, «Las fallas en la ingeniería de requisitos,» *Ingenierías USBMed,* vol. 2, nº 2, pp. 40-47, 2011.
9. J. C. Ramírez-Leal, W. J. Giraldo-Orozco y R. Anaya-Hernández, «Una propuesta metodológica para mejorar la comunicación en ingeniería de requisitos,» *Revista EIA,* vol. 13, nº 26, pp. 121-139, 02 052016.
10. P. Dillenbourg, M. J. Baker, A. Blaye y C. O'Malley, «The evolution of research on collaborative learning,» de *Learning in humans and machine: Towards an interdisciplinary learning science*, Oxford, Elsevier, 1996, pp. 189-211.
11. J. Roschelle y S. D. Teasley, «The Construction of Shared Knowledge in Collaborative Problem Solving,» de *Computer supported collaborative learning*, Berlin, Springer, 1995, pp. 69-97.
12. O. Kozar, «Towards Better Group Work: Seeing the Difference between Cooperation and Collaboration,» de *English Teaching Forum*, Washington, ERIC, 2010, pp. 16-23.
13. J. Gómez, V. Agredo-Delgado y P. H. Ruiz, «Proposal for construction of an approach for defining feature models in Software Product Lines based on Collaborative Work,» *CEUR Workshop Proceedings ,* vol. 747, 2020.
14. A. Azadegan, K. N. Papamichail y P. Sampaio, «Applying collaborative process design to user requirements elicitation: A case study,» *Computers in Industry,* vol. 64, nº 7, pp. 798-812, 06 2013.
15. J. Langan-Fox, J. Anglim y J. R. Wilson, «Mental models, team mental models, and performance: Process, development, and future directions,» *Human Factors and Ergonomics in Manufacturing & Service Industries,* vol. 14, nº 4, pp. 331-352, 2004.
16. J. Aranda, A Theory of Shared Understanding, Toronto: University of Toronto, 2010.
17. M. Glinz y S. A. Fricker, «On shared understanding in software engineering: an essay,» *Computer Science-Research and Development ,* vol. 30, nº 3, pp. 363-376, 2014.
18. H. Markus, C. Rosenkranz y R. Holten, «The Role of Shared Understanding in Distributed Scrum Development: An Empirical Analysis,» *Engineering, Computer Science,* 2016.
19. A. Hoffmann, E. A. Christiane Bittner y J. M. Leimeister, «The emergence of mutual and shared understanding in the system development process,» de *International Working Conference on Requirements Engineering: Foundation for Software Quality*, Berlin, 2013.
20. A. R. Amna, «The Use of Context-Free Grammar and Gamification to Reduce User Story Ambiguity and Raise Ambiguity Awareness: A Proposal,» de *CEUR Workshop Proceedings*, Belgium, 2020.
21. Y. Hsieh, «Culture and Shared Understanding in Distributed Requirements Engineering,» de *IEEE International Conference on Global Software Engineering (ICGSE'06)*, 2006.
22. E. A. Christiane Bittner y J. M. Leimeister, «Creating Shared Understanding in Heterogeneous Work Groups: Why It Matters and How to Achieve It,» *Journal of Management Information Systems,* vol. 31, nº 1, pp. 111-144, 2014.
23. C. Cook y N. Churcher, «Modelling and Measuring Collaborative Software,»

University of Canterbury, New Zealand, 2004.
24. M. I. Alfonso y F. Mora, «Learning software engineering with group work,» de *Proceedings 16th Conference on Software Engineering Education and Training*, Alicante, 2003.
25. G. D. Thompson, D. Petkovic y S. Huang, «Teaching distributed collaborative development techniques in a software engineering class setting,» Integrating FOSS into the Undergraduate Computing Curriculum, San Francisco, 2009.
26. F. Fazli y E. A. Christiane Bittner, «Cultural Influences on Collaborative Work in Software Engineering Teams,» de *Proceedings of the 50th Hawaii international conference on system sciences*, 2017.
27. V. Lencinas, F. Nataloni, S. Hannover Saavedra, T. G. Villanueva, S. Lacolla, M. López y I. G. Zabczuk, «Una oportunidad para generar conocimiento desde la practica profesional de bibliotecarios y archiveros,» *Revista Prefacio ,* vol. 1, nº 1, pp. 10-23, 2017.

Factores Críticos de Éxito y Medidas de Rendimiento de Procesos (BPM) para Eficientizar la Toma de Decisiones

Ángel Fabián Castillo[1]

[1] Universidad Católica de Santiago del Estero -Departamento Académico San Salvador - Jujuy
fcastillo@ucse.edu.ar

Abstract. La característica que hace a una organización inteligente es la capacidad de poder "Medir" sus procesos. Una organización inteligente es aquella flexible, de alta conectividad interna y externa, capaz de medir rápidamente posibilidades, de adaptarse a un entorno dinámico y desarrollar iniciativas frente a los nuevos estímulos que proceden de los escenarios cambiantes de la realidad sosteniendo la direccionalidad estratégica. El análisis cualitativo de los procesos es útil para obtener información sistemática sobre el procesos a través de un abanico de principios, técnicas y herramientas para identificar y documentar problemas en un proceso desde múltiples perspectivas y analizar profundamente las causas de estos problemas. Sin embargo para proporcionar una base sólida para la toma de decisiones es necesario analizar cuantitativamente los procesos de negocio, en términos de medidas de rendimiento del proceso, como el tiempo de ciclo, los tiempos de espera, costo entre otras, haciendo foco especialmente en el análisis de flujo y simulación. Las medidas de rendimiento vinculado con los factores críticos de éxito (FCE) nos proporcionan una visión clara y sólida para la toma de decisiones conforme a las estrategias de negocios, indicando el camino hacia el que deben enfocar los esfuerzos individuales y colectivos de una organización. El objetivo de este trabajo es identificar la necesidad insoslayable de vincular la determinación de las medidas de rendimiento con los Factores Críticos de Éxito con el fin de mejorar significativamente la toma de decisiones para tal propósito se muestra el trabajo de campo en el ámbito de la gestión estatal provincia realizado en el marco de la Tesis de Maestría de Ingeniería de Software.

Palabras clave. (FCE) Factor Crítico de Éxito, (BPM) Business Process Management, Análisis cualitativo-cuantitativo.

1 Introducción

Los indicadores de negocio (medidas de rendimiento) deben ser una representación coherente de la estrategia de negocio a través de objetivos claramente vinculados entre si. Estos objetivos dan origen a los indicadores de resultado que deben corresponderse con el logro de los mencionados objetivos y metas que de ellos se deriven. Por lo tanto el logro de los objetivos y las metas está respaldada por un

conjunto de indicadores de desempeño que permiten evaluar los procesos recorridos para el logro de los mencionados objetivos y metas.
Los resultados deben traducirse, finalmente, en logros de procesos que maximizan el agregado de valor tanto del proceso como del negocio.
Se requiere entonces un conjunto de indicadores que reflejen las cosas que se necesitan "hacer bien" (Factores Críticos de Éxito) para avanzar claramente hacia el cumplimiento de los objetivos y estrategias del negocio. El progreso de las acciones que propicien el logro de los objetivos, pueden y deben ser supervisados mediante medidas de rendimiento, como así también supervisar la evolución de los Factores Críticos de Éxito. El propósito de la definición de los Factores Críticos de Éxito y de la vigilancia de su evolución, es canalizar acciones y esfuerzos y redefinir prioridades en la gestión estratégica del negocio.
Se requiere de un sistema de medición que dispare e impulse al cambio "La medición motiva determinados comportamientos, asociados tanto al logro como a la comunicación de los resultados organizacionales, de los equipos y de los individuales"[1]. De allí que es fundamental la definición de indicadores que generen los comportamientos esperados, particularmente aquellos que orienten a la gestión para adaptarse con la velocidad adecuada ante un entorno en permanente y acelerado cambio.
El objetivo de este trabajo es identificar la necesidad insoslayable de vincular la determinación de las medidas de rendimiento de procesos con los Factores Críticos de Éxito con el fin de mejorar significativamente la toma de decisiones, haciendo foco en el uso de los FCE para las medidas de rendimiento del análisis de flujo y simulación.
El presente trabajo se organiza de la siguiente manera:
La sección 2 detalla el concepto de FCE y el uso de las herramienta para su vinculación con los procesos.
La sección 3 muestra el concepto de medidas de rendimiento y sus formas.
La sección 4 describe la utilización de FCE para la determinación de medidas de rendimiento. 4.2 muestra la parte inicial del trabajo experimental donde se aplica el concepto de FCE vinculado con las medidas de rendimiento en el análisis del proceso modelo As Is.
La sección 5 se expresan las conclusiones

2 Factores Críticos de Éxito (FCE)

El concepto de Factor Crítico de Éxito fue utilizado, por primera vez, como vinculante entre el Management y la Tecnología Informática, por John Rockart (MIT), en un famoso artículo de la Harvard Business Review (década del 70): "Los altos directivos definen sus necesidades de Información"
"Se puede expresar que Factores Críticos de Éxito son aquellas "cosas" que si funcionan bien o si se verifica su ocurrencia, "todo funciona bien". Son los aspectos sobre los que fija su atención un directivo para estar relativamente seguro que, su área de gestión avanza hacia el cumplimiento de sus metas." [2].

La toma de decisiones de los niveles directivos de la organización están alineadas con los Factores Críticos de Éxito (FCE). Algunos autores interpretan a los Factores Críticos de Éxito como núcleos destacados de las metas de gestión constituyéndose en los indicadores que conforman el "tablero de control" de un área de gestión.
Los FCE de un nivel jerárquico, frecuentemente, son metas de gestión de los niveles inferiores.
Los FCE deben ser cuantificables, (Ej: "Disminuir el ausentismo docente desde el 20 % al 8 % en los próximos nueve meses" es una meta de gestión del jefe de personal y un FCE para el Ministro de Educación).

2.1 Matriz "Procesos/Factores Críticos de Éxito"

La finalidad de matriz es identificar los procesos relacionados con una mayor cantidad de Factores Críticos de Éxito. Un proceso puede estar involucrado en más de un FCE. "Se establece si existe relación entre el Factor Crítico de Éxito y un proceso determinado estableciendo una ponderación de relación: "3". Muy fuerte relación entre el proceso y el Factor Crítico de Éxito considerados:"2", relación de tipo intermedio:"1". Muy débilmente relacionados.

En la última columna se registra un "peso" ponderado de los vínculos entre el proceso considerado y los Factores Críticos de Éxito." [2]. Ver Fig. 1

FACTORES / PROCESOS	Factores Críticos de Éxito													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Diligenciar Tareas Pasivas / Livianas en el Nivel Medio y Superior.	2	2	0	0	3	0	2	3	3	0	0	0	0	0
6 Diligenciar Reconocimiento de Servicios de Docentes en el Nivel Medio y Superior.	0	0	0	0	3	0	2	3	3	0	2	0	0	0
9 Designar Docentes en Horas Cátedras en el Nivel Medio y Superior.	0	0	3	1	2	3	2	3	3	0	3	0	0	2
10 Diligenciar Cómputo de Servicios Docente en el Nivel Medio y Superior.	0	0	0	0	3	0	2	3	3	0	2	0	0	0
11 Otorgar Licencia Docente por Razones Particulares en el Nivel Medio y Superior.	0	3	1	0	3	1	2	3	3	0	2	0	0	0
12 Otorgar Licencia Docente por Cargos de Mayor Jerarquía en el Nivel Medio y Superior.	0	0	1	0	3	1	2	3	3	0	2	0	0	0
13 Diligenciar Tareas Pasivas / Livianas en el Nivel Inicial y Primario.	2	2	0	0	3	0	2	3	3	0	0	0	0	0
19 Diligenciar Reconocimiento de Servicios de Docentes en el Nivel Inicial y Primario.	0	0	0	0	3	0	2	3	3	0	2	0	0	0
22 Designar Docente Suplente y/o Interino en el Nivel Inicial y Primario.	0	0	3	1	2	3	2	3	3	0	3	0	0	2
23 Designar Docente Titular en el Nivel Inicial y Primario.	0	0	3	0	2	3	2	3	3	0	3	0	0	2
24 Diligenciar Prolongación de Jornada Habitual de Docente en el Nivel Inicial y Primario.	0	0	0	0	2	1	2	3	2	0	1	0	0	0
25 Diligenciar Cómputo de Servicios de Docente en el Nivel Inicial y Primario.	0	0	0	0	3	0	2	3	3	0	2	0	0	0
26 Otorgar Licencia de Docente en el Nivel Inicial y Primario.	0	3	1	0	3	1	2	3	3	0	2	0	0	0

Fig. 1. Ejemplifica la "Matriz FCE/Procesos" realizada en el proyecto área Ministerio de Educación

2.2 Matriz "Cantidad de FCE/Calidad de la implementación actual"

La aplicación de esta herramienta es determinar las prioridades y el grado de criticidad de los procesos de acuerdo con el estado de las medidas de rendimiento establecidas: "¿Qué Procesos considerar en primer lugar? Aquellos que tengan una

mala implementación actual y que afecten a una mayor cantidad de FCE." [2]. Ver Fig. 2

22				P22	P9
21					P23
18				P26	P11
15			P13	P1-P12	
13			P25		P6-P10
11	P24				
FCE ponderados acumulados por Proceso / Calidad de la Implementación Actual	A	B	C	D	E

Fig. 2. Ejemplifica la "Matriz FCE/Calidad de Implementación"

3 Medidas de Rendimiento

El Ingeniero de software, Tom De Marco, dice: "no se puede controlar lo que no se puede medir". Así que antes de comenzar a analizar cualquier proceso en detalle, es importante definir claramente la medida de rendimiento del proceso (también llamado métricas de rendimiento de proceso) que se usará para determinar si un proceso está en "buena forma" o en "mal estado". Se deben considerar dos aspectos en el análisis de rendimiento el aspecto cualitativo del proceso y el cuantitativo en función de los factores críticos de éxito para tener una base sólida de toma de decisiones.

La medidas de rendimiento esenciales permiten evaluar el proceso en tres dimensiones: tiempo, costo y calidad se deben identificar las medidas de rendimiento más relevantes en función de los FCE.

El análisis de flujo con la simulación permite determinar medidas de desempeño como:

- **Tiempo de Ciclo** = Tiempo de Procesamiento + Tiempo de Espera
- **Eficiencia de Tiempo de Ciclo** = Tiempo de Procesamiento / Tiempo de Ciclo
- **Costo por Instancia** = Costo de Procesamiento + Costo de Desperdicios
- **Utilización de Recursos** = Tiempo dedicado del recurso al trabajo del proceso / Tiempo disponible del recurso para el trabajo del proceso
- **Rendimiento del Proceso =** Modelo del Proceso + Desempeño de cada Tarea

"El cálculo del tiempo de ciclo, mediante el análisis de flujo, es el tiempo medio que transcurre entre la momento en que comienza el proceso y el momento en que finaliza. Por extensión, decimos que El tiempo de ciclo de una tarea es el tiempo promedio que transcurre entre el momento en que la tarea comienza y el momento en que se completa." [3]. De acuerdo al modelo del procesos el tiempo de ciclo variará según se trata de un proceso secuencial, o con Gateway XOR, Gateway AND o un ciclo de repetición

Marlon Dumas[3] Ecuaciones de análisis de flujo para el tiempo de ciclo. Ver Fig. 3

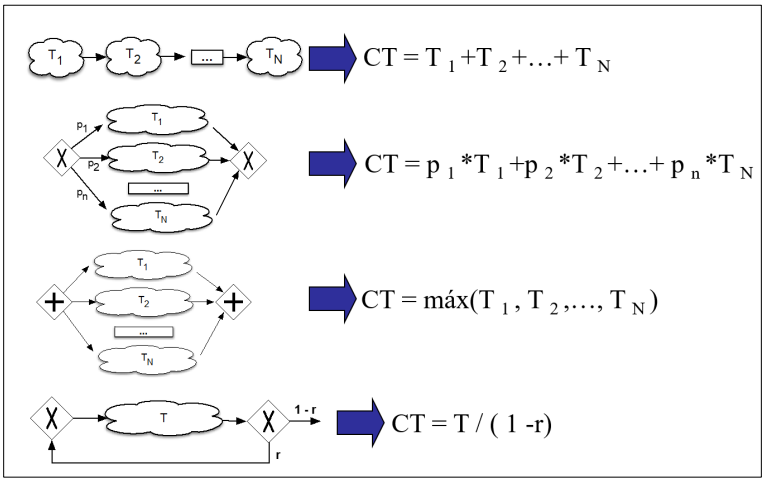


Fig. 3. Ecuaciones Tiempo de Ciclo de Marlon Dumas - "Fundamentals of Business Process Management. 2013th Edition"

Simulación modelo As Is: La simulación de procesos a través de una herramienta de simulación de procesos permite generar un gran número de instancias hipotéticas de un proceso, ejecutando estas instancias paso a paso, y registrando las métricas en cada paso en esta ejecución. La salida de un simulador incluye los registros de la simulación, así como estadísticas de tiempos de ciclo, tiempos de espera promedio y utilización media de los recursos.

Los pasos en el procesos de simulación son:

1. Modelar el Proceso As Is
2. Definir un Escenario
3. Ejecutar la Simulación
4. Analizar los Resultados Obtenidos
5. Repetir para Escenarios Alternativos

4 Utilización del concepto de Factor Crítico de Éxito para facilitar la definición de indicadores

El concepto de Factor Crítico de Éxito puede suministrar una importante ayuda para definir los indicadores y ser incluidos en el cuadro de mando integral facilitando la posibilidad de monitoreo de la evolución de los indicadores asociados a cada uno de los Factores Críticos de Éxito.

Una secuencia posible, para la mencionada definición de indicadores, es la siguiente:

a.Relevar, para cada uno de los más altos niveles de la Gestión sus los Factores Críticos de Éxito. El aporte de las máximas autoridades es esencial. “Esas cosas que deben funcionar bien” deben estar concretamente referidas a los programas y proyectos de la Gestión Estratégica

b.Determinar un conjunto de indicadores claves que permitan medir la evolución de cada Factor Crítico de Éxito a través del tiempo y establecer el procedimiento para la obtención y procesamiento de la información relacionada con cada uno de ellos.

c.Cuando algún indicador sale de un cierto rango considerado normal dispara un alerta temprana, otorgando solida información para la toma de decisiones

d.Utilizar datos cuantitativos y cualitativos considerando que un mismo Factor Crítico de Éxito puede tener varios indicadores.

“Un buen indicador debe tener las siguientes cualidades:

- Ser operacional: enfocado a la acción y con posibilidades de suministrar información útil para el control.
- Ser comparable contra un estándar deseado de logros.
- Ser confiable, es decir que pueda evaluar, con seguridad, fenómenos de tipo cualitativo y cuantitativo.
- Poder ser definido oportunamente para incidir en decisiones relacionadas con la categoría que miden.
- Poder ser medido fácilmente.
- Posibilitar el análisis en términos de la información que requiere y sus fuentes de información interna y externa.” [1]

4.1 Balance y equilibrio dinámico entre los Factores Críticos de Éxito utilizados para definir los indicadores.

Obtener un conjunto balanceado y equilibrado de Factores Críticos de Éxito es importante para asegurar la consistencia entre los indicadores, derivados de dichos Factores Críticos de Éxito y que serán incluidos en el cuadro de mando integral para la toma de decisiones.

a.Utilizar un enfoque “top-down” para identificar los Factores Críticos de Éxito genéricos y de alto nivel (aquellos aspectos que, si se verifican, inciden positivamente en el éxito de cualquier Gestión).

b.Luego debería recurrirse a un enfoque “bottom-up” concentrándose en los componentes de la demanda. Es decir se contesta la pregunta ¿Qué debería funcionar bien? (obviamente dentro del ámbito de lo factible) para satisfacer cada uno de los segmentos (componentes) de la demanda del entorno.

c.Tomando como base los resultados obtenidos mediante los enfoques “top-down” y “bottom-up”, se procede a condensar los factores hasta llegar a máximo ocho/diez Factores Críticos de Éxito de la Gestión vinculadas con las perspectivas de tiempo, costo y calidad, es vital tener una visión estratégica, y sintetizadora. Se está hablando de “signos vitales”, es decir, aquellas variables realmente importantes para la sustentabilidad y competitividad de la organización.

d.El sistema total de Factores Críticos, debe someterse a un balanceo final utilizando un enfoque de lo particular a lo general para descartar inconsistencias y hacer las reducciones finales. El criterio empleado consiste en revisar que cada factor identificado en un nivel particular contribuya decisivamente al logro de otro factor a nivel más general y que en cada nivel horizontal no se presenten redundancias.

e.Todo el sistema debe evaluarse periódicamente. La vigencia de la Estrategia de negocio, la relevancia de los Factores Críticos de Éxito, la pertinencia de los indicadores de gestión, la calidad de los procesos. Los Factores Críticos de Éxito operativos (nivel de direcciones y departamentos) cambian más rápidamente que los Factores Críticos de Éxito macro del nivel estratégicos.

4.2 Trabajo Experimental

A continuación, se describe la parte inicial del trabajo experimental del descubrimiento del Proceso Designación de médicos por concurso Ministerio de Salud donde se aplica el concepto de FCE vinculado con las medidas de rendimiento en el análisis del proceso modelo As Is.

Objetivo del Proceso: El objetivo del proceso es Diligenciar la Designación por Concurso del Personal Médico del Ministerio de Salud de la Pcia. de Jujuy

Alcance: Este proceso se inicia cuando el Hospital/Unidad solicita designar por concurso un cargo y finaliza con la firma del Decreto de Designación por parte del Sr. Gobernador.

Unidades Involucradas: En el proceso interviene las siguientes unidades:

- •Hospital / Unidad
- •Dirección de Hospitales
- •Sub Dirección de RRHH
- •Dirección Pcial de Regulación Sanitaria
- •Dirección Pcial de Personal
- •Dirección Pcial de Presupuesto
- •Dirección General de Asuntos Jurídicos
- •Fiscalía de Estado
- •Secr. General de la Gobernación
- •Despacho Ministerio de Salud

Factores Críticos Asociados al Proceso que Surgen de Matriz FCE/Procesos

a. Garantizar la asistencia médico-preventiva y la correcta accesibilidad a los servicios de salud en los distintos niveles de gestión.

b. Fortalecer la descentralización de la Atención Primaria de la Salud (APS), logrando que el 80% de la atención de salud pueda ser realizada en los CAPS y hospitales de referencias.
c. Consolidar el Servicio de Emergencias Médicas Provincial a fin de lograr la atención de las urgencias y emergencias de manera más eficiente y eficaz
d. Consolidar el desarrollo de los recursos humanos del sector salud, adaptada a las demandas del sistema y la comunidad, teniendo como eje central al ciudadano/paciente.

Análisis Cuantitativo de la Simulación dinámica del Modelo As Is
Para la simulación del proceso se utilizó la herramienta iGrafx Process 2015.

Medidas de Rendimiento vinculadas a los FCE

- Promedio de Ciclo
- Promedio de Tempo de Trabajo
- Promedio de Tiempo de Espera

Resultado General de la simulación:

- Tiempo de simulación, en meses 37,74
- Total de trámites simulados 144

	Promedio de Ciclo	Promedio de Tiempo de Trabajo	Promedio de Tiempo de Espera
144	17,87	0,26	17,62

Medidas de Rendimiento desagregadas por áreas.

- Promedio de Ciclo
- Promedio de Tempo de Trabajo
- Promedio de Tiempo de Espera

Resultados de la simulación por áreas

- Tiempo de simulación, en meses 37,74
- Total de trámites simulados 144

Estadísticas de transacción (Días)				
	Nº	Prom. Ciclo	Prom. Trabajo	Prom. Esp
Dirección Pcial de Personal	144	82,31	0,69	81,61
Dirección Pcial de Presupuesto	144	50,30	0,13	50,18
Fiscalía de Estado	144	70,94	0,13	70,81
Jefatura de Gabinete	144	11,03	0,12	10,91

Estadísticas de transacción (Días)				
	Nº	Prom. Ciclo	Prom. Trabajo	Prom. Esp
Ministerio de Hacienda - Asesoría Legal	144	53,77	0,25	53,52
Ministerio de Salud/Dirección Gral de Asuntos Jurídicos	144	35,41	0,13	35,28
Ministerio de Salud/Dirección Pcial de Regulación Sanitaria	144	66,82	2,30	64,52
Ministerio de Salud/Dirección Pcial. de Hospitales	144	20,94	0,55	20,39
Ministerio de Salud/Hospital /Unidad que Solicita	144	0,63	0,00	0,63
Ministerio de Salud/Jef. de Despacho del Ministerio de Salud	144	23,22	2,44	20,78
Ministerio de Salud/Secr. de Coordinación de Atención de la Salud	144	8,90	0,23	8,67
Ministerio de Salud/Sub Dirección de RRHH	144	38,74	0,60	38,14
Secretaría Gral. de Gobernación	144	73,16	0,13	73,04

Los resultados de la simulación antes mostrada corresponden a procesos de desarrollo "normal" sin ningún tipo de inconvenientes o impugnaciones, los cuales se ejecuta en promedio en aproximadamente 1 año y medio **(Normal),** las medidas de rendimiento nos ofrecen un panorama claro respecto del proceso en estudio y de manera detallada por área, permitiendo elaborar la Matriz FCE/Calidad de Implementación e iniciar la etapa de optimización de procesos. La relación FCE y medidas de rendimiento permiten realizar una toma de decisiones respaldada fuertemente por el análisis cuantitativo robusto.

5 Conclusiones

De lo expuesto sobre la determinación de los indicadores para procesos BPM asociados con los FCE, se destaca que los Factores Críticos de Éxito constituyen la base para determinar gran parte de las necesidades de información (Indicadores de Rendimiento) de los máximos niveles de la organización de acuerdo con sus funciones, responsabilidades, incumbencias y relación, como así también definen un

procedimiento para la formalización y regularización de la recolección, transmisión, procesamiento y presentación de información realmente crítica en el soporte a la toma de decisiones, limitando los costos asociados con la recolección de datos no significativos.
Por otro lado las medidas de rendimiento asociadas con los Factores Críticos de Éxito otorgan una base sólida para eficientizar la toma de decisiones de manera tal que tanto la definición de los FCE y las medidas de rendimiento se diseñen de manera dinámica que cambia cuando cambia la estrategia, el medio ambiente, la demanda y/o la estructura organizativa.
El objetivo es definir la información crítica necesaria, las fuentes de información y los procesos para obtenerla regularmente de manera sistemática.
Los trabajos de campo realizados en proceso de optimización de la gestión estatal de la Provincia de Jujuy, de los cual se expuso una muestra inicial, ponen en clara evidencia la necesidad de abordar los proyectos complejos con una definición clara y elocuente de los FCE para posteriormente asociar las medidas de rendimientos adecuadas, el análisis de los resultados obtenidos en las simulación (en este caso en particular para el trabajo de campo) respecto de los indicadores de rendimiento brinda a los actores decisorios sólidas bases para la toma de decisiones ya que se puede establecer la relación FCE / Medidas de Rendimiento mostrando las tareas, procesos y áreas críticas a optimizar.
Sin embargo, existe una zona gris cuando se trata en particular con gestiones estatales complejas y es la alineación de los FCE y las metas de gestión, habitualmente a cargo de los mandos intermedios de la gestión quienes son generalmente los protagonistas de la implementación de las medidas de rendimiento, si no existe un a correcta alineación se produce un desbalanceo entre FCE y medidas de rendimiento.

Referencias

1. Dr. Roberto Uzal: “Lineamientos metodológicos para la reingeniería de procesos utilizando principios y criterios de la reingeniería de sistemas de información, Anales del Congreso Argentino de Informática. Mar del Plata, 1997”.
2. Kaplan R., Norton D. "El Cuadro de Mando Integral", Edit. Gestión 2000, 1997.
3. Marlon Dumas - “Fundamentals of Business Process Management. 2013th Edition” Springer 2013 ISBN-13: 978-3642434730

Doctoral Simposium

Formal methods for knowledge extraction and reuse from heterogeneous sources for semantic interoperability of distributed architectures

Nicolás Leutwyler

Université de Lorraine, Nancy 54000, France

Abstract. The tendency in industry, manufacturing, and agriculture nowadays goes towards adopting the Industry 4.0 practices. Additionally, Internet of Things (IoT) has seen a huge increase in its usage over the last decade, and companies are eager to profit from the advantages it has offers. Between these tendencies, the usage of data as a means to increase productivity, or similarly, to minimize loss in production is found. In those lines, Formal Concept Analysis (FCA) is a clusterization method whose output is based on patterns of concepts (sets of objects and attributes). Some extensions such as Relational Concept Analysis have arisen to tackle the use case in which there are relations between seemingly different objects, which is something FCA cannot do. However, the area of automatically using the conceptual data resulted from these methods is still immature in the sense of formalization and usage. In this Ph.D., the goal is to work in expanding the boundaries of knowledge regarding the existing algorithms, mainly looking for optimizations, and extending their current capabilities.

1 Introduction

The Ph.D., focus is on the creation of mathematical models and the implementation of intelligent sensors, or cyber-physical systems (CPS), to enrich the data layer coming from the field. One of the most relevant scientific challenges is the lack of mathematical formalization of the system models and the resulting information systems, as well as the definition of the semantics of the concepts and relationships they apply, to ensure their common understanding and facilitate their interoperability by minimizing semantic losses.

The challenge of this research project is double: on the one hand, to model data from heterogeneous sources and, on the other hand, to study the problems posed by model-based engineering in cooperative systems. It is about cooperation in “systems of actors” willing to interoperate. Cooperative systems are now organized in networks, i.e., in the form of complex systems.

A known method for knowledge extraction is the Formal Concept Analysis (FCA) [7]. Given a set of objects, with certain attributes each, the method clusterize them in pairs (X, Y) where all the objects in X have all the attributes in Y , and each attribute in Y is held by all the objects in X. The pair is called a formal concept, and the

inclusion of either the first component or the second one between formal concepts forms a partially ordered set called concept lattice [2]. This is one of the most used methods for association rules mining [3]. Moreover, one of the extensions that allows to understand relations between different (heterogeneous) types of contexts is the Relational Concept Analysis, which is known to lack some capabilities such as the intuitive representation of ternary relations [12].

2 State of the art

The complex systems envisaged will be composed of CPS networks, intelligent sensors whose purpose is to retrieve data by inserting the context and thus form information networks [10, 6] which consists of relying on different types and levels of abstraction or models. Formal concept analysis (FCA) [7] is a useful and powerful method for formally describing the links between any objects (that form a context), in particular between objects that convey knowledge. This method is based on the lattice theory [2], which can be used to solve problems of interoperability assessment between information systems within companies. An extension of the FCA framework was introduced in [9] and is called Relational Concept Analysis (RCA).
RCA focuses on datasets that are compatible with Entity-Relationship Models [1] or, alternatively, Resource Description Framework (RDF) [4]. Linked open data has been recognized as a valuable source of general data mining information, and knowledge graphs are a method for formalizing this knowledge [11].
This provides a method for extracting conceptual knowledge from multirelational data. Information extraction is part of the field of study called data mining [8], information that can be related to each other can be studied through the methods of multirelational data mining (MRDM) [5] that deals with multi-contextual data. The RCA method is not limited to knowledge extraction from separate contexts, but aims to express knowledge by interoperating the semantics of different contexts, i.e., in addition to extracting knowledge from different contexts, it also extracts knowledge from a specific context, the data contained in the other contexts are used to enrich knowledge extraction.

3 Problem Statement and Contribution

The contributions of this PhD project should include: (a) a state-of-the-art research, (b) development tools, extensions and mathematical models towards the usage of knowledge extraction, (c) the inclusion of the developed tools into the ´Ecole de Ski Fran¸caise (ESF) enterprise's software.

4 Research Methodology and Approach and Evaluation Plan

Considering the already stated problem, the methodology of the Ph.D., is divided into three main parts:

(a) systematic literature review (SLR) and the understanding of the state of the art, (b) using the outcome of the SLR to pursue a research goal and publish the advances, (c) writing the final report including a compendium of the previous works in the lines of the project, a development of each of the contributions, and a guideline to continue adding value to the field in the same topics.

Since the project involves both mathematical and software contributions, the evaluation plan has to consider them differently. On the one hand, the mathematical contributions would be evaluated by their proofs and the level of maturity they reach. On the other hand, the software would be evaluated by the users, whom can later provide insight on how useful or complete the software is to their expectations.

5 Preliminary results, Conclusions and Learned Lessons

So far, the project contains a small contribution, which is an obvious extension to the RCA algorithm to allow mining data represented with ternary relations. This is something that was already done in different ways, the novelty of the contribution would be that the way in it is implemented is more intuitive and natural than the other ones.

The generalization of the method (from ternary to n-ary) is still a work in progress, and has already been rejected in a conference.

In terms of learned lessons, I could say that, firstly, I learned the importance of having a plan when reading scientific papers. Secondly, to be clear about each contribution is way more important when writing.

Lastly, that it is also important to have a plan for publications.

References

1. Peter Pin-Shan Chen. “The entity-relationship model—toward a unified view of data”. In: ACM Transactions on Database Systems 1.1 (Mar. 1976), pp. 9–36. issn: 0362-5915. doi: 10.1145/320434.320440.
2. Rudolf Wille. “Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts”. In: Ordered Sets. Ed. by Ivan Rival. Dordrecht: Springer Netherlands, 1982, pp. 445–470. isbn: 978-94-009-7798-3. doi: 10.1007/978-94-009-7798-3_15.
3. Rakesh Agrawal, Tomasz Imieli´nski, and Arun Swami. “Mining association rules between sets of items in large databases”. In: Proceedings of the 1993 ACM SIGMOD international conference on Management of data. SIGMOD ’93. New York, NY, USA:

Association for Computing Machinery, June 1993, pp. 207–216. isbn: 978-0-89791-592-2. doi: 10.1145/170035.170072.
4. Eric Miller. “An Introduction to the Resource Description Framework”. In: Bulletin of the American Society for Information Science and Technology 25.1 (1998). eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bult.105, pp. 15–19. issn: 1550-8366. doi: 10.1002/bult.105.
5. Saso Dzeroski. “Multi-relational data mining: an introduction”. In: ACM SIGKDD Explorations Newsletter 5.1 (July 2003), pp. 1–16. issn: 1931-0145. doi: 10.1145/959242.959245.
6. Gérard Morel et al. “Manufacturing Enterprise Control and Management System Engineering: paradigms and open issues”. en. In: Annual Reviews in Control 27.2 (Jan. 2003), pp. 199–209. issn: 13675788. doi: 10.1016/j.arcontrol.2003.09.003.
7. Uta Priss. “Formal concept analysis in information science”. In: Annual Review of Information Science and Technology 40.1 (Sept. 2007), pp. 521–543. issn: 00664200. doi: 10.1002/aris.1440400120.
8. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. Introduction to Information Retrieval. ISBN: 9780511809071 Publisher: Cambridge University Press. July 2008. doi: 10.1017/CBO9780511809071.
9. Mohamed Rouane-Hacene et al. “Relational Concept Analysis: Mining Concept Lattices From Multi-Relational Data”. In: Annals of Mathematics and Artificial Intelligence 67 (Jan. 2013). doi: 10.1007/s10472-012-9329-3.
10. Olivier Cardin. “Contribution a la conception, l’evaluation et l’implémentation de systémes de production cyber-physiques”. Habilitation a diriger des recherches. Université de Nantes, Dec. 2016. https://tel.archives-ouvertes.fr/tel-01443318.
11. Petar Ristoski and Heiko Paulheim. “RDF2Vec: RDF Graph Embeddings for Data Mining”. In: The Semantic Web – ISWC 2016. Ed. by Paul Groth et al. Cham: Springer International Publishing, 2016, pp. 498–514. isbn: 978-3-319-46523-4. doi: 10.1007/978-3-319-46523-4_30.
12. Priscilla Keip et al. “Effects of Input Data Formalisation in Relational Concept Analysis for a Data Model with a Ternary Relation”. In: Formal Concept Analysis. Ed. by Diana Cristea, Florence Le Ber, and Baris Sertkaya. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 191–207. isbn: 978-3-030-21462-3. doi: 10.1007/978-3-030-21462-3_13.

Arquitecturas seguras a partir de requerimientos patrones de seguridad y vulnerabilidades

Miguel Solinas[1] and Leandro Antonelli[2]

[1] Laryc - FCEFyN, Universidad Nacional de Córdoba, Córdoba, AR
miguel.solinas@unc.edu.ar
[2] Lifia - Fac. de Informática, Universidad Nacional de La Plata, La Plata, AR
lanto@lifia.info.unlp.edu.ar

Abstract. La construcción de software seguro continúa siendo un desafío que trata los requerimientos de seguridad como no funcionales y los resuelve en etapas tardías del proceso de desarrollo de software. En un contexto de Continuous Software Engineering (CSE), considerada como la práctica corriente para la construcción de software, raramente existe una etapa de análisis de una arquitectura que garantice minimizar riesgos de vulnerabilidades conocidas. Esto conduce a la construcción de software inseguro que demanda acciones de detección posterior de vulnerabilidades. En el mejor de los casos serán eliminadas con un gran costo de recursos. O directamente se libera software con vulnerabilidades conocidas que tendrán que ser eliminadas en una etapa de actualización en versiones posteriores. Este trabajo describe las ideas seminales para enfrentar el desafío de construir un método y un sistema de recomendación de arquitecturas seguras de software.

Keywords. Software Engineering, Vulnerabilities, Security Patterns.

1 Introducción

Bajo las prácticas del Continuous Software Engineering (CSE)[1], las principales actividades orientadas a mejorar la seguridad se enfocan en revisiones de código y el uso de herramientas automatizadas de análisis, estático o dinámico, de vulnerabilidades sobre el artefacto ya construido. No es habitual utilizar una visión holística de la seguridad, aplicando por ejemplo, acciones que contemplen la especificación de requerimientos funcionales. Aún reconociendo que los costos de resolver problemas de seguridad en etapas tardías del proceso de desarrollo de software es varios órdenes de magnitud más costoso que solucionar los problemas de seguridad en etapas tempranas del mismo proceso.

2 Estado del arte

Hace al menos 10 años que la problemática de construcción de software seguro está vigente [2]. Lo que no se tuvo en cuenta en aquellos trabajos, fue el registro de vulnerabilidades. En ese momento sumaban 53.263 entradas y en mayo de este año han superado las 176.000. En todo este tiempo se han hecho esfuerzos para tomar fallas observadas en el mundo real que existen en el código, abstraerlas y agruparlas en clases comunes que representan vulnerabilidades potenciales más generales. Esto redujo ese número a una estructura relativamente adecuada para ser accesible y útil a un conjunto diverso de audiencias y propósitos.

El concepto original de lo que luego se convertiría en la lista de Common Vulnerabilities and Exposures (CVE) [4] fue presentado en enero de 1999 [3]. Han pasado más de 20 años de registro y existe un “conocimiento” contenido en la terna CVE, Common Weakness Enumeration (CWE) [5] y Common Attack Pattern Enumerations and Classifications (CAPEC) [6] que aun no se ha consolidado como una "tabla periódica" de exposiciones pero esos registros empiezan a estudiarse para mejorar de la seguridad del software.

El problema que se presenta en los escenarios de desarrollo propuestos por las metologías ágiles y las CSE es la falta de conocimiento, herramientas e interés por incorporar la seguridad como un derivado de la experiencia de vulnerabilidades conocidas [1].

Por otro lado, en el área de conocimiento de la seguridad del software los patrones de seguridad son una ayuda para hacer cumplir principios de seguridad en el diseño ya que encapsulan conocimiento experto en una formato reutilizable que es la base para la construcción de arquitecturas mas robustas, seguras y validadas. En su documentación, pueden hasta incluir código a fin de ayudar a automatizar etapas de implementación. Los patrones de seguridad siguen siendo un condimento para proporcionar seguridad a la construcción de software.

3 Planteo del problema y contribuciones

Las preguntas que surgen es ¿por qué no se generaliza la utilización de aquellos patrones? ¿dan cobertura a todos los problemas de seguridad? ¿contemplan alguna relación con las exposiciones que generan las vulnerabilidades conocidas? La respuesta a esta última pregunta conduce a una solución eficáz y eficiente para mejorar la seguridad del software.

4 Metodología y enfoque de investigación

Se utilizará una metodología analítica experimental con un enfoque cuantitativo a fin

de generalizar resultados a partir de un caso de estudio básico y su evolución hacia escenarios mas complejos. La novedad en el abordaje del problema de construcción de software seguro, está en incorporar los registros de vulnerabilidades, desde sus clases, conjuntamente con patrones propios del dominio para mejorar la seguridad de una arquitectura en una etapa temprana del proceso de desarrollo.

5 Plan de evaluación

Se hará un caso de estudio para evaluar la aplicabilidad de la propuesta y cuestionarios a los involucrados para evaluar la usabilidad.

6 Resultados preliminares o intermedios

No hay resultados aun por ser una etapa muy temprana.

7 Conclusiones y lecciones aprendidas

Si los registros de vulnerabilidades y las clases comunes que representan están evolucionando en la dirección de una "tabla periódica" de exposiciones, será posible descubrir un método y construir un sistema para que esas clases puedan ser vinculadas con requerimientos y patrones de seguridad a través de una ontología. De no ser posible su mapeo directo hoy, dado que el número de patrones de seguridad es reducido, al menos será posible generar recomendaciones esenciales para el diseño de arquitecturas mas seguras, desde etapas tempranas del proceso de desarrollo de software en un contexto de CSE.

8 Ph.D. Stage

Early Stage. El resumen que aquí se presenta se está completando para presentarse como propuesta en el Doctorado en Informática en la UNLP.

Agredecimientos a mi director de Maestría, el Dr. Leandro Antonelli quien después de diez años tuvo la generosidad de escuchar mis ideas, aconsejarme y alentarme para enfrentar el desafío de un Doctorado.

References

1. Roshan N. Rajapakse, Mansooreh Zahedi, M. Ali Babar, Haifeng Shen,

Challenges and solutions when adopting DevSecOps: A system- atic review, Information and Software Technology, Volume 141, 2022, https://doi.org/10.1016/j.infsof.2021.106700.
2. Solinas, M., Elicitación y trazabilidad de requerimientos utilizando patrones de se- guridad, Tesis Maestría, 2012, https://doi.org/10.35537/10915/4211.
3. Mann, David E. and Christey Steven M., Towards a Common Enumeration of Vul- nerabilities, 2nd Workshop on Research with Security Vulnerability Databases on January 21-22, 1999 at Purdue University in West Lafayette, Indiana, USA
4. CWE https://cwe.mitre.org/index.html. Last accessed 13 May 2022.
5. CVE https://www.cve.org/. Last accessed 13 May 2022.
6. CAPEC https://capec.mitre.org/. Last accessed 13 May 2022.

Alineación de glosarios del dominio

Domain Glossary Alignment

Paola Grijalva Arriaga[1,2][0000-0003-2616-417X], Leandro Antonelli [2][0000-0003-1388-0337] and Pablo Thomas[3][0000-0001-9861-987X]

[1] Escuela de Ingeniería en Computación e Informática, Facultad de Ciencias Agrarias, Universidad Agraria del Ecuador, Av. 25 de Julio y Pio Jaramillo, Guayaquil, Ecuador
[2] LIFIA - Facultad de Informática, Universidad Nacional de La Plata, 50 esquina 120, Buenos Aires, Argentina
[3] LIDI - Facultad de Informática, Universidad Nacional de La Plata, 50 esquina 120, Buenos Aires, Argentina
pgrijalva@uagraria.edu.ec
leandro.antonelli@lifia.info.unlp.edu.ar
pthomas@lidi.info.unlp.edu.ar

Resumen. Durante la etapa de especificación de requerimientos se hace uso de descripciones conceptuales del Universo del Dominio (UdD) a través de varias representaciones como los Glosarios. Las organizaciones tienen documentos que permiten a los ingenieros de requerimientos obtener una descripción de procesos y procedimientos que aportan al conocimiento del dominio. Por otro lado, organizaciones externas poseen también documentación que aporta al mismo dominio, pero no necesariamente utilizando los mismos términos o vocabulario. Es por ello, que es necesario una alineación de los vocabularios para poder compararlos y obtener una descripción más completa y consistente. Los glosarios son herramientas que se utilizan durante esta etapa y permiten la descripción del dominio. El objetivo de este trabajo doctoral es definir un proceso de alineación de glosarios del dominio representados a partir del Léxico Extendido del Lenguaje (LEL), integrando heurísticas y métodos semánticos, léxicos, con el fin de hallar similitudes, diferencias u omisiones. Se realiza una investigación documental de la literatura con respecto a las técnicas utilizadas para alinear dominios específicos y los utilizados para alcanzar la completitud en glosarios. Luego, se realizará la creación del método que contemple heurísticas e integre técnicas de similitud semánticas y léxicas. La evaluación del proceso se realizará mediante un experimento utilizando un caso de estudio. Y a través de un caso de estudio se validará el resultado del proceso con la alineación de LELs de dominios relacionados.

Abstract. During the requirements specification stage, conceptual descriptions of the Domain Universe (UdD) are used through various representations such as Glossaries. Organizations have documents that allow requirements engineers to obtain a description of processes and procedures that contribute to domain knowledge. On the other hand, external organizations also have documentation that contributes to the same domain, but not necessarily using the same terms or vocabulary. For this reason, it is necessary to align the vocabularies in order to compare them and obtain a more complete and consistent description. Glossaries are tools used during this stage and allow the description of the domain. The objective of this doctoral work is to define a process of aligning domain glossaries represented from the Extended Language Lexicon (LEL), integrating heuristics and semantic, lexical methods, in order to find similarities, differences or omissions. A documentary research of the literature is carried out regarding the techniques used to align specific domains and those used to achieve completeness in glossaries. Then, the creation of the method that contemplates heuristics and integrates semantic and lexical similarity techniques will be carried out. The evaluation of the process will be carried out through an experiment using a case study. And through a case study, the result of the process will be validated with the alignment of LELs of related domains.

1. Introducción/Motivación

La definición de requerimientos es una construcción gradual desde el estudio del dominio del problema hasta la captación de los diferentes requerimientos de las partes interesadas. La completitud es una característica deseada, sin embargo, es muy difícil lograr. Ridao et al. [1] indican que obtener un modelo de requerimientos completo es una meta inalcanzable, y la sola estimación del grado de completitud alcanzado es muy difícil. Una de las herramientas utilizadas y muy necesaria durante la fase de elicitación de requerimientos, son los glosarios. Martin Glinz [2] define un Glosario como una colección de definiciones de términos que son relevantes en algún dominio, que contiene referencias cruzadas, sinónimos, homónimos, siglas y abreviaturas. Arora et al.[3] indican que un glosario es una parte importante de cualquier documento de requerimientos de software. Hace explícitos los términos técnicos en un dominio y proporciona definiciones para ellos, ayudando a mitigar la imprecisión y la ambigüedad. Actualmente, es muy común la interacción de los sistemas informáticos, en donde cierto sistema brinda servicios para que consuman otros sistemas. En este marco, contar con dos LEL, en donde cada uno describa un dominio diferente, pero con perspectiva de que haya un borde de conexión, permite identificar fronteras entre los dos dominios de forma tal de que permita la interoperabilidad. En este caso, es crítico identificar la intersección, superposición y solapamiento de los glosarios del dominio, para que los sistemas se comuniquen y puedan intercambiar información.

El objetivo de este trabajo doctoral es definir un proceso de alineación de glosarios del dominio representados a partir del Léxico Extendido del Lenguaje (LEL), con el fin de hallar similitudes, diferencias u omisiones. Este proceso, brindaría solución a

tres situaciones. Por un lado, permite mejorar la calidad de los LELs, ya que permite reutilizar un LEL ya construido para enriquecer uno nuevo. Por otro lado, este proceso, permitiría analizar descripciones de dominios diferentes, buscando un borde de conexión, para identificar fronteras entre los dos dominios que permita la interoperabilidad. Finalmente, la alineación de dos descripciones del mismo dominio representados en diferentes glosarios permite unificar el lenguaje como ocurre por ejemplo cuando se necesita en iniciativas de certificación de calidad donde la organización debe alinear la descripción de sus prácticas a las descripciones de los estándares. La Figura 1 describe la arquitectura de la iniciativa propuesta.

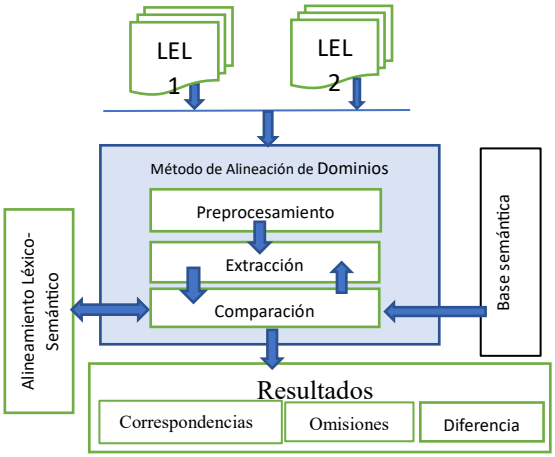


Fig. 1: Arquitectura propuesta

En la figura 1 se puede observar cómo sería el funcionamiento del método de alineación. Internamente estaría compuesto de un preprocesamiento que permite preparar el documento para que pueda ser analizado por las medidas de similitud léxica-semántica, la extracción que contendrá heurísticas y la comparación que estará relacionado con las medidas léxicos- semánticas y una base semántica con términos del dominio. Como resultado del proceso, se espera las correspondencias, omisiones y diferencias entre los LELs.

2. Estado del Arte

Se revisa información relacionada con proyectos de alineación de dominios, utilizando glosarios de dominio LEL, así como también trabajos referentes a la alineación utilizando ontologías basados en el uso de la semántica, léxica y estructural. En el caso de los Glosarios LEL, se ha encontrado documentación donde la alineación consiste en el desarrollo de heurísticas, en técnicas de inspección que permiten mejorar la completitud y disminuyen las omisiones como se puede observar

en [3],[4],[5],[6],[7]. Por otro lado, utilizando alineación de ontologías se encuentra trabajos que permiten alinear dominios utilizando técnicas de similitud semántica, estructural y léxica, como se puede observar en [8],[9],[10],[11],[12],[13].

3. Metodología y enfoque de investigación

La metodología para el desarrollo del proyecto doctoral se puede observar en la Figura 2.

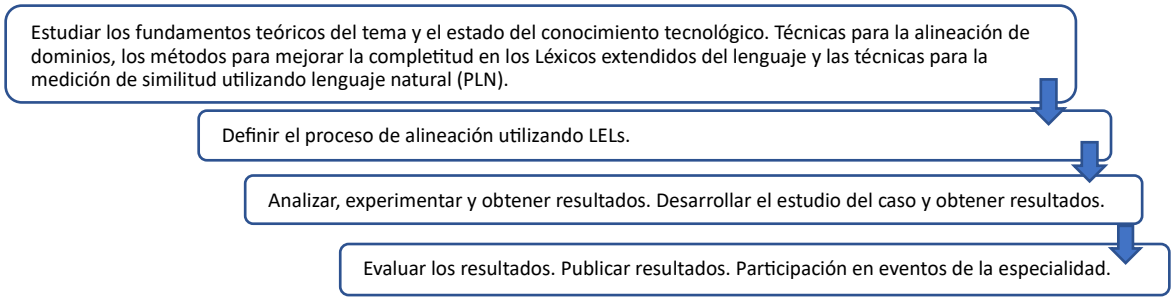


Fig. 2. SEQ Figura * ARABIC 2 Metodología de investigación y desarrollo

Primero, se realizará una investigación documental de la literatura con respecto a la temática, caracterizando el estado actual de la ciencia con respecto a los métodos utilizados para la alineación de dominios y los utilizados para alcanzar la completitud en los léxicos extendidos del lenguaje. Luego, se desarrollará el proceso de alineación integrando las técnicas que se adecuen a las naturales del LEL. La evaluación del proceso de alineación de LELs a través de la experimentación, obteniendo como resultados las heurísticas y técnicas de alineación que brinden mejores resultados, con la ejecución del proceso. Luego, se aplicará el modelo a un caso de estudio, utilizando LELs relacionados a un mismo dominio para identificar sus similitudes, diferencias u omisiones. Finalmente, con la colaboración de un ingeniero de requerimientos y usuarios expertos en el área se validará los resultados obtenidos del modelo.

4. Plan de Evaluación

Mediante un experimento se evaluará las técnicas de alineación de similitud semánticas-léxicas aplicadas en glosarios LEL, seleccionando en cada ámbito las técnicas con mayor precisión en sus resultados. Para su ejecución, se formarán conjunto de símbolos de cada LEL y en cada uno se definirán subconjuntos con la descripción correspondiente a cada símbolo. Para cada ámbito léxico o semántico se aplicará a los conjuntos y subconjuntos las medidas de distancia y similitud utilizadas en este contexto. Para la ejecución del experimento, se utilizará librerías estadísticas y de aprendizaje automático incluidas en Python.

A través de un caso de estudio se validará el proceso de alineación en base a los resultados obtenidos de la ejecución. Esta validación será realizada por un ingeniero de requerimientos y expertos en el área, a través de una IDE que facilitará visualización de resultados. El caso de estudio se enmarcá en el contexto de acreditación de las universidades y escuelas politécnicas del Ecuador (IES), que periódicamente son evaluadas para obtener la certificación que acredite la calidad mínima requerida para su funcionamiento. Para cada proceso de acreditación el Consejo de aseguramiento de la calidad (CACES) proporciona a las IES el modelo de evaluación, con el cuál serán evaluadas. Se considerará por un lado el LEL desarrollado del manual de políticas y procesos de un área sujeta a evaluación, como es el caso de investigación de la IES. Y el otro LEL estará relacionado con la función sustantiva de investigación del Modelo de evaluación para el aseguramiento de la calidad.

5. Resultados preliminares

Se realiza una investigación documental de la literatura con respecto a la temática, caracterizando el estado actual de la ciencia con respecto a las técnicas utilizadas para la alineación de dominios y los utilizados para alcanzar la completitud en los léxicos extendidos del lenguaje. La tabla 1 se observa la distribución de la clasificación documental seleccionada para análisis, las medidas más utilizadas en las diferentes técnicas y cómo han sido utilizadas estas técnicas.

Tabla 1: Distribución de la Clasificación

Herramienta	Técnica de alineación	Medidas utilizadas	Métodos
Ontología	Similitud léxica [11] Similitud Lingüística [7] Similitud semántica [21] Similitud estructural [9]	**Léxico:** Jaro Winkler, Levenshtein Distance, Jaro Distance, Hamming Distance. Medidas Rougel-L; **Semántico**: Euclidean, Coeficiente de Dice, Coeficiente de Jaccard, Resnik (res), Lin (lin) Jiang y Conrath, Leacock y Chodorow (lch), Wu y Palmer (wup), Path length (path) **Estructural:** similitud Herencia (MSI) Hermanos (MSS)	Experimentos y casos de estudios.**Semántico y léxico:** mapeos de referencia, ampliación de herramientas o prototipos, indexación estructural, emparejamiento **Estructural:** Biblioteca de patrones
Glosarios	Semánticos [5] Estructurales [13] Sintácticos [9]		Experimentos y Casos de Estudios con aplicación de Heurísticas, Variantes de Inspección y Mapas conceptuales

Los resultados de la revisión bibliográfica sobre las técnicas para la alineación de dominios utilizando glosarios y ontologías, muestran que entre las más utilizadas están a nivel semántico, léxico y estructural. Los estudios han concentrado mayor esfuerzo en el uso de ontologías para encontrar la similitud entre diferentes dominios,

utilizando varias técnicas como Euclidean, Manhathan, Coeficiente Jaccad con expansión en sinónimos, similitud coseno, empleándolas en proyectos relacionados de alineación utilizando máquinas de traducción, construcción automática de resúmenes, atribución de autoría, pruebas de lectura comprensivas, recuperación de información, que necesitan medir el grado de similitud entre dos textos dados. Por otro lado, con los glosarios de dominio (LEL), los trabajos se han concentrado en mejorar la completitud y la calidad de estos, mejorando las heurísticas y técnicas estructurales sobre el mismo glosario en análisis. Los trabajos de la revisión bibliográfica realizada se enfocan a encontrar similitudes, omisiones y errores dentro de un solo contexto.

6. Etapa doctoral

Temprana

7. Conclusiones

Los resultados de la revisión sobre las técnicas para la alineación de dominios utilizando glosarios y ontologías, muestran que entre las más utilizadas están a nivel semántico, léxico y estructural. Los estudios han concentrado mayor esfuerzo en el uso de ontologías para encontrar la similitud entre diferentes dominios, utilizando varios métodos como Euclidean, Manhathan, Coeficiente Jaccad con expansión en sinónimos, similitud coseno, en proyectos relacionados a máquinas de traducción, construcción automática de resúmenes, atribución de autoría, pruebas de lectura comprensivas, recuperación de información, que necesitan medir el grado de similitud entre dos textos dados. Por otro lado, utilizando LELs, se busca mejorar la completitud y la calidad de estos, mejorando las heurísticas y técnicas estructurales, creando mapas conceptuales con la información concerniente del dominio enfocados a encontrar similitudes, omisiones, errores dentro de un solo LEL. Cómo trabajo futuro, se creará un método de alineación de dominios, que recepte los LELs a ser alineados. Creando heurísticas e integrando medidas de similitud léxicas y semánticas y conexión a base semántica, los cuales serán probados mediante un prototipo para estos fines.

Referencias

1. M. Ridao y J. H. Doorn, «Estimación de completitud en modelos de requisitos basados en lenguaje natural», *WER 2006 - 9th Work. Requir. Eng.*, pp. 146-152, 2006, [En línea]. Disponible en: http://www.inf.puc-rio.br/wer/WERpapers/artigos/artigos_WER06/ridao.pdf.
2. M. Glinz, «A glossary of requirements engineering terminology», *Stand. Gloss. Certif. Prof. Requir. Eng. Stud. Exam, Version*, vol. 1, n.º May, p. 56, 2011.
3. C. Arora, M. Sabetzadeh, L. Briand, y F. Zimmer, «Automated Extraction and Clustering of Requirements Glossary Terms», *IEEE Trans. Softw. Eng.*, vol. 43, n.º 10, pp. 918-945, oct. 2017, doi: 10.1109/TSE.2016.2635134.

4. C. Litvak, G. Rossi, y L. Antonelli, «Conflict Management in the Collaborative Description of a Domain Language (S)», jul. 2018, n.º July, pp. 524-577, doi: 10.18293/SEKE2018-106.
5. A. Sebastián, G. D. S. Hadad, y E. Robledo, «Inspección centrada en Omisiones y Ambigüedades de un Modelo Léxico», *CIbSE 2017 - XX Ibero-American Conf. Softw. Eng.*, pp. 71-84, 2017.
6. A. Sebastián, G. D. S. Hadad, y D. Raffo, «Evaluación de Variantes de Inspección en un Modelo Léxico», *An. do WER 2019 - Work. em Eng. Requisitos*, pp. 1-6, 2019.
7. G. D. S. Hadad, C. S. Litvak, y J. H. Doorn, «Heurísticas para el modelado de requisitos escritos en lenguaje natural», *CACIC 2014 - XX Congr. Argentino Ciencias la Comput.*, 2014, [En línea]. Disponible en: http://sedici.unlp.edu.ar/handle/10915/42337.
8. D. R. Chandranegara y R. Sarno, «Ontology Alignment using combined similarity method and matching method», en *2016 International Conference on Informatics and Computing (ICIC)*, 2016, n.º Icic, pp. 239-244, doi: 10.1109/IAC.2016.7905722.
9. G. P. Kuntarto, Y. Alrin, y I. P. Gunawan, «The Key Role of Ontology Alignment and Enrichment Methodologies for Aligning and Enriching Dwipa Ontology with the Weather Concept on the Tourism Domain», en *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, oct. 2019, pp. 1-6, doi: 10.1109/ICICoS48119.2019.8982437.
10. N. Karam, A. Khiat, A. Algergawy, M. Sattler, C. Weiland, y M. Schmidt, «Matching biodiversity and ecology ontologies: challenges and evaluation results», *Knowl. Eng. Rev.*, vol. 35, p. e9, mar. 2020, doi: 10.1017/S0269888920000132.
11. B. B. Alhassan, «Extending an Ontology Alignment System with a Lexical Database», *Sci. Res. J.*, vol. III, n.º I, pp. 12-17, 2015.
12. I. Ouali, F. Ghozzi, R. Taktak, y M. S. Hadj Sassi, «Ontology Alignment using Stable Matching», *Procedia Comput. Sci.*, vol. 159, pp. 746-755, 2019, doi: 10.1016/j.procs.2019.09.230.
13. H. Karimi y A. Kamandi, «A learning-based ontology alignment approach using inductive logic programming», *Expert Syst. Appl.*, vol. 125, pp. 412-424, jul. 2019, doi: 10.1016/j.eswa.2019.02.014.

Intelligent formal analysis of heterogeneous data for semantic web

Mariano Julián Ferreirone[1,2,3], Mario Lezoche[1,2], Diego Torres[3,4], and Hervé Panetto[1,2]

[1] Univesité de Lorraine, Nancy, France

[2] CRAN, Nancy, France

[3] LIFIA, CICPBA-Facultad de Informática, UNLP

[4] Depto. CyT, UNQ, Bernal, Argentina

Abstract. The semantic web is having a main role in the development of many sectors of the industry, and is becoming necessary for being competitive in the market. The tourism in France is not an exception, and as part of their reorganization, the knowledge formalization modelling is required. This work is expected to improve the interoperability between the information represented on different knowledge formalization schema levels.

Keywords. Knowledge formalization, Semantic Interoperability, etourism.

1 Introduction

In the ”Grand Est” region, a new regional tourism policy has emerged with the elaboration of a Regional Tourism Development Plan (RTDP). The project aims to develop a digital tool to pave the way for a global, intelligent and connected system allowing the collection and processing of data in real time, in order to extract knowledge for tourism purposes. This data and information will then be aggregated and transformed by automatic inference systems to formalize the existing implicit knowledge. This formalization work will enable the use of data and information to provide more accurate and real-time services to all socio-economic actors. For the organization, a systems engineering approach [7] which consists in relying on different types and levels of abstraction or models will be used.

These models should express and formalize not only the ”structural” aspect of the system components, but also their behaviour [5], which may be limited by the specific requirements of the system domain (business rules) and the interoperability protocol(s), which may impose strict rules to provide inter-operable systems with properties such as autonomy, confidentiality and transparency[6].

The objective of this project is to model data from heterogeneous sources and to study the problems posed by model-driven engineering cooperative systems.

More specifically, the effort is going to be focused on the lack of formalization in the inter-operation of systems by minimizing semantic losses. Two popular formalization methods are going to be studied from the point of view of their interoperability: ontology and knowledge graphs. Since they can be found with different levels of schema, for example RDF (Resource Description Framework) or OWL (Web Ontology Language), it's interesting for the community to develop a way of connecting them minimizing the loss of semantics.

2 State of the art

Some interesting works related to combining knowledge formalizations with different schema levels or to increase the schema level on existent representations are as follows. In the tool named CoGui [3], the transformation is done from RDF to conceptual graphs, and the resulting OWL file is exported to different languages. When referring to RDF, currently, OWL rules, constraints and type disjunctions are ignored. Another interesting concept is the ontology alignment or ontology matching, which is based on generating a set of correspondences between concepts, properties or instances of different structured KGs, with the objective of unifying them into a new one [2]. Furthermore, Shapes Constraint Language [1] is a language for validating RDF graphs against a set of conditions, and adding formalization to RDF knowledge graphs. The characteristics of the property graphs [8] represent also an interesting model to take into account since they add some descriptive properties to the graph.

3 Problem Statement and Contributions

The contribution of this PhD project should include: a state of art research based on a systematic literature review methodology, the development of the semantic interoperability method, between heterogeneous data sources, which could be represented as a program or a set of rules that will be able to combine existent knowledge formalization models of different schema level.

4 Research Methodology and Approach

The research methodology is a process composed of the following steps: 1) Survey based on the state of art related to the combination between Knowledge formalization models of different schema level, guided by a defined literature review methodology named Scoping Literature Review (SLR) [4]. 2) Development of an approach to combine existent knowledge formalization models of different schema level. 3) Testing of the new approach using real data from an example of the corresponding knowledge representation. 4) Evaluation of the results, analyzing the equivalence of the semantics. 5) The conclusion of the advancements done in the last iteration, look for opportunities to define following challenge, write a scientific article

if corresponds. 6) Generalization of the created novel approach. 7) Develop prototype of the program. The process will be iterative from the step 2 to the step 5.

5 Evaluation Plan

The evaluation plan consists in doing several tests with different knowledge graphs (or ontologies) as inputs, measuring the semantic similarity between the input and output knowledge formalization model examples that are going to be part of the process. This will be done in order to be aware if there's a loss of semantics in the process. Furthermore, the time of execution is going to be noted and analyzed as well.

6 Conclusions and Learned Lessons

For the moment, the learned lessons are based on understanding how to find and read scientific articles, how to structure a review with a systematic methodology and concepts related to the semantic web and knowledge representation.

7 Ph.D. Stage

Early.

References

1. Shapes constraint language (SHACL). Tech. rep., W3C (Jul 2017), https://www.w3.org/TR/shacl/
2. Ardjani, F., Bouchiha, D., Malki, M.: Ontology-alignment techniques: Survey and analysis. International Journal of Modern Education and Computer Science 7, 67–78 (11 2015). https://doi.org/10.5815/ijmecs.2015.11.08
3. Baget, J.F., Chein, M., Croitoru, M., Fortin, J., Genest, D., Gutierrez, A., Leclére, M., Mugnier, M.L., Salvat, E.: RDF to Conceptual Graphs Translations. In: CS-TIW: Conceptual Structures Tool Interoperability Workshop. vol. LNAI, p. 17. Springer, Moscow, Russia (Jul 2009), https://hal-lirmm.ccsd.cnrs.fr/lirmm-00410621
4. Franciosi, C., Voisin, A., Miranda, S., Riemma, S., Iung, B.: Measuring maintenance impacts on sustainability of manufacturing industries: from a systematic literature review to a framework proposal. Journal of Cleaner Production 260, 121065 (2020). https://doi.org/10.1016/j.jclepro.2020.121065
5. Maier, M.W.: Architecting principles for systems-of-systems. Systems Engineering 1, 267–284 (1998).
6. Panetto, H., Zdravkovic, M., Jardim-Goncalves, R., Romero, D., Cecil, J., Mezgár, I.: New perspectives for the future interoperable enterprise systems. Computers in Industry 79, 47–63 (2016)
7. Ramos, A.L., Ferreira, J.V., Barcel´o, J.: Model-based systems engineering: An emerging approach for modern systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42(1), 101–111 (2012). https://doi.org/10.1109/TSMCC.2011.2106495

8. Tomaszuk, D., Angles, R., Thakkar, H.: Pgo: Describing property graphs in rdf. IEEE Access 8, 118355–118369 (2020).

Reverse Engineering in Software Requirements in web applications using a LEL *

Angela Verónica Granizo Rodríguez[1][0000−0001−8117−275X], Leandro Antonelli[2][0000−0003−1388−0337], Sergio Firmenich[2,3][1234−5678−9012], and Diego Firmenich[4][0000−0002−7212−4454]

[1] Escuela Superior Politécnica de Chimborazo. Panamericana Sur km 1 y 1/2 . Riobamba. Ecuador
agranizo@espoch.edu.ec
sergio.firmenich@lifia.info.unlp.edu.ar
[2] LIFIA, Facultad de Informática, UNLP. La Plata. Calle 50 y 120, S/N. Buenos Aires. Argentina.
lanto@lifia.info.unlp.edu.ar
[3] CONICET. Godoy Cruz 2290. Ciudad Autónoma de Buenos Aires. Argentina.
[4] Departamento de Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco. Argentina.
diego.firmenich@gmail.com

Abstract. This document explores and describes the state of the art of Reverse Engineering in the specification of Software Requirements for web applications using an LEL (Language Extended Lexicon). A review of the scientific literature on this topic was carried out with the aim of investigating how the application of Reverse Engineering in the specification of software requirements would improve their quality, with the support of the creation of a tool for web augmentation, that is, a web extension that allows creating a glossary of terms or LEL of web applications, this with the purpose of knowing their operation for maintenance and improvement purposes.

Keywords. Software Requeriments, Reverse Engineering, Language Extended Lexicon.

* Universidad Nacional de La Plata

1 Introducción/Motivación

La mayoría de los problemas que se dan en la entrega de los sistemas tienen que ver con conceptos erróneos de requisitos en la captura de requerimientos, lo que provoca falencias en su calidad. Por ello, es indispensable un proceso de ingeniería de requisitos (RE) para cumplir con los objetivos de tiempo, costo y calidad. Cuando las aplicaciónes web ya fueron construídas o en sistemas ya existentes o heredados, ya se

plasmaron los requerimientos y si estos tenían errores, para lograr conocer su funcionamiento para mantenimiento y mejoramiento se puede recurrir a Ingeniería Inversa; extraer los requerimientos de software. En este caso aparece la necesidad de contar con un LEL, que facilite el conocimiento del dominio de la aplicación. De acuerdo a los antecedentes descritos anteriormente, se puede plantear la siguiente pregunta de investigación ¿cómo se puede mejorar la calidad de los requerimientos de software en aplicaciones web aplicando Ingeniería Inversa y utilizando un LEL?

El presente trabajo describe el estado del arte y la propuesta de un enfoque que utilice Ingeniería Inversa en la especificación de requerimientos de software de aplicaciones web utilizando un LEL (Léxico Extendido del Lenguaje). Por lo tanto, vamos a crear herramientas de aumentación web como una extensión web que permita extraer contenidos del DOM que sirvan como referencia para la especificación de requerimientos de software.

2 Estado del arte

Las Empresas que poseen sistemas informáticos obsoletos se enfrentan a una encrucijada; continuar utilizando los sistemas heredados, realizar sus cambios requeridos y sus costos se incrementarán inevitablemente, por otro lado, reemplazar sus sistemas heredados con sistemas nuevos, esto también tendrá un costo y puede suceder que los nuevos sistemas no provean apoyo efectivo como los sistemas heredados. Lo mismo ocurre con las aplicaciones web que son sistemas, pero funcionan en la world wide web y son el motivo de investigación. Se puede referir varias razones del costo de mantenimiento de sistemas heredados o aplicaciones web con cierta antigüedad, por ejemplo: diferentes equipos han implementado distintas partes de la aplicación web, es decir existirán estilos de programación no consistentes. Todo o parte de la aplicación web se implementó usando un lenguaje de programación que luego ya es obsoleto, con dificultad se puede encontrar personal con conocimientos de los mismos o consultoría externa costosa, para dar mantenimiento. La documentación de la aplicación web no es apropiada y no está actualizada, la única existente es el código fuente o únicamente se cuenta con la versión ejecutable.

De acuerdo a este preámbulo, la ingeniería inversa de software es la forma de recuperar los requisitos o las especificaciones de diseño de un sistema, su objetivo es extraer muchos tipos de información del software existente y usarla para la renovación, mejor comprensión y ayuda en el mantenimiento de los sistemas heredados [1]. La ingeniería inversa es el proceso de analizar un sistema para identificar sus componentes y sus interrelaciones y crear representaciones del mismo en otra forma o en un nivel más alto de abstracción [2]. Las técnicas de ingeniería inversa existentes se centran en extraer la información de las funciones centrales y de las reglas negocio del software heredado, que se llama primary concerns [3].

Por otro lado, un proceso de ingeniería de requisitos es crucial para cumplir con los objetivos de tiempo, costo y calidad. Es necesario recuperar los requisitos del resultado de ingeniería inversa del sistema heredado y al integrar este resultado en la fase de requisitos del ciclo de vida del software, es posible tener una mejor obtención de requisitos [1].

Así también el Léxico Extendido del Lenguaje (LEL) o glosario de términos en la elicitación de requerimientos se ha propagado entre desarrolladores e investigadores en la Ingeniería de Requisitos. El glosario es un artefacto de especificación de requisitos de software utilizado para describir los elementos centrales en el contexto de la aplicación mediante la descripción de su vocabulario. El LEL apoya la construcción de un glosario y clasifica los símbolos en Sujeto, Objeto, Verbo y Estado, estos a su vez se decriben con los atributos de noción e impacto [4].

A continuación se describe una revisión de la literatura de varios autores; en primer lugar Wang et al. (2006), presentaron un enfoque de Ingeniería Inversa para recuperar el diagrama de clase UML del código fuente de Java [5]. Fahmi et al. (2007), expusieron un modelo revisado del proceso de reingeniería tradicional y también describieron la racionalidad del modelo propuesto. Han discutido sobre la ingeniería inversa de software, la ingeniería de requisitos, sus prácticas y actividades básicas. Mencionan que, si los requisitos del sistema heredado se conocen y se consideran al hacer la nueva especificación de requisitos, seguramente se puede mejorar la fase de requisitos. Argumentan la necesidad de obtener requisitos de los resultados de la ingeniería inversa [1]. Su et al. (2008), propusieron un framework de Ingeniería Inversa de software orientado a aspectos para la solución de comprensión y evolución de problemas de propiedades transversales (crosscutting properties) en sistemas heredados. El objetivo de este enfoque fue recuperar preocupaciones transversales del sistema a nivel de requisitos [3]. Aman et al. (2013), presentan un framework de ingeniería inversa de XML a UML para la generación de especificación de requisitos de software [6].

Todos los trabajos relacionados son muy relevantes y como se puede notar el uso de la Ingeniería Inversa de ha hecho presente de distintas formas, pero nadie se enfoca en el uso del LEL, si bien hacen Ingeniería Inversa en requerimientos de software, no se apoyan en el uso del LEL que es mi enfoque. Como propuesta, se plantea incluir el LEL en la especificación de requerimientos de aplicaciones web aplicando Ingeniería Inversa para conocer su funcionamiento con fines de mantenimiento y mejoramiento.

3 Planteamiento del problema

Como problema podemos describir, que la complejidad en la extracción de requerimientos desde sistemas heredados hace que los mismos tengan falencias en su calidad, puesto que no se cuenta con un documento de requerimientos y peor aún documentación con un léxico extendido del lenguaje (LEL), es decir que este enfoque permitirá mejorar la calidad de los requerimientos de software. Para alcanzar este objetivo se aplicará Ingeniería Inversa a aplicaciones web para conocer su funcionamiento con fines de mantenimiento y mejoramiento de las mismas. La contribución, es que no existe un enfoque que utilice Ingeniería Inversa para especificar requerimientos utilizando un LEL; por tanto, vamos a crear herramientas de aumentación web como una extensión web que permita extraer contenidos del DOM que sirvan como referencia para la especificación de requerimientos de software.

4 Metodología y enfoque de investigación

El enfoque de esta investigación es cuantitativo. Dentro del diseño de la investigación utilizaremos el método experimental, pues vamos a seleccionar dos grupos uno de control y otro experimental para probar la herramienta desarrollada. El tipo de investigación, es correlacional con dos variables involucradas; la independiente que será el enfoque de extracción de requerimientos de aplicaciones web y la variable dependiente la calidad de los requerimientos. La idea es verificar si la correlación entre estas variables es significativa.

5 Plan de evaluación

Dentro del enfoque propuesto, se está desarrollando la herramienta de aumentación web, es decir la extensión web de una aplicación ya construida, misma que permite crear el LEL. También se está trabajando en una revisión sistemática acerca de Ingeniería Inversa en Requerimientos de Software. Posteriormente se realizará un experimento con analistas de desarrollo a los cuales se dividirá en dos grupos, con y sin el enfoque y veremos los resultados.

6 Etapa doctoral

Early Stage -ES-

References

1. S. A. Fahmi and H.-J. Choi, “Software Reverse Engineering to Requirements,” 2007, Accessed: Mar. 05, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/4420580/.
2. E. J. Chikofsky and J. H. Cross, “Reverse Engineering and Design Recovery: A Taxonomy,” IEEE Softw., vol. 7, no. 1, pp. 13–17, 1990, doi: 10.1109/52.43044.
3. Y. Su, X.-W. Zhou, and M.-Q. Zhang, “Approach on Aspect-Oriented Software Reverse Engineering at Requirements Level,” in 2008 International Conference on Computer Science and Software Engineering, 2008, vol. 2, Accessed: Mar. 06, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/4722062/.
4. M. Urbieta, L. Antonelli, G. Rossi, and J. C. S. do Prado Leite, “The impact of using a domain language for an agile requirements management,” Inf. Softw. Technol., vol. 127, p. 106375, Nov. 2020, doi: 10.1016/J.INFSOF.2020.106375.
5. X. Wang and X. Yuan, “Towards an AST-Based Approach to Reverse Engineering,” 2006, Accessed: Mar. 06, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/4054609/.

6. H. Aman and R. Ibrahim, “Reverse Engineering: From Xml to Uml for generation of software requirement specification,” 2013, Accessed: Mar. 16, 2022. [Online]. Available:https://ieeexplore.ieee.org/document/6637575/.

Methods and tools for the abstraction of object models in web content by end users *

Alex Tacuri[1,2][0000−0003−3159−5556]

[1] Escuela Superior Politécnica de Chimborazo. Panamericana Sur km 1 y 1/2 . Riobamba. Ecuador
[2] LIFIA, Facultad de Informática, UNLP. La Plata.
atacuri@espoch.edu.ec

Abstract. The main idea is to organize the data exposed in Web site interfaces so that they can be processed or used efficiently, either by a person or automated by a machine through an algorithm. The structure or model that will be used in this approach is that of a graph, where each node will be an object abstracted from the DOM. From the point of view of external structures, currently the information on the web is very little combinable, to be able to establish connections or think about object relationships, which leads to not being able to process information effectively and efficiently, In addition, automating a task becomes complex, so providing a user with a model in which they can integrate this data and information will greatly alleviate this lack of integration.

Keywords. Web search, End-user Programming, Web augmentation.

* Universidad Nacional de La Plata

1 Introducción/Motivación

La web actualmente presenta un sinnúmero de sitios web que presentan información muy importante sobre temas de actualidad, pero si reflexionamos un momento, muchos de estos hablan de un mismo tema, con particularidades de cada uno, que si los integramos podemos alcanzar una mejor compresión del mismo; si bien es cierto, esto se lo realiza de manera inconsciente, puesto que, al momento de obtener lo mejor de un sitio, procedemos a buscar en la web otra información relacionada para seguir enriqueciendo el contenido. El objetivo central de este enfoque es primeramente obtener datos de sitios web por medio de Search APIs, mismas que crean una estructura de cómo acceder al contenido de un sitio en específico, para posteriormente integrarlo en un modelo de datos relacionados, con el fin de ofrecer el resultado de

esta integración a través del mismo navegador por medio de una extensión web; todo esto se realiza al momento de invocar a esta capa de servicio. Además recalcar que todo este procedimiento puede ser realizado por usuarios sin habilidades de programación.

2 Estado del Arte

Pensando en la integración y reutilización de información, existen modelos de desarrollo de aplicaciones que se fundamentan en el proceso de concepción orientado a objetos, uno de ellos es el modelo OOHDM (del inglés, «Object Oriented Hypermedia Design Methodology») mismo que en su primera fase, se debe realizar un diseño conceptual de la aplicación hipermedia, con objetos de dominio o clases y relaciones entre ellos que posteriormente pasan a un diseño de navegación, diseño de interfaces e implantación [1]. Otra metodología es WebML, que de igual forma en su primera fase se fundamenta en la creación de un modelo conceptual de objetos que sirve para la navegación y composición de la aplicación, para su posterior presentación [2]. Desde la perspectiva de que un usuario final pueda construir estructuras externas, sería ideal contar con un modelo de objetos que permita basarse en la extracción, vinculación de información de manera reutilizable. Para realizar este modelo podemos aprovechar que el desarrollo de aplicaciones web se basan fundamentalmente en un diseño basado en objetos, los cuales tienen, como fin, aliviar la complejidad y la reutilización de los mismos, que luego de un proceso de integración o composición, llegan a presentarse al usuario final como instancias de los objetos mediante una interfaz de usuario [3]. Sirviéndonos de esta funcionalidad, y dado que estas solicitudes son finalmente presentadas al usuario de alguna manera en la UI, se puede crear un modelo de objetos en el cliente mediante la extracción de la información de ellas, pensando en la reutilización de estos en diversas aplicaciones.

Una vez que el usuario a abstraído información de la web, se propone crear un editor de modelo de objetos, donde se podrán establecer propiedades y relaciones entre ellos, mismos que son provenientes de distintas páginas web a través de web scraping y automatización de la navegación, pero definido a nivel de objetos del dominio. Con este proceso se alcanza la integración e interoperabilidad de las aplicaciones en una capa creada por el usuario final, mismo que podrá interactuar con el modelo, a través de un lenguaje de consultas y una interfaz interactiva que permita la afinación del modelo, y el acceso a objetos e instancias de sitios que hayan sido agregados. Para la elaboración de modelos de objetos web se piensa utilizar técnicas de abstracción de contenidos web con web scraping y aumentación web en el lado del cliente para la definición de editores visuales. Ahora veamos cómo se encuentran estas técnicas en los últimos años; dentro de la WA han aparecido varios interesantes enfoques para considerar. En este año (2020), Hertel et al. presentaron un enfoque donde incorporaron a todos los involucrados en un sitio web a que puedan participar en las actividades de rediseño web con WA y Desarrollo por Usuarios Finales, del inglés «End-User Development» [4]. Gonzales et al. expusieron otro enfoque donde crearon un framework para facilitar el acceso de los usuarios con discapacidad visual dentro del navegador apoyando la accesibilidad web de los sitios [5]. Adicionalmente el mismo autor et al. desarrollaron una api para que los programadores que no están

familiarizados con SPARQL puedan hacer uso de Linked Open Data y puedan acceder de forma más fácil a esta red de información [6]. Sottile et al. (2019), presenta la creación de Aumentos de Web Semánticos [7] y Firmenich et al. desarrollaron una plataforma para interfaces de usuario distribuidas (DUI, del inglés «Distributed User Interface») del lado del cliente, construida sobre los cimientos del aumento web y el desarrollo del usuario final con la finalidad de sincronizar varias interfaces de usuario final [8]. Fernández et al. presentaron a "Logikós" un desarrollo que permite que los usuarios tengan una herramienta para la toma de decisiones, de múltiples criterios en diferentes sitios web [9]. Bosetti et al. presentaron una herramienta de visualización para describir las operaciones fundamentales necesarias para visualizar datos semiestructurados en la Web [10]. En el año 2018, Firmenich et al. presentaron un enfoque para definir y hacer evolucionar los requisitos de aumento web utilizando prototipos visuales enriquecidos y descripciones textuales, que se pueden mapear automáticamente en artefactos de software en ejecución. Bosetti et al. desarrollaron el enfoque para la Aumentación Web Móvil (MoWA, del inglés «Movil Web Augmentation»), En el año 2017, Aldalur et al. identificaron a WA como una tecnología prometedora para EUD [11]. En el 2013, Capra et al. presentaron búsquedas subrogantes con imágenes es decir agregando imágenes a una búsqueda para un mejor análisis de búsqueda [12]. En todos estos enfoques, radica la idea de tener algún tipo de abstracción y/o extracción del contenido público en la Web, pero ningún ataca la problemática específica que es poder programar las estructuras externas basadas en modelos de objetos complejos, que van más allá de solo crear una indirección al elemento del DOM concreto.

3 Planteamiento del problema/Contribuciones

Actualmente la extracción de datos de páginas web (*web scraping*) es utilizado con diferentes propósitos; pero un problema que se ha encontrado es que los datos que se obtuvieron de diferentes fuentes no se encuentran integrados al momento de la extracción; proceso que se lo realiza posteriormente a la abstracción del contenido de la web. En este sentido, nuestro enfoque pretende extraer contenido e integrarlo de forma automática pero no solo por usuarios que dominan programación sino también por usuarios finales sin experiencia en desarrollo de sistemas. Para alcanzar este objetivo se pretende elaborar herramientas con aumentación web que permitan abstraer el contenido web (Search APIs), integrar el contenido y formar modelos de objetos que permitan un mejor procesamiento de la información de diferentes fuentes de datos heterogéneas. Adicionalmente se ha logrado crear una capa de servicio en el navegador que permite que otras aplicaciones puedan consumir datos de esta capa, un ejemplo de este consumo es una aplicación que consume datos de un modelo que integra 3 buscadores: Google Escolar, Springer y DBLP; permitiendo ejecutar búsquedas en estas 3 Search APIs e integra los resultados de cada uno.

4 Metodología de Investigación y Enfoque

Dos grupos uno de control y otro experimental para probar las herramientas desarrolladas. El tipo de investigación bajo la cual se va a manejar la investigación es correlacional puesto que vamos a tener dos variables involucradas; la independiente que será el enfoque de desarrollo de modelos web a partir de abstracciones de la web , y la variable dependiente mejorar la integración de información por parte del usuario final. La idea es verificar si la correlación entre estas variables es significativa. La investigación se va a realizar en la Universidad Nacional de la Plata – Argentina y en Escuela Superior Politécnica de Chimborazo – Ecuador.

5 Plan de Evaluación

Dentro del desarrollo de este enfoque se han realizado los siguientes componentes: una herramienta de abstracción de contenidos web basado en webscraping (Search API), un editor de modelos de objetos que permita crear relaciones y configuraciones entre los objetos previamente abstraídos, una consola interactiva que me permita buscar e interactuar con instancias de objetos basados en el modelo desarrollado por el usuario final utilizando tags semánticos y propiedades, definiendo para esto un lenguaje de consultas sobre el modelo y una aplicación de aumentación web tipo mashup que consume la capa de servicio que se encuentra en el browser como extensión web y permite realizar la consulta del modelo.

Actualmente se está trabajando en realizar el experimento con usuarios finales con y sin experiencia en programación y poder mostrar de forma cuantitativa la validez de las herramientas desarrollados y el enfoque como tal. Posteriormente se pretende utilizar técnicas de machine learning para soportar el proceso de construcción de modelos de objetos o tareas autómatas durante el desarrollo del enfoque.

6 Resultados Preliminares o Intermedios

Se público el artículo “ANDES: An approach to embed search services on the Web browser” en “Computer Standards & Interfaces” sobre Search APIs, indexada en ScienceDirect. DOI: https://doi.org/10.1016/j.csi.2022.103633

Actualmente se desarrolló la capa de servicios que se ejecuta como extensión del browser y es consumida por otra aplicación para hacer consultas, para el experimento vamos a utilizar 3 Search APIs: Springer, DBLP y Google Scholar.

7 Conclusiones y Lecciones Aprendidas

El tiempo en integrar información de varios sitios de un mismo contenido se reduce significativamente logrando mejorar la productividad al momento de buscar información.

8 Etapa Doctoral

Middle.

References

1. J. Molina-Ríos and N. Pedreira-Souto, “Comparison of development methodologies in web applications,” Information and Software Technology, vol. 119. Elsevier B.V., Mar-2020.
2. D. Granada, J. M. Vara, M. Brambilla, V. Bollati, and E. Marcos, “Analysing the cognitive effectiveness of the WebML visual notation,” Softw. Syst. Model., vol. 16, no. 1, pp. 195–227, Feb. 2017.
3. A. Abouzahra, A. Sabraoui, and K. Afdel, “Model composition in Model Driven Engineering: A systematic literature review,” Information and Software Technology, vol. 125. Elsevier B.V., p. 106316, Sep-2020.
4. H. Hertel, A. Dittmar, and D. Linke, “Meta-level support for facilitating participation in website (re-)design activities,” in Proceedings of the 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems, 2020, pp. 1–6.
5. C. González-Mora, I. Garrigós, S. Casteleyn, and S. Firmenich, “A web augmentation framework for accessibility based on voice interaction,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020, vol. 12128 LNCS, pp. 547–550.
6. C. González-Mora, I. Garrigós, and J. Zubcoff, “An apification approach to facilitate the access and reuse of open data,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020, vol. 12128 LNCS, pp. 512–518.
7. C. Sottile, S. Firmenich, and D. Torres, “An End-User Semantic Web Augmentation Tool,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, vol. 11553 LNCS, pp. 239–243.
8. S. Firmenich, G. Bosetti, G. Rossi, M. Winckler, and J. M. Corletto, “Distributed Web browsing: supporting frequent uses and opportunistic requirements,” Univers. Access Inf. Soc., vol. 18, no. 4, pp. 771–784, Nov. 2019.
9. A. Fernández, G. Bosetti, S. Firmenich, and P. Zaraté, “Logikós: Augmenting the Web with Multi-criteria Decision Support,” in Lecture Notes in Business Information Processing, 2019, vol. 348, pp. 123–135.
10. G. Bosetti, S. Firmenich, M. Winckler, G. Rossi, U. C. Fandos, and E. EgyedZsigmond, “An end-user pipeline for scraping and visualizing semi-structured data over the web,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, vol. 11496 LNCS, pp. 223–237.
11. I. Aldalur, M. Winckler, O. Díaz, and P. A. Palanque, “Web Augmentation as a Promising Technology for End User Development,” in New Perspectives in EndUser Development., F. Paternò and V. Wulf, Eds. Springer International Publishing, 2017, pp. 433–459.

12. R. Capra, J. Arguello, and F. Scholer, “Augmenting web search surrogates with images,” in 22nd ACM International Conference on Information and Knowledge Management, CIKM’13, San Francisco, CA, USA, October 27 - November 1, 2013, 2013, pp. 399–408.

Method and engineering tools for in-vehicle information systems (In-Vehicle Information Systems), focusing on the risk of driver distraction

Carlos Carvajal[1], Andrés Rodriguez[1] and Alejandro Fernández[1]

[1] Lifia - Fac. de Informática, Universidad Nacional de La Plata, La Plata, AR
carlos.carvajal@lifia.info.unlp.edu.ar
andres.rodriguez@lifia.info.unlp.edu.ar
casco@lifia.info.unlp.edu.ar

Abstract: Thousands of deaths a year are attributed to driver distractions while driving. The automotive industry has worked for decades maturing development guidelines and complex evaluation scenarios for the acceptance of infotainment centers as an integral part of the vehicle. The massification of smart wearable devices with a gigantic ecosystem of mobile applications not designed for the particularities of the automotive environment, easily bypasses the precautions of safe driving schemes. There is then, a niche of developers without the formal knowledge and economic support to adequately enter the automotive applications development. Cost-benefit methods and tools accessible in the context of software engineering are proposed to support the development of in-vehicle information systems in accordance with the requirements of current regulations and standards.

1. Introducción/Motivación

El conductor, el auto y el camino. En un mundo ideal es el contexto y con quienes exclusivamente se debería interactuar al conducir. Primero aparecieron los acompañantes humanos con sus potenciales conversaciones, luego funcionalidades de apoyo y bienestar como controles de aire acondicionado, radio/casetera/CD/DVD. La evolución automotriz continuó con la inclusión de pantallas para información/entretenimiento y sistemas de navegación GPS. Durante décadas estos elementos potencialmente distractivos dispararon alertas en la industria automotriz y agencias regulatorias gubernamentales, que formularon guías de diseño para normar y controlar la fabricación de dispositivos que pudiesen competir por la atención del conductor. Coincidiendo con el lanzamiento del primer iPhone en el año 2007, 3 factores se comenzaron a alinear y empujaron en aumentar el riesgo de distracción del conductor: estandarización de teléfonos inteligentes con pantallas táctiles, ubiquidad del acceso móvil de Internet y consolidación de ecosistemas para el desarrollo y

distribución generalizada de aplicaciones móviles. Bajo este contexto se acuñó la expresión: "dispositivos nómadas", que no son parte original del auto, son portables, y con la posibilidad de ejecución de software ideado para un contexto en el que la atención del usuario se maximiza. Esta situación genera nuevos retos para la industria automotriz y agencias gubernamentales de control.

Por nuestro lado planteamos aportar con un enfoque de ingeniería de software que soporte a los desarrolladores de sistemas de información para el vehículo que no cuentan con el apoyo económico para cumplir con los costosos y rigurosos métodos de evaluación formales. Para esto, proponemos poner foco en etapas tempranas de desarrollo en donde el prototipado puede estar incluso aún en el papel. Métodos de interacción humano – computador plantean la predicción del tiempo de ejecución de tareas dentro del contexto de una aplicación móvil. Adaptaciones de estas metodologías se han implementado para el vehículo. La idea es la de aportar una herramienta que permita la evaluación y comparación de diferentes opciones de escenarios de interfaces, y así el desarrollador tenga pistas tempranas antes invertir tiempo y dinero en etapas posteriores con prototipos funcionales.

2. Estado del arte

Los conocidos como métodos analíticos son útiles para realizar predicciones sobre la usabilidad de los productos sin necesidad de prototipos robustos y pruebas con usuarios, que muy probablemente pueden tener un alto costo en dinero y tiempo[1][2][3]. La evaluación empírica de un sistema de información vehicular podría llevar semanas o meses, mientras que un enfoque analítico permite predecir los tiempos de las tareas de las aplicaciones en mucho menos tiempo[4]. Los métodos analíticos son adecuados para su aplicación en las primeras fases en el ciclo de vida del producto debido a su baja demanda de recursos [5].

El performance de una tarea es una medida representativa utilizada para investigar interacciones entre el usuario y un sistema computacional [6]. Parte fundamental de estos estudios ha sido el predecir el tiempo de ejecución de tareas en función de variables como el contexto y característica del usuario. El método KLM es basado en el modelo GOMS de metas (Goals), operadores (Operators) y métodos (Methods) desarrollado por Card et al. [7]. El concepto radica en que el tiempo de ejecución de tareas puede ser predicho mediante la definición de conceptos como: tipo de operadores, tiempo de operador y el número de repeticiones del operador. El método divide las acciones requeridas para una tarea en múltiples tareas unitarias, y se basa en el tiempo correspondiente predefinido para cada tarea unitaria. El método KLM (keystrokelevel model) como su nombre lo indica, fue originalmente ideado para interacciones con teclados. Varias adaptaciones se han realizado y verificado en el contexto de interfaces táctiles[8][9], para esto se requirieron añadir operadores específicos para escenarios como: zoom, deslizar y tocar. En el contexto automotriz operadores adicionales han sido introducidos validando la adaptación del modelo KLM [10][11][12]. Como ejemplos de nuevos operadores tenemos: movimiento de mano entre volante y la pantalla táctil, movimiento de la atención entre la mirada al frente al conducir y la pantalla táctil, entre los principales.

3. Planteamiento del problema y contribuciones

Si bien es cierto la industria automotriz y entidades gubernamentales de control han trabajado durante décadas en metodologías, guías de diseño para normar y controlar el riesgo de distracción del conductor. Así mismo, los sistemas operativos móviles predominantes han creado entornos listos para el automóvil en sus ecosistemas de aplicaciones[13][14]. Sin embargo, la facilidad actual para que desarrolladores de todo tipo creen aplicaciones distribuibles de forma generalizada, sin los controles para el contexto automotriz es muy alto.

Los métodos analíticos resultan como una solución útil para soporte de los desarrolladores de aplicaciones móviles ya que permiten validar interfaces y flujos de tareas sin necesidad de avanzar en el desarrollo de prototipos funcionales o invertir en demandantes pruebas empíricas. Sin embargo, la modelización de la interacción humano computador potencialmente tiene un alto costo en términos del conocimiento necesario para crear los modelos, el aprendizaje, la comprensión de complejas teorías bases de cada modelo, y el tiempo que se puede tomar para generar los resultados. Justamente la contribución que se plantea es el concepto de una herramienta que permita abstraer los detalles de parametrización, configuración y resultados para el desarrollador o diseñador. En general, el objetivo es que no se tenga necesidad de invertir en conocimiento detallado de métodos predictivos. Así mismo se plantea que la herramienta contribuya con alertas cuando una tarea o interfaz evaluada, cuantitativamente presente resultados no acordes con escenarios de aceptación estandarizados en el entorno automotriz[15].

4. Metodología y enfoque de la investigación

Se plantea abordar los siguientes temas:

- Análisis del estado del arte de las diferentes aristas que comprenden los temas de investigación propuestos.
- Definición de brecha de investigación en la cual contribuir de forma original.
- Definición de prueba de concepto acorde a los modelos seleccionados.
- Representación de resultados con la metodología y herramientas propuestas.
- Demostración de resultados.

5. Plan de evaluación

6. Resultados preliminares o intermedios

Carvajal C., Rodríguez A., Fernández A., "The growing and risky industry of nomadic apps for drivers", 2020. http://sedici.unlp.edu.ar/handle/10915/119031

7. Conclusiones y lecciones aprendidas

Consideramos que todavía nos encontramos en una etapa muy preliminar de desarrollo de la presente tesis doctoral como para presentar conclusiones maduras y definitivas.

En el camino recorrido se ha evaluado diferentes líneas de desarrollo que en preliminarmente sonaban prometedoras. Revisando en retrospectiva, un factor en común que han tenido los descartes es el entender que han sido conceptos casi sin explorar por otros investigadores. En su momento esto se validó como un tema interesante en el contexto doctoral, pero a la larga la experiencia que ha quedado es que la falta de mayores investigaciones ha resultado en teorías poco útiles y con mínima practicidad para ser implementadas.

Una grata sorpresa ha resultado el conocer la existencia de una comunidad radicada en Latinoamérica que trabaja en el estudio de sistemas de información en el vehículo y sus desafíos. Interesantes oportunidades de colaboración y trabajo en equipo se presentan.

8. Etapa doctoral

Etapa intermedia.

Referencias

1. D. D. Salvucci, "Modeling tools for predicting driver distraction," *Proc. Hum. Factors Ergon. Soc.*, pp. 1149–1152, 2005.
2. M. Pettitt, G. Burnett, and A. Stevens, "An Extended Keystroke Level Model (KLM) for Predicting the Visual Demand of In-Vehicle Information Systems," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2007.
3. C. Harvey and N. A. Stanton, "Modelling the hare and the tortoise: predicting the range of in-vehicle task times using critical path analysis," *Ergonomics*, vol. 56, no. 1, pp. 16–33, 2013.
4. D. Manes, P. Green, and D. Hunter, "Prediction of destination entry and retrieval times using keystroke-level models," *Umtri*, 1997.
5. D. E. Kieras and D. E. Meyer, "An Overview of the EPIC Architecture for Cognition and Performance with Application to Human-Computer Interaction," vol. 5, no. 5, 1995.
6. S. C. Lee, S. H. Yoon, and Y. G. Ji, "Modeling task completion time of in-vehicle information systems while driving with keystroke level modeling," *Int. J. Ind. Ergon.*, vol. 72, no. June, pp. 252–260, 2019.
7. S. K. Card, T. P. Moran, and A. Newel, "The KLM for User Performance Time with Interactive Systems," *Commun. ACM*, vol. 23, no. 7, pp. 396–410, 1980.

8. A. D. Rice and J. W. Lartigue, “Touch-Level Model (TLM): Evolving KLM-GOMS for Touchscreen and Mobile Devices,” pp. 1–6, 2014.
9. A. Lee, K. Song, H. B. Ryu, J. Kim, and G. Kwon, “Fingerstroke time estimates for touchscreen-based mobile gaming interaction,” *Hum. Mov. Sci.*, vol. 44, pp. 211–224, 2015.
10. M. Pettitt, “An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems,” *Hum. Factors Comput. Syst.*, 2007.
11. C. Purucker, F. Naujoks, A. Prill, and A. Neukum, “Evaluating distraction of in-vehicle information systems while driving by predicting total eyes-off-road times with keystroke level modeling,” *Appl. Ergon.*, vol. 58, pp. 543–554, 2017.
12. S. Stefan, B. Pfleging, D. Kern, and A. Schmidt, “Support for modeling interaction with automotive user interfaces,” *Proc. 3rd Int. Conf. Automot. User Interfaces Interact. Veh. Appl. AutomotiveUI 2011*, pp. 71–78, 2011.
13. Google, “Android Auto.” [Online]. Available: https://www.android.com/auto/. [Accessed: 28-Nov-2019].
14. Apple Inc., “iOS - CarPlay - Apple.” [Online]. Available: https://www.apple.com/ios/carplay/. [Accessed: 28-Nov2019].
15. P. Green, “Estimating Compliance with the 15-Second Rule for Driver-Interface Usability and Safety,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 43, no. 18, pp. 987–991, 1999.

Adaptive gamification of citizen science projects

María Dalponte Ayastuy[1,2][0000-0002-1412-5694] and Diego Torres[1,2][0000-0001-7533-0133]

[1] Depto CyT, Universidad Nacional de Quilmes
Roque Saenz Peña 352, Bernal, Buenos Aires, Argentina
mdalponte@unq.edu.ar
[2] LIFIA,CICPBA-Facultad de Informatica, Universidad Nacional de La Plata
50 y 120, La Plata, Buenos Aires, Argentina
diego.torres@lifia.info.unlp.edu.ar

Abstract. Mass collaboration mediated by technology is now commonplace (Wikipedia, Quora, TripAdvisor). Online mass collaboration is also present in science in the form of Citizen Science. These collaboration models, which have a large community of contributors coordinated to pursue a common goal, are known as Collaborative systems. Gamification is a strategy to convene participants to CS projects. However, it cannot be generalized because of the different users' profiles, and so it must be tailored to the users and playing contexts. This Ph.D. project approaches the problem of adapting gamification to the user's profile, project objectives, and global status.

Keywords. Citizen Science, Gamification, Adaptive Gamification, Collaborative Project.

1 Introduction

Citizen science encompasses a range of methodologies that encourage and support the contributions of the public to the advancement of scientific and engineering research and monitoring. The public contributes co-identifying research questions; co-designing/conducting investigations; co-designing/building/testing low-cost sensors; co-collecting and analysing data; co-developing data applications; and collaboratively solving complex problems [13].

Citizen science has become widely known in recent years thanks to the ubiquity of technology through communication technologies and the mass use of smartphones. In addition, there is a growing number of scientific projects and volunteers that collect data through their daily used resources. Consequently, research interest is awakened for the design, development, and implementation of the technologies that are needed for the exercise of citizen science [12].

As a strategy to reach the most significant number of people, considering the multiple cultural characteristics (origin, language, gender, age, among others) can be approached by scientific developments in the area of HCI (human computer interaction), and particularly gamification [12]. Gamification is the application of game strategies in spaces or areas whose nature is not playful [6].

The natural use of gamification is in citizen science projects [8], and there are already examples of gamified experiences. Although in some of these approaches, the exact related game mechanisms have different impacts on different people, the use of gamification elements may be more valued by some volunteers than by others. Some found it motivating and rewarding, while others ignored it or made them stop participating in the project. For some time, the HCI has been working on the formalization of playability heuristics and models of the components of games and game experiences [6].

Despite the rapid growth of the gameful design research area, and the current level of success in the user engagement that it reveals, these findings are not general in terms of the domain. Neither can be generalized to all users. The one-size-fits-all approach presents several limitations because of the different motivations, personalities, needs, or values of the users [1]. The design of game environments that are appropriate for everyone must consider a personalization or adaptation of the game elements and mechanics offered for each volunteer in each case. This adaptation should recognize the cultural aspects of the people and the interaction between them. Currently, the research stream on adaptive gamification is taking care of the gamification that each particular user needs in a particular moment, tailoring the gamification to the users and contexts[9].

For example, adaptation can be made on many aspects: the game storytelling, the game difficulty, the content generation, the guidance or hinting on the goals, the presentation, the curriculum sequencing, among others [7]. Nevertheless, the existing adaptive gamification approaches are not directly applicable to citizen science, given that they do not necessarily focus on the community aspect or project's objectives.

2 State of the art

A systematic mapping was carried out [11]To identify representative studies related to adaptive gamification and CLCS. The review allowed identifying different proposals for the scope (standard, ad-hoc or flexible) and the user model's versatility (dynamic vs. static). However, it was found that in most of the cases, the model is neither defined nor explicitly specified.

As a result of the evaluation, there were found different adaptation points of view, such as difficulty adaptation, storytelling adaptation, community-based adaptation, or gamification elements adaptation, where the goals/challenges and points are the most used. The user modeling is also important for an adaptation strategy and must be considered the scope of the model (standard, ad-hoc or flexible) and the versatility

(dynamic vs. static). The aspect that deserves further research is the adaptability considering the community, focusing on features that have not yet been worked on, such as cultural diversity, gender, and multiplicity of knowledge. Also, it is interesting to develop an approach of community modeling in community-aware adaptive gamification. These findings are compiled in [5].

3 Problem Statement and Contributions

The contributions of this Ph. D. project should include: a) a state of the art research, b) a conceptual framework modeling user behavior profile as well as the citizen science ecosystem and the game elements and mechanics applicable to these projects, c) an adaptation device based on machine learning techniques, to tailor the gamification to the users and community.

4 Methodology and Evaluation Plan

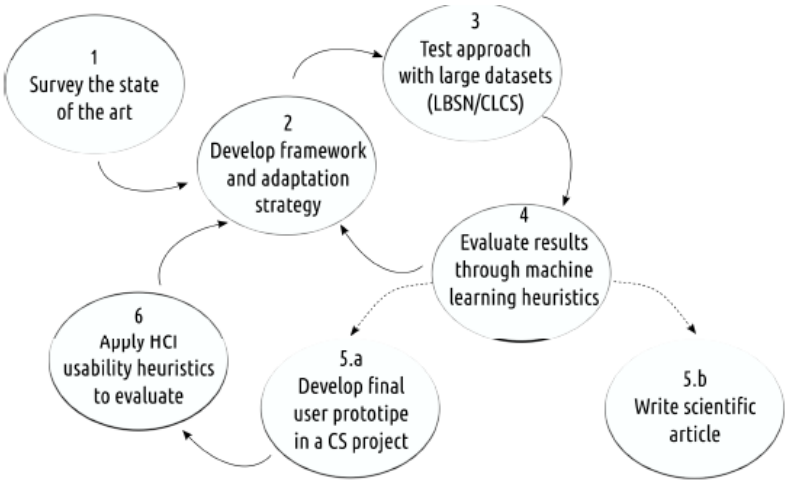


Fig. 1. Methodology

The main objective of this Ph.D. project is to develop and design an adaptive gamification approach in the context of citizen science projects. This objective can be decomposed into other finer-grained objectives. Firstly, to build a conceptual

framework as a set of modeling elements and relationships that can describe the users' behavior through their interaction activity (among them and with the system); the citizen science community ecosystem; and the applicable game elements and mechanics to these collaborative projects. Lastly, to design an adaptation strategy based on machine learning techniques to tailor the game experience using the user behavior, community, and game elements models. The research methodology is a process with the following steps: 1) to survey state of the art in the game elements adaptability research area, related to collaborative collecting projects and citizen science, in particular, 2) Development of an approach over the conceptual framework and the adaptation strategy, 3) testing the approach on historical data of large scale datasets from Location-Based Social Networks (LBSN) or Collaborative Location Collecting Systems projects (CLCS), 4) to evaluate results by applying machine learning quality heuristics, 5) to consider these partial results to give place to new concerns and the potential writing of a scientific article, 6) to develop a prototype for end-users where this adaptation strategy is incorporated to a citizen science application, 7) to conduct heuristic evaluations for usability in HCI [10]. Steps 2 to 7 are part of an iterative process (see Figure 1).

5 Preliminary or Intermediate Results

The project has reached the following milestones:

- Adaptive Gamification in Collaborative systems, a Systematic Mapping Study: a study of the published research on the application of adaptive gamification to collaborative systems

The study focuses on works that explicitly discuss an approach to personalization or adaptation of the gamification elements in this system. It employs a systematic mapping design in which a categorical structure for classifying the research results is proposed based on the topics that emerged from the papers' review. Published in Computer Science Review journal [5]

- Adaptive gamification in collaborative location collecting systems: a case of traveling behavior detection

This work is focused on the first steps to detect users' behavioral profiles related to spatial-temporal activities in the context of collaborative location collecting systems. This article was accepted in IX Conference on Cloud computing, Big Data, and Emerging Topics.

- Adaptive gamification in collaborative location collecting systems: a case of traveling behavior detection

This work is focused on the first steps to detecting users' behavioral profiles related to spatial-temporal activities in the context of CLCS. Specifically, this article introduces: (1) a strategy to detect patterns of spatial-temporal activities, (2) a model to describe the spatial-temporal behavior of users based on (1), and a strategy to

detect users' behavioral patterns based on unsupervised clustering. The approach is evaluated over a Foursquare dataset. The results showed two types of behavioral atoms and two types of users' behavioral patterns. [4]

- Relevance of non-activity representation in traveling user behavior profiling for adaptive gamification

This work presents two approaches of traveling user behavior profiling: a raw series built up with categorical data that describes the activity of the user in a period, and a timed series that is an enhanced version of the first that includes a representation of the non-activity time frames. This article seeks to analyze what aspects the different representations can contribute to describing user behavior categories to offer a tailored gamification strategy. Conference paper published in the XXI international conference in HCI [2].

- A model of adaptive gamification in collaborative location-based collecting systems

This article presented an automatic game challenge generation approach for CLCS. The needs and characteristics of the CLCS are presented, such as the space-time objectives and the space-time user behavior, to later be valued during the automatic generation of game challenges. The contributions are a user profile model considering the space-time behavior and challenge completion, a model for the different types of challenges applicable in CLCS, a model for the CLCS objectives and coverage, and a strategy for the application of Machine Learning techniques for adaptation. Conference paper published in the 3rd International Conference on Artificial Intelligence in HCI [3].

6 Conclusions and Learned Lessons

The project's next step is to develop the game challenge recommendation system on synthetic data or from an existing CLCS. The recommendation must then be incorporated as a gamification strategy into an end-user prototype.

7 Ph.D. Stage

This PhD project can be considered in a middle stage.

References

1. Böckle, M., Novak, J., Bick, M.: Towards adaptive gamification: a synthesis of current developments. In: Proceedings of the 25th European Conference on Information Systems (ECIS). Guimarães, Portugal (2017)
2. Dalponte Ayastuy, M., Torres, D.: Relevance of non-activity representation in traveling user behavior profiling for adaptive gamification. In: Proceedings of the XXI International Conference on Human Computer Interaction. Interacción '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3471391.3471431.
3. Dalponte Ayastuy, M., Torres, D., Fernández, A.: A model of adaptive gamification in collaborative location-based collecting systems. In: Degen, H., Ntoa, S. (eds.) Artificial Intelligence in HCI. pp. 201–216. Springer International Publishing, Cham (2022)
4. Dalponte Ayastuy, M., Torres, D.: Adaptive gamification in collaborative location collecting systems: a case of traveling behavior detection. Journal of Computer Science and Technology 22(1), e05 (Apr 2022). https://doi.org/10.24215/16666038.22.e05
5. Dalponte Ayastuy, M., Torres, D., Fernández, A.: Adaptive gamification in Collaborative systems, a systematic mapping study. Computer Science Review 39, 100333 (2021). https://doi.org/10.1016/j.cosrev.2020.100333
6. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: Defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. p. 9–15. MindTrek '11, Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/2181037.2181040
7. Göbel, S., Wendel, V.: Personalization and adaptation. In: Serious Games, pp. 161–210. Springer (2016).
8. Kapp, K.M.: The Gamification of Learning and Instruction Fieldbook: Ideas into Practice. Pfeiffer & Company, Bloomsburg, Pennsylvani, 1st edn. (2013).
9. Klock, A.C.T., Da Cunha, L.F., de Carvalho, M.F., Rosa, B.E., Anton, A.J., Gasparini, I.: Gamification in e-learning systems: A conceptual model to engage students and its application in an adaptive e-learning system. In: International Conference on Learning and Collaboration Technologies. pp. 595–607. Springer (2015)
10. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 249–256. CHI '90, Association for Computing Machinery, New York, NY, USA (1990). https://doi.org/10.1145/97243.97281
11. Petersen, K., Vakkalanka, S., Kuzniarz, L.: Guidelines for conducting systematic mapping studies in software engineering: An update. Information and Software Technology 64, 1–18 (Aug 2015). https://doi.org/10.1016/j.infsof.2015.03.007
12. Preece, J.: Citizen science: New research challenges for human–computer interaction. International Journal of Human–Computer Interaction 32(8), 585–612 (2016). https://doi.org/10.1080/10447318.2016.1194153
13. Vohland, K., Land-zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., Wagenknecht, K. (eds.): The Science of Citizen Science. Springer Nature (2021). https://doi.org/10.1007/978-3-030-58278-4

Integrated Requirements Engineering Framework with Intelligent Systems and Semantic Reasoning for the Extraction of Application Vocabularies

Alexandra Corral[1][0000-0002-9225-1346], Luis E.Sanchez[2][0000-0003-0086-1065] and Leandro Antonelli[3][0000-0003-1388-0337]

[1] University of the Armed Forces ESPE, Sangolquí, Ecuador
macorral@espe.edu.ec

[2] GSyA Research Group, University of Castilla-La Mancha, Ciudad Real, Spain
luise.sanchez@uclm.es

[3] Lifia – Facultad de Informatica, National, University of La Plata, La Plata, Argentina
lanto@lifia.info.unlp.edu.ar

Abstract. Domain discovery is an essential activity in the early stages of software development, problems of understanding and communication are common in development teams, due to the excessive handling of information from different sources, generating great effort in: consensus, interpretation and transfer of useful information for the application. For this reason, it has been seen the need to merge artificial intelligence with requirements engineering processes that mimic human behavior, through the creation of a vocabulary for the application, improving the Extended Lexicon of Language LEL, using Natural Language Processing, ontologies for semantic reasoning and heuristic rules to improve completeness.

Keywords. Software Engineering, Software Requirements, Artificial Intelligence Techniques, Natural Language Processing.

1 Introducción

Los problemas comunes en el equipo de desarrollo, es la falta de comunicación y comprensión del dominio del Discurso. Integrar y procesar excesiva información de diferentes fuentes puede generar errores de interpretación del problema.

El conocimiento humano es una de las fuentes principales para obtener información preliminar en el conjunto de requerimientos iniciales. Elicitar el conocimiento humano es una tarea que implica de mucho esfuerzo en equipos de desarrollo de software para entender el dominio de la aplicación.

Dominios como por ejemplo la agricultura, requiere en gran medida del conocimiento experto, siendo varios factores que influyen en la diversidad del conocimiento para el desarrollo de aplicaciones como: conocimiento humano, conocimiento ambiental, conocimiento tecnológico y de infraestructura, convertidos en sistemas con terminología compleja de entender y consensuar entre los diferentes stakeholders.

El descubrimiento y conocimiento del dominio generalmente están dados en fases tempranas del desarrollo de software, por lo que la integración de aplicaciones no convencionales provenientes de la inteligencia artificial que imiten un comportamiento similar al de la inteligencia humana, podría beneficiar al equipo de desarrollo en la resolución de problemas comunes en la ingeniería de requisitos.

De esta manera se pretende generar un Framework para la construcción de un vocabulario de la aplicación utilizando técnicas de inteligencia artificial sobre dominios específicos que permitan acortar brechas entre el conocimiento humano y el conocimiento de la aplicación

2 Estado del Arte

La creación de vocabularios para la aplicación es relativamente nueva desde el punto de vista semántico sobre conceptos de un determinado dominio. Algunas investigaciones han dado sus aportes sobre el uso del PLN para extraer conceptos relevantes de especificaciones de requerimientos para ayudar a generar modelos UML como [1] sin establecer semántica. Estudios recientes, sobre la generación de vocabularios están en [2], cuyo aporte se basa en la extracción de entidades, frases nominales y verbales utilizando el Lenguaje natural, para la transformación de modelos ya existentes, genera algoritmos basados en patrones y cuya evaluación de precisión es a través de métricas que no supera al 0,69 %.

3 Declaración del Problema y Contribuciones

Estudios anteriores extraen conceptos relevantes utilizando PLN, el análisis semántico se centra en la identificación de sinónimos, homónimos, antónimos, hipónimos, entre otros, sin establecer una semántica en un contexto determinado, que permitan aproximar al impacto y connotación de la aplicación.

En este sentido la contribución pretende resolver dicho problema dando mejoras sobre:

- La generación de patrones para el análisis complejo de las estructuras gramaticales, extraídas del lenguaje natural de los *stakeholders*, daría lugar a la transformación de conceptos del dominio del problema a conceptos de la aplicación mediante la utilización de PLN. Dichos patrones permitirán:

—Analizar la completitud sobre conceptos como sujetos y objetos constituidos en el LEL

—Analizar acciones del discurso del dominio a través de la categorización de verbos, permitiendo ampliar símbolos del LEL.

- Determinar la semántica del vocabulario utilizando ontologías sobre dominios específicos y a través de relaciones aproximar a escenarios, generando connotación sobre el impacto de la aplicación.
- Generar Heurísticas para determinar la completitud del vocabulario.

4 Metodología y enfoque

La metodología de la investigación, para el desarrollo del Framework de construcción del vocabulario en sistemas inteligentes con razonamiento semántico, se basa en la aplicación del método científico, mismo que permitirá obtener los resultados esperados a través de la deducción e inducción sobre la temática, ya que revisando estados del arte, se seleccionó los más relevantes, para disponer de los insumos necesarios a fin de procesar el lenguaje natural, generando información para el análisis lingüístico de patrones de estructuras construyendo de este modo el vocabulario extendido. En la elaboración de los procesos y/o protocolos es indispensable la observación directa previa de comportamiento del proceso manual y automatizado del uso del lenguaje, para que antes de concretar su formalización se establezca las métricas de evaluación que finalmente logren concretar el entendimiento del dominio y entendimiento del lenguaje común del Steakholder.

5 Plan de Evaluación

El presente informe de simposio doctoral, es de etapa temprana; el cumplimento de la hipótesis se validará en función de la calidad de resultados obtenidos en cumplimiento de los patrones establecidos que logren las mejores estructuras gramaticales, a través del uso de ontologías de dominio extraer el vocabulario adecuado, para que sea validado a través de métricas y revisión de expertos.

6 Conclusiones

La contribución pretende a través de un Framework la construcción de vocabularios para la aplicación sobre dominios específicos, mediante la incorporación de inteligencia y semántica en las etapas iniciales de la ingeniería de requisitos.

Estudios preliminares no han demostrado aún una precisión en la extracción de conceptos del dominio ni han establecido la semántica por completo para generar vocabularios, lo que nuestra propuesta tiene pertinencia para dar solución a estos problemas.

References

1. A. A. Almazroi, L. Abualigah, M. A. Alqarni, E. H. Houssein, A. Q. M. AlHamad, y M. A. Elaziz, «Class Diagram Generation from Text Requirements: An Application of Natural Language Processing», en *Deep Learning Approaches for Spoken and Natural Language Processing*, Springer, 2021, pp. 55-79.
2. P. Danenas, T. Skersys, y R. Butleris, «Natural language processing-enhanced extraction of SBVR business vocabularies and business rules from UML use case diagrams», *Data & Knowledge Engineering*, vol. 128, p. 101822, 2020.

A process to improve Collaborative Work through shared understanding in problem-solving activities

Vanessa Agredo Delgado[1,2][0000-0003-0870-6895] and Cesar A. Collazos[2][0000-0002-7099-8131]

[1]Corporación Universitaria Comfacauca - Unicomfacauca, Popayán, Colombia
vagredo@unicomfacauca.edu.co
[2]Universidad del Cauca, Popayán, Colombia
ccollazo@unicauca.edu.co

Abstract. Collaborative work is becoming increasingly complex and assertive communication is necessary to solve problems in collaborative activities, where the actors must coordinate so that the group works effectively and efficiently. However, achieving true collaboration is not an easy task, there are many factors that influence its achievement, where many of these have been investigated, but the analysis of critical cognitive factors is very scarce, and more those that allow good communication and consequently good coordination. One of them is shared understanding since when working collaboratively there must be common knowledge and understanding, which works as a joint reference base to work effectively and efficiently. Therefore, this work seeks to define a process to improve collaborative work through the construction, monitoring, and assistance of shared understanding. The process has had several versions and each of them has been validated in different contexts, obtaining different types of results, both in terms of its specification and formalization, as well as in terms of its use related to ease of use, feasibility, and usefulness in the construction of shared understanding, which has allowed improving the aspects that have been identified. Through a final validation in a case study, it was determined that the process does improve collaborative work, however, it is still necessary to have technological support elements and it is necessary to lighten the elements that compose it to make it more agile to use.

Keywords. Collaborative work, Shared understanding, Problem-solving activities, Improved CSCW.

1 Introduction

Working collaboratively is not an easy task [1], wrongly it has been believed that having the technological infrastructure guarantees effective collaboration [2], so it

arises the Computer-Supported Collaborative Work (CSCW) concept, which is a multidisciplinary research field that focuses on tools and techniques to support multiple people to achieve a common goal [3]. But to ensure effective collaboration, some external factors should be further analyzed [4], such as the design of the activities, tasks and steps that compose it [3]. In this sense, Collazos [5] in his research divided Computer-Supported Collaborative Learning into 3 phases (work that served as the basis for this research, where the process elements were adapted and improved). The Pre-Process phase begins with the activity design and specification, in the Process phase, the collaboration activity is executed, and finally, the Post-Process phase, a review is carried out to verify the achievement of the proposed objective. Improvements have been made to collaboration in the context of learning [5], [6], [7], and in different aspects of collaborative work [8], [9], [10]. with particular attention paid to the processes and tools provided to aid communication and interaction; but the critical cognitive aspects that ensure that the team works collaborative- ly effectively and efficiently are often absent [11]. One of these cognitive processes is shared understanding, which refers to when group members share a perspective (mutual agreement) or can act in a coordinated manner [12].

Considering the above, it was proposed in this research to define a process (In two levels: the conceptual level that defines the how and a technological level that provides technological support to achieve it) to improve CSCW through the construction, monitoring, and assistance of shared understanding in problem-solving activities. The main objective is to improve collaborative work, making use of the shared understanding benefits in a specific context of problem-solving activities.

2 Problem statement

One of the main collaborative work problems is that collaboration success is hard to achieve [13]. At the same time, collaboration does not occur as easily as one may expect [4], and it is difficult for all the members of a group to participate effectively in the development of the idea with all the other members, and even more so with people who are geographically distant [14]. In this sense, collaboration is defined as "… a coordinated, synchronized activity that is the result of a continued attempt to construct and maintain a shared conception of a problem" [15]. It follows from this definition that, for collaboration to occur, there must be a shared understanding of the problem being solved, thus, being an important determinant of the performance of collaborative groups [16]. Therefore, seeking to improve collaborative work, may consist of finding the application of techniques that support the creation of a shared understanding in heterogeneous groups, it is expected that these groups gain efficiency in their work and produce better group results [17].

Considering the literature, there are some problems related to shared understanding, lit- tle attention has been paid to the systematic development of the processes that lead to shared understanding [18]. and the specific patterns that lead to its construction are not known [19]. Therefore, the practitioners need guidance on how to evoke the processes deliberately and repeatedly [19].

Considering this, the research question was: *How to construct, monitor, and assist the Shared Understanding for improving the CSCW in problem-solving activities?*

3 Research Methodology and Approach

To achieve the objectives of the project, the scientific method described by Bunge [20] was selected as the research framework, which was executed in an iterative and incremental manner. Also, this framework was adapted, and the three cycles defined by the multi-cycle action research methodology with bifurcation [21] were used. The first cycle refers to the Conceptual Cycle, where the research topic is identified, the analysis of the relevant literature is performed, a plan and design of the research project is made, and as a mile-stone, the problem statement is obtained. The second cycle refers to the Methodological Cycle, where the steps for the definition of the process are executed and what was planned in the previous cycle is implemented; it is here where the research disciplines are executed, the main activities proposed by Bunge [20], and the defined and validated process is obtained as a milestone. Finally, the third cycle refers to the Evaluation Cycle, where the research is supervised, and the validated hypothesis is obtained as a milestone.

4 Evaluation Plan

To validate the process, five iterations were carried out, where each version of the process was validated, considering the results obtained in the previous iterations:

- First iteration: The existing processes and elements of a collaborative learning activity were analyzed. Subsequently, with the review of the literature and the identified opportunities for improvement, the elements of collaborative work and those that allow the construction of a shared understanding were analyzed. With this, the first version of the process was defined, which contained 2 phases, the Pre-Process phase, and the Process where shared understanding is built [22]. This version was validated through an exper iment with a group that used the process and a control group that did not, validating its feasibility and usefulness [23]. In addition, an exploratory study was conducted to validate if promotes and improves shared understanding [24].
- Second iteration: Version 2 of the process was defined, which was validated by experts in software and process engineering, who validated the syntax and semantics of the process, in such a way that some errors were identified in the process specification made in SPEM 2.0 [25] were identified and a validation was also performed with AVISPA-Method [26] to make a visual analysis of the process model [27].
- Third iteration: A third version of the process was created, which was called THUNDERS (CollaboraTive work through shared UNDErstanding in pRoblems-solving activities), was applied in an academic context to validate whether THUNDERS promotes and improves shared understanding in a problem-solving activity. This version was also applied in a requirement engineering context,

validating its completeness and usefulness [28].

- Fourth iteration: Version 4 was generated and was subjected to validation by experts in collaboration issues, in order to select the tasks that are or are not mandatory in the execution of the process, with the objective of lightening and simplifying it, and allowing to obtain new processes for specific contexts, being as extensive or light as required, depending on the characteristics of such contexts.
- Fifth iteration: Corrections and updates were made, thus generating version 5, which was validated in a case study, with a group that used THUNDERS and a control group that did not use it, analyzing that its application does improve collaborative work.

5 Conclusions

As a result of this research, the characterization and materialization of the process were obtained, in which different elements of collaborative work are conceptualized, related, collected, and proposed, such as: instruments, strategies, measurement mechanisms, tools involved, and necessary in the construction of shared understanding in problem-solving activities, with formation of heterogeneous groups, from the moment a collaborative activity is designed, executed and the fulfillment of the solution of the problem and the proposed objectives is validated. Proposing a formal and enriched process with activities, tasks, steps, roles, flows, and work products (inputs, outputs, assistance documents), which when applied supports the improvement of collaborative work through the construction, monitoring, and assistance of shared understanding, obtaining better results and achieving the objective of the activity.

In the process of building THUNDERS, several iterations were carried out that allowed the construction in an iterative and incremental way, achieving that the process in the versions that were built was feasible, useful, complete, promoted, and improved shared understanding, and finally, it was obtained that the complete process improved collaborative work, applying it in the context of education and software development. This provided the community with empirical evidence on the construction of shared understanding, its measurement, and the strategies to achieve it, in collaborative problem-solving activities.

6 Ph.D. Stage

Late Stage

References

1.D. Johnson, R. Johnson and K. Smith, "Cooperative Learning: Improving University instruction by basing practice on validated theory,"
Journal on Excellence in University Teaching, vol. 25, no. 4, pp. 1-26, 2013.
2.P. Dillenbourg, "What do you mean by collaborative learning?," in

Collaborativelearning: Cognitive and Computational Approaches., P. Dillenbourg, Ed., Oxford, Elsevier, 1999, pp. 1-19.
3. J. Grudin, "Computer-supported cooperative work: History and focus," *Computer,* vol. 27, no. 5, pp. 19-26, 1994.
4. N. Rummel and H. Spada, "Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer- mediated settings," *The Journal of the Learning Sciences,* vol. 14, no. 2, pp. 201-241, 2005.
5. C. A. Collazos, "Diseño de actividades de aprendizaje colaborativo asistido por computador," *Revista educación en ingeniería,* vol. 9, no. 17, pp. 143-149, 2014.
6. C. A. Collazos, J. Muñoz Arteaga and Y. Hernández, Aprendizaje colaborativo apoyado por computador, LATIn Project, 2014.
7. V. Agredo Delgado, C. A. Collazos and P. Paderewski, "Descripción formal de mecanismos para evaluar, monitorear y mejorar el proceso de aprendizaje colaborativo en su etapa de Proceso," Popayán, 2016.
8. P. B. Lowry, C. C. Albrech, J. D. Lee and J. F. Nunamaker, " Users' experiences in collaborative writing using Collaboratus, an Internet-based collaborative work," *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference,* pp. 243-252, 2002.
9. P. B. Lowry and J. F. Nunamaker, "Using Internet-based, distributed collaborative writing tools to improve coordination and group awareness in writing teams," *EEE Transactions on Professional Communication,* vol. 46, no. 4, pp. 277-297, 2003.
10. A. S. Garcia, J. P. Molina, D. Martınez and P. Gonzalez, "Enhancing collaborative manipulation through the use of feedback and awareness in CVEs," *Proceedings of the 7th ACM SIGGRAPH international Conference on Virtual-Reality Continuum and Its Applications in industry,* 2008.
11. J. F. DeFranco, C. J. Neill and R. B. Clariana, "A cognitive collaborative model to improve performance in engineering teams—A study of team outcomes and mental model sharing," *Systems Engineering,* vol. 14, no. 3, pp. 267-278, 2011.
12. P. Van den Bossche, W. Gijselaers, M. Segers, G. Woltjer and P. Kirschner, "Team learning: building shared mental models," *Instructional Science,* vol. 39, no. 3, p. 283–301, 2011.
13. J. Grudin, "Why CSCW applications fail: problems in the design and evaluationof organizational interfaces," *Proceedings of the 1988 ACM conference on Computer-supported cooperative work,* pp. 85-93, 1988.
14. D. Persico, F. Pozzi and L. Sarti, "Design patterns for monitoring and evaluating CSCL processes.," *Computers in Human Behavior,* vol. 25, no. 5, pp. 1020-1027, 2009.
15. J. Roschelle and S. D. Teasley, "The Construction of Shared Knowledge in Collaborative Problem Solving," in *Computer supported collaborative learning* , Berlin, Heidelberg, Springer, 1995, pp. 69-97.
16. J. Langan-Fox, J. Anglim and J. R. Wilson, "Mental models, team mental models, and performance: Process, development, and future directions," *Human Factors and Ergonomics in Manufacturing & Service Industries,* vol. 14, no. 4, pp. 331-352, 2004.
17. S. Mohammed, L. Ferzandi and K. Hamilton, "Metaphor no more: A 15-year review of the team mental model construct.," *Journal of Management,* vol. 36, no. 4, pp. 876-910, 2010.
18. E. A. Christiane Bittner and J. M. Leimeister, "Why Shared Understanding Matters--Engineering a Collaboration Process for Shared Understanding to Improve Collaboration Effectiveness in Heterogeneous Teams," *System Sciences (HICSS),* vol. 46th Hawaii International Conference, pp. 106-114, 2013.
19. E. A. C. Bittner and J. M. Leimeister, "Creating Shared Understanding in Heterogeneous Work Groups: Why It Matters and How to Achieve It," *Journal of Management Information Systems,* vol. 31, no. 1, pp. 111-144, 2014.

20. M. Bunge, "Philosophy of Science," *México: From Problem to Theory. New Brunswick, Transaction. Siglo Veintiuno Editores,* 2002.
21. F. J. Pino, M. Piattini and G. Horta Travassos, "Managing and developing distributed research projects in software engineering by means of action-research," *Revista Facultad de Ingeniería Universidad de Antioquia,* pp. 61-74, 2013.
22. V. Agredo-Delgado, P. H. Ruiz, A. Mon, C. A. Collazos and H. M. Fardoun, "Towards a process definition for the shared understanding construction in Computer-Supported Collaborative Work," in *Human-Computer Interaction*, Springer, 2020, pp. 263-274.
23. V. Agredo-Delgado, P. H. Ruiz, A. Mon, C. A. Collazos, F. Moreira and H. M. Fardoun, "Validating the shared understanding construction in computer-supported collaborative work in a problem-solving activity," in *World Conference on Information Systems and Technologies*, 2020.
24. V. Agredo-Delgado, P. H. Ruiz, C. A. Collazos and F. Moreira, "An exploratory study on the validation of THUNDERS: a process to achieve shared understanding in problem-solving activities," 2019.
25. OMG, "Software & Systems Process Engineering Metamodel (SPEM)," *OMG Std,* vol. 2, pp. 18-71, 2007.
26. M. C. Camacho, J. A. Hurtado and P. Ruiz, "Un método incremental para el análisis visual de modelos de proceso oftware," *Gerencia en Tecnologia Informatica,* vol. 15, no. 43, pp. 79-91, 2016.
27. V. Agredo-Delgado, P. H. Ruiz, C. A. Collazos and A. Mon, "Building shared understanding with THUNDERS," 2020.
28. V. Agredo-Delgado, P. H. Ruiz, L. E. Garzón, J. España and C. A. Collazos, "A strategy for building shared understanding in requirements engineering activities," 2021.

Design decisions of collaborative tools that reduce theimpact of Fake News

Edison Caiza[1], Alejandro Fernández[1] and Diego Torres[1]

[1] Universidad Nacional de La Plata, Calle 50 &, Av. 120, La Plata, Buenos Aires, Argentina

Abstract. The phenomenon of dissemination and consumption of fake news brings with it the permanent risk of harm to people, companies, or entities referenced in the content of the messages or publications. An important aspect to consider in the proliferation of disinformation is the design of collaborative tools in which fake news are disseminated; which can be improved through knowledge of the factors that facilitate or prevent its propagation. Thus, this research work proposes the construction of a reference model of the impact of fake news, which allows obtaining metrics of the damage they can cause and the factors that increase or reduce it, in order to provide a catalog of design recommendations/decisions that reduce said impact.

Keywords. Fake News, Collaborative tools, Impact, Reference Model.

Introducción

Los espacios colaborativos dada su naturaleza permiten la participación de varios individuos que contribuyen a un objetivo: los cuales mantienen interacción entre sí. Bajo este escenario, si los aplicativos que permiten dicha interacción y colaboración no poseen los controles adecuados para moderar la información que se comparte, permanece latente el riesgo de difundir información errónea o falsa, o cualquier contenido asociado al término “fake news”; y, con ello, la materialización de daños a personas, empresas o entidades referenciadas en ella. Ante esta situación, resulta importante comprender la relación existente entre las características de las fake news y los espacios colaborativos en los que estas se difunden, a fin de levantar/elaborar un catálogo de decisiones de diseño que facilite la implementación de aplicativos que reduzcan el impacto de las fake news.

Las medidas adoptadas por los gigantes de las redes sociales y los sistemas de mensajería, como Facebook, Whatsapp y Twitter, han desacelerado la difusión de desinformación; sin embargo, no son suficientes, debido a que cada vez son más los interesados en difundir contenidos erróneos. Como parte de estas medidas, Facebook, por ejemplo, promueve la información creíble y elimina la información errónea de sus plataformas, además del cierre de cuentas falsas y de medios fraudulentos y la restricción de publicidad para entidades maliciosas. WhatsApp, por su parte, ha limitado la facilidad para reenvío de mensajes en su plataforma. [1] Y, Twitter

permite etiquetar un tweet como sospechoso; además de encargarse de la limpieza de bots y del control de tendencia de temas políticos.

Con base expuesto, se entiende que no existen medidas de solución definitivas y que se demanda de la implementación de medidas de solución alternativas, desde otras perspectivas: tales como el diseño de los espacios colaborativos. Es de aquí que nace la motivación de este trabajo de investigación. El primer hito a concretar, como parte de la propuesta, es la definición de un modelo referencial de los daños (impacto) que provocan la difusión y el consumo de desinformación. En lo posterior, y a partir del modelo referencial establecido, se extraerán las métricas visibles que faciliten el cálculo o la representación de la criticidad de la materialización del daño de las fake news. Y, finalmente, a partir de las representaciones de criticidad obtenidas, se trabajará en el análisis de los factores que incrementan o reducen el grado de criticidad y se construirá un catálogo de decisiones de diseño (controles) para la implementación de espacios colaborativos.

Estado del Arte

Desafortunadamente, el encontrarse expuesto a información inexacta puede conducir a ciertos problemas. Los lectores, después de consumir declaraciones inexactas, muestran claros efectos de esos contenidos en sus decisiones. La exposición a este tipo de información genera confusión acerca de lo que es verdadero, dudas sobre entendimientos precisos y la subsiguiente dependencia de falsedades. [2] Son incontables los eventos que se pueden citar como resultado de la materialización del daño provocado por la desinformación. La muerte, la escalada del conflicto, la hostilidad política y el pánico social son algunos de los efectos adversos de la proliferación de las fake news. [3] En trabajos como [4], por ejemplo, se identifican datos alarmantes del consumo de información falsa respecto del "mito popular de que el alcohol altamente concentrado podría desinfectar al cuerpo y matar al coronavirus" (cerca de 800 personas muertas, 5876 hospitalizadas y 60 ciegos luego de beber metanol). Otro ejemplo del daño a la integridad física de las personas es el asesinato o linchamiento de personas inocentes provocado por la falta de educación digital de quienes consumen información inexacta. En el ámbito social, por citar un último ejemplo, las fake news afectan de sobre manera a las decisiones democráticas de los países, influenciando de forma negativa en la opinión de los electores. [5]

Pese a que existen demasiados ejemplos de materialización de los daños provocados por las fake news, una búsqueda preliminar en revistas y artículos indexados, además de sitios especializados en desinformación, dan cuenta de la falta de métricas estandarizadas para evaluación del impacto de las fake news; lo que representa un problema y un reto para la implementación de controles adecuados, en espacios colaborativos, que reduzcan los riesgos asociados al consumo de desinformación. [6][7]

Planteamiento del Problema y Contribuciones

La proliferación de fake news en los espacios colaborativos es motivo de estudio de un sin fin de trabajos de investigación, en distintas disciplinas de la ciencia; debido principalmente a los efectos negativos (daños) que estas producen en las personas, empresas y entidades. La literatura disponible se enfoca principalmente en el estudio de la problemática social y en la búsqueda u optimización de los métodos de detección; sin embargo, es importante abordar también la relación existente entre las características e impacto de las fake news y las decisiones de diseño de los espacios colaborativos en los que estas se difunden, a fin de establecer los controles necesarios para reducir el impacto de su consumo.

Bajo este contexto, se han planteado las siguientes preguntas de investigación:

P1: ¿Se puede determinar con precisión el impacto de las fake news?

P2: ¿Se puede determinar con precisión la relación existente entre la caracterización de las fake news y las decisiones de diseño de espacios colaborativos?

P3: ¿Se puede contribuir en la toma de decisiones para construcción de espacios colaborativos saludables, en los que se reduzca o impida la propagación de fake news?

Para responder a estas preguntas, el trabajo de investigación planteado propone levantar un conjunto de recomendaciones/decisiones de diseño para implementación de espacios colaborativos "seguros": a través de la búsqueda de información relevante en literatura y la construcción y validación de un modelo de referencia del impacto de las fake news.

En vista de que no se han identificado artículos en los que se aborde la lucha contras las fake news desde la perspectiva del diseño de los espacios colaborativos, la contribución de este trabajo de investigación se centrará en la discusión de la estandarización de métricas del impacto de las fake news y de la forma en que estas se encuentran relacionadas con el diseño de las aplicaciones de interacción social. [5]

Metodología y enfoque de investigación

La ejecución de este trabajo de investigación se dividirá en tres fases:

1. ***Construcción de un Modelo de Referencia del impacto de las fake news***, que permita la extracción de métricas que faciliten la explicación de la relación entre el impacto de las fake news y las decisiones de diseño de los

espacios colaborativos en los que estas se difunde. Para implementación del modelo se empleará la Metodología de Investigación de Diseño (DRM) [8]; que, una vez evaluada, se muestra como la más apta para este tipo de representación.

2. ***Compilación de un catálogo de decisiones de diseño,*** que se obtendrá a partir del análisis de la relación existente entre las métricas de impacto de las fake news y el diseño de las aplicaciones colaborativas.

3. ***Evaluación del catálogo de decisiones de diseño,*** para validación de consistencia de las recomendaciones levantadas. Esta fase se describe a detalle en el punto "Plan de evaluación" de este documento.

Plan de evaluación

El foco de evaluación de este trabajo de investigación será el catálogo de decisiones de diseño, para lo cual se combinarán las siguientes estrategias:

1. Las recomendaciones/decisiones de diseño se encontrarán soportadas principalmente en la literatura disponible y a través de la recolección de datos empíricos.

2. Para las recomendaciones/decisiones de diseño que sean factibles, se validará su asertividad a través de experimentación o por medio del juicio de expertos.

3. Las recomendaciones/decisiones de diseño que resulten demasiado complejas de experimentación y no se cuente con el juicio del experto, se validarán a través de la percepción de los usuarios (mediante opinión o encuestas).

Resultados preliminares o intermedios

Como parte del trabajo realizado, se pueden citar los siguientes resultados preliminares:

1. Se ha determinado el Estado del Arte de las Fake News y de los medios en los que se difunden; a través de lo cual se ha conseguido la publicación de un artículo, a modo de revisión sistemática de literatura.

2. Se ha conseguido una primera aproximación al modelo de referencia de impacto de las fake news, a través de la síntesis de literatura y la representación gráfica con DRM.

Conclusiones y Lecciones Aprendidas

El desarrollo y resultados de este trabajo de investigación contribuirán a la lucha en contra de la difusión de contenidos erróneos y de desinformación, desde la perspectiva de evaluación del daño que pueden ocasionar y el enfoque en las recomendaciones para diseño de espacios colaborativos un tanto más seguros para compartir información. El modelo de impacto de fake news que se obtenga como resultado de este trabajo, puede servir como base para la discusión en trabajos futuros en los que se analicen métricas y decisiones de diseño de espacios colaborativos.

Etapa del Doctorado

Doctoral ES

Referencias

1. Nsoesie, O. Elaine, Oladeji, Olubusola (2020). Identifying Patterns to Prevent the Spread of Misinformation during Epidemics, The Harvard Kennedy School (HKS) Misinformation Review, Volume 1, Special Issue on COVID-19 and Misinformation.
2. Rapp, David & Salovich, Nikita. (2018). Can't We Just Disregard Fake News? The Consequences of Exposure to Inaccurate Information. Policy Insights from the Behavioral and Brain Sciences. 5. 232-239. 10.1177/2372732218785193.
3. Apuke, Oberiri & Omar, Bahiyah. (2020). FAKE NEWS PROLIFERATION IN NIGERIA: CONSEQUENCES, MOTIVATIONS, AND PREVENTION THROUGH AWARENESS STRATEGIES. Humanities & Social Sciences Reviews. 8. 318-327. 10.18510/hssr.2020.8236.
4. Islam, M. S., Sarkar, T., Khan, S. H., Mostofa Kamal, A., Hasan, S. M. M., Kabir, A., Yeasmin, D., Islam, M. A., Amin Chowdhury, K. I., Anwar, K. S., Chughtai, A. A., & Seale, H. (2020). COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis, The American Journal of Tropical Medicine and Hygiene, 103(4), 1621-1629.
5. Machado, Caio & Kira, Beatriz & Narayanan, Vidya & Kollanyi, Bence & Howard, Philip. (2019). A Study of Misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections. WWW '19: Companion Proceedings of the 2019 World Wide Web Conference. 1013-1019. 10.1145/3308560.3316738.
6. Xinyi Zhou and Reza Zafarani. 2018. Fake News: A Survey of Research, Detection Methods, and Opportunities. ACM Comput. Surv. 1, 1 (December 2018).
7. S. B. Parikh, V. Patil, R. Makawana and P. K. Atrey, "Towards Impact Scoring of Fake News," 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2019, pp. 529-533, doi: 10.1109/MIPR.2019.00107.
8. Lucienne T.M. Blessing, Amaresh Chakrabarti, DRM, a Design Research Methodology, Springer-Verlag London 2009, XVII, 397 páginas, https://doi.org/10.1007/978-1-84882587-1.

Conversational interfaces and the Web

Gonzalo Ripa
LIFIA, Facultad de Informática, UNLP
gripa@lifia.info.unlp.edu.ar

Abstract. Conversational User Interfaces (CUI), both in voice interaction and chatbot ways, have become an important and popular way of interacting with current applications. These kinds of interfaces are nowadays spreading around the Web for multiple purposes, although the vast majority of Web sites do not provide them yet. There is still a huge gap between what Web applications allow users to do via Graphical User Interfaces (GUI), and what CUI offers. Also, there are unanswered questions about CUI design and its usability impacts on web browsing. Although other approaches for CUIs creation, the proposed approach transfers the responsibility of creating the conversational interfaces to the users of the Web browser, instead of remaining an implementation decision taken by the Web site owners. In this sense, an End-User Development (EUD) environment was designed to allow end users to define conversational interfaces by themself, based on annotations of UI elements from Web Sites. In the case of Chatbots, plugged into the target Web sites, and in the case of Voice Interfaces (VUI), creating applications for smart speakers' devices based on third-party Web sites content. The underlying ideas and the reference architecture could be used to create tools that allow fast prototyping of CUIs, which could be important to conduct exploratory studies about the impact of using them in Web applications.

Keywords. Conversational interfaces, chatbots, voice user interfaces, end-user development, Web contents.

1 Introduction

Conversational interfaces are becoming a relevant way to access content and services that are normally delivered by Web applications and GUIs. Despite the progress in Model Driven Engineering [1] and Multi-Modal User Interfaces [2], most of the Web sites are not developed with conversational technologies in mind. In particular, the use of VUIs may occur in different contexts: many mobile applications are now able to react to voice commands, and also new uses of voice-based conversation arose, for instance, in the area of Web accessibility based on voice commands. Going to chatbots, those that are embedded in the Website UI may be created ad-hoc [1], or by some middleware based on the annotation of the HTML document [3]. Although these are valuable contributions, it would be ideal that conversational interfaces could

be created automatically from the website GUI. Recent papers [4] [2] have coined the idea of “conversational Web browsing” i.e. “a dialog-based, natural language interaction with websites". Although it seems a promising idea we are still far from materializing this concept. The creation of a conversational interface still depends on application developers, and their specification is not trivial and the cost could be high for many Web site owners. Before focusing on an automatic “conversational transcoding” to create conversational interfaces from GUI, it is required to better understand the different impacts of this interaction in the context of conversational Web browsing and particularly attend to the difference in this interaction when it is done in computers and mobile devices, specially in the case of Chatbots, where the role can change drastically.

2 State Of The Art

Despite the existence of valuable uses of voice interaction, like some approaches based on improving Web accessibility [5], or another one related to performing Web actions with the voice [6], more complex conversational interfaces are still missing to access Web contents and services. Beyond receiving a voice/text command and giving a response based on some specific Web content, the conversational interface may propose an interactive way to deal with that response, explore more from there, etc. These kinds of actions can be achieved by offering conversational interfaces between applications and users. Recently, some new research has arisen in this regard. Prior works have defined the concept of conversational Web interaction [4] and conversational Web browsing [7]. The former proposes a conceptual framework to generate Chatbots for conversational interaction with websites through annotation techniques. In the latter, the authors show an implementation of their prior architecture. It is worth mentioning that both articles are based on an architecture that depends on HTML annotations that must be defined by the application developers. In contrast, here is proposed an annotation based approach that allows end users and developers to add annotations to any Web site on the client side without requiring programming skills and without depending on Web sites owner’s annotations.Another interesting work [8] focuses on ways to build natural language interfaces for individual web APIs so that different virtual assistants don’t have to implement their interfaces, they only need to integrate the proposed framework. The nature of previous approaches makes them very different since the approach proposed here involves the design of the conversational interface as a separated model created by users, while for example, the approach proposed by Baez et al. [4] focuses on a conceptual architecture for automatizing the creation of these interfaces. In this regard, the approach proposed allows combining Web content from multiple sources, because it separates Web content access and return from the conversational interface design. This is a remarkable difference from several mentioned approaches that use a complete Web site as the input of an automatic process and implies Web sites’ developers’ intervention.

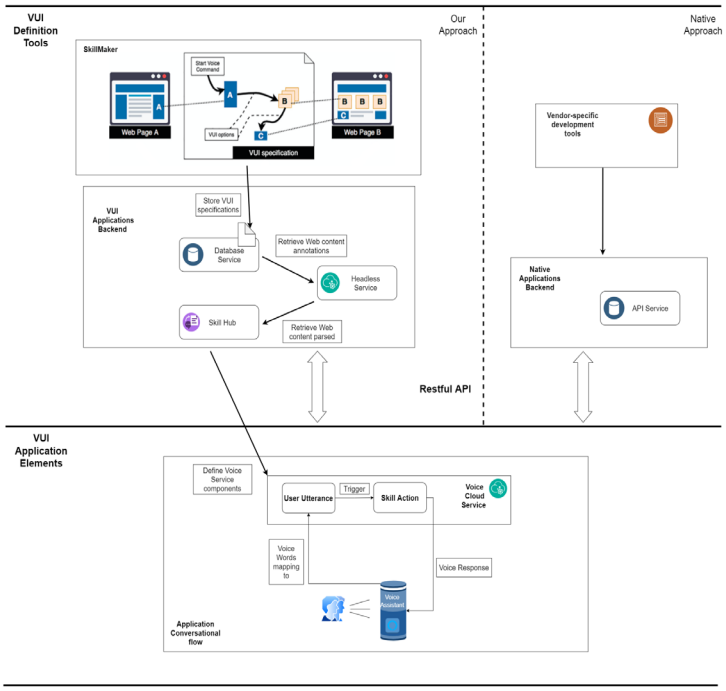
Related to the end-user development aspect of the approach, chatbot development has been tackled from this perspective, particularly for data exploration [9]. However, that work does not support voice user interfaces, and also it does not focus on Web

interaction.

3 Problem Statement And Contributions

Most Web sites and content that exist nowadays can be consumed almost exclusively through a GUI way and were not conceived to be accessed using the voice or through some other conversational interface like could be a Chatbot. What Web applications allow users to perform via Graphical User Interfaces (GUI) is limited, and CUIs could break these limits. The proposed approach, which was implemented as an end-user development environment, allows users to add conversational interfaces to any existing Web site on the client side. In this way, we move the responsibility of creating these interfaces from the Web site's owners to end users. Then, these are automatically weaved on the target Website when it is loaded in the Web browser (in the case of Chatbots) or are automatically available to be consumed in some voice-based device (in the case of VUIs). The main idea is to create conversational interfaces based on existing Web GUIs by defining a set of web element annotations. Imagine users interacting with a chatbot that offers a conversational interaction inside an e-commerce domain: end users could want to get a list of products filtered by some criteria. They must create information models using the EUD tool of this approach identifying and selecting each part of the web elements that conforms to a product for sale, i.e. annotating web contents. Figure 1 shows the overview of the approach from the point of view of VUIs creation but is analogous to chatbot conversations creation too.

Fig. 1: Overview of the idea



The most important is to understand that the EUD tool includes a GUI part (SkillMaker) where end users select and annotate parts of web sites to create VUI specifications. These specifications will be saved into a database service and processed by other back end services (VUI App Backend) which communicates via Restful API with the VUI applications running in some voice-based device: in this case, the VUI applications are created by users, unlike the native approach where these applications are created by owners.

The approach doesn't focus on AI aspects related to the detection of users' requests, but useful interfaces are generated based on simple rule-based conversations, as pointed out in [8]. Although the prototype is focused on the definition of simple conversational interfaces by end users, the underlying ideas and the reference architecture could serve as a basis to create tools for fast prototyping and implementation of these interfaces, which could be important to conduct exploratory studies about the impact of their use in Web applications and also a way to collect data for the creation of automatic transcoding of GUIs into conversational interfaces.

4 Research Methodology

The research is based fundamentally on the analysis of the associated literature (a systematic review was done) and carrying out at least two experiments during the process:

• Experimental test with users for the final product (VUI application) generated by using EUD tools.

• Experimental test with users evaluating the web content annotation process, the conversational interface (Chat bot) definition process carried out using EUD tools, and the utility of the final product (Chatbot) generated for Web and mobile environments.

For the first experiment were collected primary data combining several techniques like surveys, providing a questionnaire of 10 questions multiple-choice based on SUS (System Usability Scale) standard to measure the usability; as well the observation of the subjects performance during experiment execution, to detect difficulties that users can face.

5 Evaluation Plan

Considering that performing experiments can be expensive (we need many users and different domain profiles) but with interesting results, some challenges appear for evaluate the proposed approach with end users:

• Flexibility with which a user can annotate Web site contents.

• The usefulness of conversational interfaces generated as interaction alternatives for a Web site (with the limitations that the conversation model used has).

• The utility of conversational interfaces according to the Website domain. i.e. Are academic sites just as useful as e-commerce (a large database with products) or sites where there is very deep navigation?

• Potential usability improvements for Web sites with

weak designs.

• In the end, it will be evaluated:

– When using a conversational interface to interact with Web sites is interesting.

– What features should have the conversational interface to be useful? I.e. is it valid to show a carousel inside a chatbot to navigate objects that it returns or is it better to show them as a list?

6 Preliminary Or Intermediate Results

The first experiment mentioned in the "Research methodology" section was already carried out. The main goal was to assess whether or not there is a significant user experience difference between applications generated by the approach EUD tool, and native applications (based on Alexa or Google Assistant SDKs). The experiment aims at answering the following research question: *is the user experience equivalent for Native apps and apps generated by the EUD tool?*

For this experiment, it was decided to define and mimic existing apps from the Alexa store using the EUD tool. In this way, we were able to verify whether or not the apps made with the approach are similar to a native one that meets the same requirements. In particular, 2 use cases: weather and news domains. It was designed as a completely randomized experiment where end users (20 Argentinean subjects, 10 females and 10 males, spread between 16-78 yo and with different education levels from different formation fields) with no previous experience with Alexa service, were asked to perform a set of tasks on one of the application versions (native and EUD tool generated). For the analysis of the research question mentioned above, the System Usability Scale (SUS) was used to capture the perceived usability with both approaches, where the resultant means are pretty similar from 87,00 (native) to 88,50 (approach), which denotes not much difference in the SUS score. As a preliminary result of the evaluation carried out, applications generated using the approach provide a similar user experience to native solutions.

7 Conclusions And Learned Lessons

Conversational user interfaces have been a strong research area for a long time. Despite the process in which Web applications owners started to migrate or support this kind of interaction for new devices and interaction contexts, there is still a huge gap between what users can already do with these interfaces and common access to Web content. In fact, many Web applications are still not supporting this new kind of interaction. Beyond this, although this gap could disappear in the near future, other issues such as personalizing or adapting this kind of interface will surface, as

happened in the past with other kinds of interaction. As it was mentioned before, the approach proposed was evaluated by comparing the applications generated with our approach and equivalent native applications. In this way, we showed that the applications obtained with the approach provide a similar user experience to native solutions.

Although so far the tool was not evaluated to provide evidence of the effort required for creating applications, I strongly believe that using the EUD tool is very convenient if we compare the effort required to create native applications, moreover considering the learning curve for both approaches. In this sense, it is planned to study if the EUD tool may also be used for creating another kind of conversational interface such as Web chatbots. This adapting stage is already in development. Besides, it is planned to perform experiments to better understand how end users may use the environment and measure if end users actually are able to create their applications.

8 Ph.D. Stage

Middle Stage

References

1. Marco Brambilla, Jordi Cabot, and Manuel Wimmer. *Model-Driven Software Engineering in Practice*. Synthesis Lectures on Software Engineering. Morgan & Clay pool Publishers, 2012.
2. Nadia Elouali, José Rouillard, Xavier Le Pallec, and Jean Claude Tarby. Multimodal interaction: a survey from model driven engineering and mobile perspectives. *J. Multimodal User Interfaces*, 7(4):351–370, 2013.
3. Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle M. Lottridge. Understanding the long-term use of smart speaker assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3):91:1–91:24, 2018.
4. Marcos Báez, Florian Daniel, and Fabio Casati. Conversational web interaction: Proposal of a dialog-based natural language interaction paradigm for the web. In *Chat bot Research and Design - Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19-20, 2019, Revised Selected Papers*, volume 11970 of *Lecture Notes in Computer Science*, pages 94– 110. Springer, 2019.
5. Yevgen Borodin, Faisal Ahmed, Muhammad Asiful Is lam, Yury Puzis, Valentyn Melnyk, Song Feng, I. V. Ra makrishnan, and Glenn Dausch. Hearsay: a new generation context-driven multi-modal assistive web browser. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, pages 1233–1236. ACM, 2010.
6. Julia Cambre, Alex C. Williams, Afsaneh Razi, Ian Bick ing, Abraham Wallin, Janice Y. Tsai, Chinmay Kulkarni, and Jofish Kaye. Firefox voice: An open and extensible voice assistant built upon the web. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pages 250:1–250:18. ACM, 2021.
7. Pietro Chittò, Marcos Báez, Florian Daniel, and Boualem Benatallah. Automatic generation of chatbots for conversational web browsing. In *Conceptual Modeling - 39th International Conference, ER 2020, Vienna, Austria, November 3-6, 2020, Proceedings*, volume 12400 of *Lecture Notes in Computer Science*, pages 239–249. Springer, 2020.

8. Yu Su, Ahmed Hassan Awadallah, Madian Khabsa, Patrick Pantel, Michael Gamon, and Mark J. Encarnación. Building natural language interfaces to web apis. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pages 177–186. ACM, 2017.
9. Ludovica Piro, Giuseppe Desolda, Maristella Matera, Rosa Lanzilotti, Sara Mosca, and Emanuele Pucci. An interactive paradigm for the end-user development of chatbots for data exploration. In *Human-Computer Interaction - INTERACT 2021 - 18th IFIP TC 13 International Conference, Bari, Italy, August 30 - September 3, 2021, Proceedings, Part IV*, volume 12935 of *Lecture Notes in Computer Science*, pages 177–186. Springer, 2021.

ISBN 978-950-34-2303-5