Advances in Intelligent Systems and Computing 688

Jezreel Mejia
Mirna Muñoz
Álvaro Rocha
Yadira Quiñonez
Jose Calvo-Manzano *Editors*

Trends and Applications in Software Engineering

Proceedings of the 6th International Conference on Software Process Improvement (CIMPS 2017)

Springer

Advances in Intelligent Systems and Computing

Volume 688

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

More information about this series at http://www.springer.com/series/11156

Jezreel Mejia · Mirna Muñoz
Álvaro Rocha · Yadira Quiñonez
Jose Calvo-Manzano
Editors

Trends and Applications in Software Engineering

Proceedings of the 6th International Conference on Software Process Improvement (CIMPS 2017)

Springer

Editors

Jezreel Mejia
Centro de Investigación en Matemáticas A.C.
Unidad Zacatecas
Zacatecas
Mexico

Mirna Muñoz
Centro de Investigación en Matemáticas A.C.
Unidad Zacatecas
Zacatecas
Mexico

Álvaro Rocha
Departamento de Engenharia Informática
University of Coimbra
Coimbra
Portugal

Yadira Quiñonez
Facultad de Informática
Universidad Autónoma de Sinaloa
Mazatlán
Mexico

Jose Calvo-Manzano
Lenguajes y Sistemas Informáticos e Ingeniería de Software
Universidad Politécnica de Madrid
Madrid
Spain

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-69340-8 ISBN 978-3-319-69341-5 (eBook)
https://doi.org/10.1007/978-3-319-69341-5

Library of Congress Control Number: 2017956778

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Introduction

This book contains a selection of papers accepted for presentation and discussion at the 2017 International Conference on Software Process Improvement (CIMPS'17). This Conference had the support of the CIMAT A.C. (Mathematics Research Center/Centro de Investigación en Matemáticas), SEZAC (Secretaria de Economía del Gobierno de Zacatecas, México), AISTI (Iberian Association for Information Systems and Technologies/Associação Ibérica de Sistemas e Tecnologas de Informação), ReCIBE (Revista electrónica de Computación, Informática, Biomédica y Electrónica), DLS Venture Capital, Heltex (Una esperanza de Vida). It will take place at Palace of Conventions of Zacatecas, Zacatecas, México, from October 18 to 20, 2017.

The International Conference on Software Process Improvement (CIMPS) is a global forum for researchers and practitioners that present and discuss the most recent innovations, trends, results, experiences, and concerns in the several perspectives of Software Engineering with clear relationship but not limited to software processes, Security in Information and Communication Technology, and Big Data Field. One of its main aims is to strengthen the drive toward a holistic symbiosis among academy, society, industry, government, and business community promoting the creation of networks by disseminating the results of recent research in order to aligning their needs. CIMPS'17 built on the successes of CIMPS'12, CIMPS'13, CIMPS'14, which took place on Zacatecas, Zac, CIMPS'15 which took place on Mazatlán, Sinaloa, and CIMPS'16 which took place on Aguascalientes, Aguascalientes, México.

The Program Committee of CIMPS'17 was composed of a multi-disciplinary group of experts and those who are intimately concerned with Software Engineering and Information Systems and Technologies. They have had the responsibility for evaluating, in a 'blind review' process, the papers received for each of the main themes proposed for the Conference: Organizational Models, Standards and Methodologies, Knowledge Management, Software Systems, Applications and Tools, Information and Communication Technologies and

Processes in non-software domains (mining, automotive, aerospace, business, health care, manufacturing, etc.) with a demonstrated relationship to Software Engineering challenges.

CIMPS'17 received contributions from several countries around the world. The papers accepted for presentation and discussion at the Conference are published by Springer (this book), and extended versions of best selected papers will be published in relevant journals, including SCI/SSCI and Scopus indexed journals.

We acknowledge all those who contributed to the staging of CIMPS'17 (authors, committees, and sponsors); their involvement and support is very much appreciated.

In loving memory of our international honorary member: Angel Jordan (1930–2017).

October 2017

Jezreel Mejia
Mirna Muñoz
Álvaro Rocha
Yadira Quiñonez
Jose Calvo-Manzano

Organization

Conference

General Chairs

Jezreel Mejia	Mathematics Research Center, Research Unit Zacatecas, Mexico
Mirna Muñoz	Mathematics Research Center, Research Unit Zacatecas, Mexico

The general chairs and co-chair are researchers in computer science at the Research Center in Mathematics, Zacatecas, México. Their research field is Software Engineering, which focuses on process improvement, multi-model environment, project management, acquisition and outsourcing process, solicitation and supplier agreement development, agile methodologies, metrics, validation and verification, and information technology security. They have published several technical papers on acquisition process improvement, project management, TSPi, CMMI, multi-model environment. They have been members of the team that has translated CMMI-DEV v1.2 and v1.3 to Spanish.

General Support

CIMPS General Support represents centers, organizations, or networks. These members collaborate with different European, Latin American, and North American organizations. The following people have been members of the CIMPS Conference since its foundation for the last 6 years.

Cuauhtémoc Lemus Olalde	Head of Cimat Unit Zacatecas, Mexico
Angel Jordan	Software Engineering Institute, USA (1930–2017)

Gonzalo Cuevas Agustin	Politechnical University of Madrid, Spain
Jose A. Calvo-Manzano Villalón	Politechnical University of Madrid, Spain
Tomas San Feliu Gilabert	Politechnical University of Madrid, Spain
Alvaro Rocha	Universidade de Coimbra, Portugal

Local Committee

CIMPS established a Local Committee from the Secretariat of Economic of Zacatecas Government and the Mathematics Research Center, Research Unit Zacatecas, Mexico. The list below comprises the Local Committee members.

Secretariat of Economic

Heriberto Ortega Dominguez, Local Chair, Mexico
Arturo Ramos Hernandez, Local Co-Chair, Mexico
Carlos Fernando Bárcena Pous, Government Support, Mexico
Minerva Villegas Acuña, Public Relations, Mexico

CIMAT Unit Zacatecas

Jessica Garrido Sánchez, Logistics, Mexico
Freddy Iniguez López, Support, Mexico
Ricardo E. Melchor Velásquez, Support, Mexico
Faleg A. Peralta Martínez, Support, Mexico
Isaac Rodriguez Maldonado, Support, Mexico

Scientific Program Committee

CIMPS established an international committee of selected well-known experts in Software Engineering who are willing to be mentioned in the program and to review a set of papers each year. The list below comprises the Scientific Program Committee members.

Adriana Peña Perez-Negrón	University of Guadalajara CUCEI, Mexico
Alejandro Rodríguez Gonzalez	Politechnical University of Madrid, Spain
Alejandra Garcia Hernández	Autonomous University of Zacatecas, Mexico
Alvaro Rocha	Universidade de Coimbra, Portugal
Angel M. Garcia Pedrero	Politechnical University of Madrid, Spain

Antoni Lluis Mesquida Calafat	University of Islas Baleares, Spain
Antonio de Amescua Seco	University Carlos III of Madrid, Spain
Arturo J. Méndez Penín	University of Vigo, Spain
Benjamin Ojeda Magaña	University of Guadalajara, Mexico
Carla Pacheco	Technological University of Mixteca, Oaxaca, Mexico
Carlos Lara Álvarez	CIMAT Unit Zacatecas, Mexico
Daniel Fernandez	CENIDET, Mexico
Edrisi Muñoz Mata	CIMAT Unit Zacatecas, Mexico
Edwin León Cardenal	CIMAT Unit Zacatecas, Mexico
Elisabet Cápon	Swiss Federal Institute of Technology, Zürich (ETHZ), Switzerland
Eleazar Aguirre Anaya	National Politechnical Institute, Mexico
Fernando Moreira	University of Portucalense, Portugal
Gabriel A. Garcia Mireles	University of Sonora, Mexico
Giner Alor Hernandez	Technological University of Orizaba, Mexico
Gloria P. Gasca Hurtado	University of Medellin, Colombia
Graciela Lara Lopez	University of Guadalajara CUCEI, Mexico
Gonzalo Luzardo	Higher Polytechnic School of Litoral, Ecuador
Gustavo Illescas	National University of Central Buenos Aires Province, Argentina
Hector Duran Limón	University of Guadalajara, Mexico
Hugo Arnoldo Mitre	CIMAT Unit Zacatecas, Spain
Hugo O. Alejandrez-Sánchez	National Center for Research and Technological Development, CENIDET, Mexico
Jezreel Mejia Miranda	CIMAT Unit Zacatecas, Mexico
Jose A. Mora Soto	CIMAT Unit Zacatecas, Mexico
Jose Luis Sanchez Cervantes	Technological University of Orizaba, Mexico
Juan Manuel Toloza	National University of Central Buenos Aires Province, Argentina
Lohana Lema Moreta	University of the Holy Spirit, Ecuador
Luis Casillas	University of Guadalajara CUCEI, Mexico
Luis J. Dominguez Pérez	CIMAT Unit Zacatecas, Mexico
Magdalena Arcilla Cobián	National Distance Education University, Spain
Manuel Mora	Autonomous University of Aguascalientes, Mexico
Manuel Pérez Cota	University of Vigo, Spain
Maria de León Sigg	Autonomous University of Zacatecas, Mexico
María del Pilar Salas-Zárate	University of Murcia, Spain
Miguel Hidalgo Reyes	Polytechnic University of Morelos, Mexico
Mirna Muñoz Mata	CIMAT Unit Zacatecas, Mexico
Omar S. Gómez	Higher Polytechnic School of Chimborazo, Ecuador
Perla Velasco-Elizondo	Autonomous University of Zacatecas, Mexico

Ramiro Goncalves	University Tras-os Montes, Portugal
Raúl Aguilar Vera	Autonomous University of Yucatán, Mexico
Ricardo Colomo Palacios	Østfold University College, Norway
Lisbeth Rodriguez Mazahua	Technological University of Orizaba, Mexico
Rafael Valencia-Garcia	University of Murcia, Spain
Rory O'Connor	Dublin City University, Ireland
Santiago Matalonga	University of the West, Scotland
Sodel Vázquez Reyes	Autonomous University of Zacatecas, Mexico
Tomas San Feliu Gilabert	Politechnical University of Madrid, Spain
Ulises Juárez Martínez	Technological University of Orizaba, Mexico
Vianca Vega	Catholic University of North Chile, Chile
Victor Flores	Catholic University of the North, Chile
Victor Saquicela	University of Cuenca, Ecuador
Viviana Y. Rosales Morales	Technological University of Orizaba, Mexico
Yadira Quiñonez	Autonomous University of Sinaloa, Mexico
Yilmaz Murat	Çankaya University, Turkey

Contents

Organizational Models, Standards and Methodologies

Software Systems, Applications and Tools

Information and Communication Technologies

Organizational Models, Standards and Methodologies

ISO/IEC 29110 and curricula programs related to Computer Science and Informatics in Mexico: Analysis of practices coverage

Mirna Muñoz[1], Adriana Peña Pérez Negrón[2], Jezreel Mejia[1], Graciela Lara Lopez[2]

[1]Centro de Investigación en Matemáticas
Avda. Universidad no. 222, 98068, Zacatecas, México
{mirna.munoz, jmejia}@cimat.mx
[2] Departamento de Ciencias de la Computación, CUCEI de la Universidad de Guadalajara,
Av. Revolución No. 1500, 44430, Guadalajara, Jal., México
adriana.pena@cucei.udg.mx, graciela.lara@red.cucei.udg.mx

Abstract. Matching the software industry requirements with the academy training represents a significant problem that must be addressed. In this context, Computer Science students should acquire at the universities the knowledge required to work under organizations environment especially related to the use of models and standards for improving their quality and productivity. This is a key challenge in very small entities (VSEs) and small and medium enterprises (SMEs), because their limitations in time, budget and human resources. This paper presents an analysis of coverage between ISO/IEC 29110 standard, which is used for software industry to ensure quality in Software Engineering practices in VSE, and four academic curricula programs of higher education related to Computer Science and Informatics. The results show some gaps in the Mexican curricula programs that can be reinforced to provide engineers with the knowledge and abilities regarding the use of software engineering best practices, expected in VSE and SME using ISO/IEC 29110. Besides, the analysis shows why ISO/IEC 29110 is easily implemented in this type of organizations.

Keywords: ISO/IEC 29110, software industry, SMEs, VSEs, computer science and informatics curricula

1. Introduction

Nowadays the growth in the importance of software development provides the opportunity in software industry to produce software products and services in order to satisfy the market needs. In this context, the importance of Software development organizations, especially Small and Medium Enterprises (SMEs) and Very Small Entities (VSE), have grown and strengthened becoming a key element in the consolidation of the software industry [1,2].

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_1

According to the National Survey on Productivity and Competitiveness of Micro, Small and Medium Enterprises (SME and VSE) represent 98% of the total of software development industries in Mexico [3]. This fact highlights the importance of guaranteeing the quality of products in this type of organizations.

In order to help SME and VSE in the implementation of best practices to improve product and/or service quality and process performance, standards such as ISO/IEC 29110 engineering and management guides were developed. In México, this standard has been adopted as one of the Quality Standards that have the recognition of the government and the industry.

Therefore, providing qualified professionals able to work under quality models and standards represents a big challenge for universities. Highlighting the importance of this fact, researchers are working on providing resources focused on helping the understanding of quality models and standards such as the learning tool for understanding the project management process of ISO/IEC 29110 basic profile [4,5].

In this paper are shown the results of an analysis performed between the ISO/IEC 29110 basic profile practices and the knowledge provided by the academy taking four curricula programs of Mexican universities in Computer Science and Informatics. This analysis identified if the universities provide an adequate knowledge to Computer Science students to be integrated in organizations that work under quality standards.

The rest of the paper is structured as follows: after de Introduction, section two shows the background of this research work; section three presents an overview of ISO/IEC 29110; section four shows the obtained results; and finally, section five presents conclusions.

2. Background

Diminishing the gap among industry requirements is not a new challenge, but is one of the most critical tasks to be addressed in software education [6,7,8,9]. This situation highlights the importance of performing analysis that present the gaps that universities might have to meet the requirements of the software industry, regarding the quality models and standards. As reported in [6], one way is the analysis of the coverage between curricula programs of universities and models or standards used in software industry.

As background of this research, the authors have performed a similar analysis between curricula programs related of Computer Science and Informatics of Mexican universities and the Moprosoft Model in [10]. Moreover, the study was improved in [11] and [12], showing a depth study among universities curricula programs and the requested knowledge regarding the project management best practices.

Due to the increasing number of Mexican organizations getting the ISO/IEC 29110 certifications, 33 of 43 VSEs certified in ISO/IEC 29110 are Mexican [13]. There are also national programs focused on supporting the organization for the certification of ISO/IEC 29110. Therefore, the authors of this paper decided to perform the analysis between the ISO/IEC 29110 standard and the curricula programs related of Computer Science and Informatics of Mexican universities.

3. ISO/IEC 29110

The ISO/IEC 29110 series of standards and guides was specifically developed for developing software VSEs [2], but not critical software. A VSE could be an enterprise, organization and projects that have up to 25 people. Therefore, the standard goal is to give a solution related to the poor standard adoption in this type of organizations [7], and covering the specific need of very small organizations regarding risk levels, business models and situational factor such as criticality or uncertainty environments [7,9].

The standard is organized by profiles: entry, basic, intermediate and advance [14]. This paper is focused on the basic profile that targets VSEs developing a single product by a single team.

The basic profile is composed by two processes: the project management process and the implementation process, briefly described in Table 1.

Table 1. ISO/IEC 29110 processes

Process	Purpose	Activities
Project management	Establishes and carries out the tasks related to a project implementation in a systematic way, so that the project's objectives are complying with the expected quality, time and costs [15].	Project Planning Project Plan execution Project Assessment and Control Project Closure
Software Implementation	Performances in a systematic way the activities related to the analysis, design, construction, integration and test according to the requirements specified of new or modified software products [15].	Initiation Analysis Design Construction Integration and tests Delivery

4. Methodology for the analysis of curricula programs

For the performance of the analysis of coverage of the four curricula programs of Mexican universities, the methodology used in previous studies [10, 11, 12] was taken as based and adapted.

The tailored methodology is composed of 4 phases: (a) analyze task description provided by the ISO/IEC 29110 standard in its basic profile identifying the required knowledge; (b) identifying the program goal and subjects; (c) analyze each subject goal and units content in order to identify the provided knowledge and; (d) compare each task provided by ISO/IEC 29110 and its required knowledge with the knowledge provided by each subject.

As in previous studies, the scale of values was established in the range of 0 to 4 as follows:

- *0*: the knowledge provided through the subject does not have knowledge related to the ISO/IEC 29110 practice. It means the practice has no coverage.
- *1*: the knowledge provided through the subject is minimal and indirectly related to the ISO/IEC 29110 practice. It means that the practice has a low level of coverage.
- *2*: The knowledge provided through the subject is generic and useful to perform the ISO/IEC 29110 practice. It means that the practice has a medium level of coverage.
- *3*: The knowledge provided through the subject directly supports the performance of the ISO/IEC 29110 practice. It means that the practice has a high level of coverage.
- *4*: The knowledge provided through the subject is specific and directly related to the requirement to perform the ISO/IEC 29110 practice. It means that the practice has a complete coverage level.

It is important to mention that in this study a practice refers to a task in the standard ISO/IEC 29110.

The methodology execution was performed as follows: the four authors execute the steps of the methodology individually. Then, a crosschecking with the obtained results was performed, and finally, a set of meetings was performed to get agreements when values were different. Next section shows the obtained results.

5. Results

It is important to mention that this study was focused on both, the project management process and the implementation process. The analyses performed are focused on two aspects:

- *Analysis of curricula program coverage by activity*: to establish the coverage level by activity. For this, two steps were performed; first, individual value of coverage level by practice was set. Then was applied the next coverage value formula:

  Coverage percentage value by activity = $\boldsymbol{\Sigma}$*i...n practices coverage value / (number of practices * maximum coverage)……………………………………(1)*

  Where the maximum level of coverage or complete coverage is 4.

- *Analysis comparing the four curricula program coverage by ISO/IEC 29110 process*. Establishing the coverage level by process. For this, the next formula was applied:

  Coverage percentage by process = $\boldsymbol{\Sigma}$*i...n Coverage percentage value by activity / number of activities)……………………………………………………….(2)*

Next sections present the obtained results.

5.1 Analysis of curricula programs coverage by activities

For the performance of the analysis presented in this section, the formula (1) was applied.

5.1.1 Informatics Engineering curricula program

The Informatics curricula program aims to create and maintain creative and innovative solutions regarding the information systems. The results, presented in Table 1, show that graduates of this program are better qualified to carry out the practices involved in software implementation such as: establish an implementation environment, define and analyze requirements, develop software architectural and detailed design, develop software components and integrate them, define and perform unit test, manage changes, perform validation and verification, document user, operation and maintenance manuals and deliver the product. Besides, activities of project management such as: monitoring the project progress and delivering the product have a high coverage.However,these graduates needs to improve their training in activities related to the development of project plan such as: define scope, estimate size, time and resources of tasks necessary to perform the project, identify risks, configuration management, manage change request, perform meetings, and develop a control strategy.

Table 1. Coverage of activities by Informatics engineering curricula program

Activities	% Coverage
Project Planning	71
Project Plan execution	75
Project Assessment and Control	88
Project Closure	88
Initiation	93
Analysis & Design	94
Construction	96
Integration and tests	93
Delivery	96

5.1.2 Software engineering curricula program

The software engineering curricula program aims to train professionals in process development and the evolution of large and small scale software systems that solve problems in different areas, using appropriate tools to optimize time and cost.

The results, presented in Table 2, show that graduates of this program are well qualified to carry out all practices involved to manage, assess and control a project such as: define scope, estimate size, time and resources of tasks necessary to perform the project, identify risks, configuration management, manage change request, perform meetings, develop a control strategy; monitor the project progress and deliver the product; as well as in the practices related to the software construction such as: establish an implementation environment, define and analyze requirements, develop software architectural and detailed design, develop software components and integrate them, define and perform unit test, manage changes, perform validation and verification, document user, operation and maintenance manuals, and deliver the product.

Table 2. Coverage of activities by the software engineering curricula program

Activities	% Coverage
Project Planning	98
Project Plan execution	100
Project Assessment and Control	100
Project Closure	100
Initiation	88
Analysis & Design	98
Construction	100
Integration and tests	100
Delivery	100

5.1.3 Computer Science curricula program

The degree in Computer Science curricula program aims to train professionals with analytical skills, critical skills, creativity and leadership to provide computational solutions in organizations applying information technology and communications.

Table 3 shows the results obtained of analyzing the degree in Computer Science curricula program. The graduates of this program are better qualified to carry out the practices of software implementation such as: establish an implementation environment, develop software components and integrate them, define and perform unit test, manage changes, perform validation and verification, and document user, operation and maintenance manuals and deliver the product.However,these graduates needs to improve their training in activities related to project management such as: define scope, estimate and size, time and resources of tasks necessary to perform the project, identify risks, configuration management, manage change request, perform meetings, and develop a control strategy. As well as in practices related to software implementation such as: define and analyze requirements, software analysis and design, develop software architectural, and detailed design.

Table 3. Coverage of activities by Computer Science curricula program

Activities	% Coverage
Project Planning	58
Project Plan execution	46
Project Assessment and Control	33
Project Closure	50
Initiation	38
Analysis & Design	38
Construction	85
Integration and tests	84
Delivery	89

5.1.4 Computer engineering curricula program

The Computer engineering curricula program aims to train professionals with analytic capacities, critical to provide creative solutions to the regional and state development using computer technology, and promoting socials values as well as the environmental care.

Table 4 shows the results obtained of analyzing the degree in computer engineering curricula program. The graduates of this program are better qualified to carry out the practices related to project management such as: define scope, estimate size, time and resources of tasks necessary to perform the project, identify risks, configuration management, manage change request, perform meetings, develop a control strategy as well as in activities related to define and analyze requirements, develop software architectural and detailed design, develop software components and integrate them, define and perform unit test, manage changes, perform validation and verification. However, these graduates needs to improve their training in activities related to project management such as: monitor the project progress; as well as in practices related to software implementation such as: establish an implementation environment and document user, operation and maintenance manuals, and deliver the product.

Table 4. Coverage of activities by Computer engineering curricula program

Activities	% Coverage
Project Planning	87
Project Plan execution	79
Project Assessment and Control	58
Project Closure	88
Initiation	63
Analysis & Design	83
Construction	89
Integration and tests	70
Delivery	63

5.2 Analysis of the four curricula programs coverage level by process

This analysis establishes the coverage level of curricula programs regarding the two process of the ISO/IEC 29110 basic profile. To perform this analysis the formula (2) was applied.

Table 5 shows the results of comparing the four curricula programs as follows: (column a) Informatics engineering curricula program; (column b) Software engineering curricula program; (column c) Computer Science curricula program; and (column d) Computer engineering curricula program.

Table 5. Summary ISO/IEC 29110 processes coverage by curricula programs

Process	(a)	(b)	(c)	(d)
Project management	82%	100%	47%	78%
Software Implementation	93%	97%	62%	75%

The findings of the analysis performed are summarized as follows:

- *Project management process*: the purpose of this process is establishing and carrying out the tasks of the software implementation project in a systematic way, allowing complying the project's objectives with the expected quality, time and cost [16].
  Interpretation of process coverage: only the Software engineering curricula program (column b) fully covered the project management process, this means that these graduates are qualified to carry out this process. Graduates of the Informatics engineering curricula program (column a) and the Computer engineering curricula program (column d) have high level of knowledge to carry out the practices involved in project management. However, graduates of the Computer Science curricula program (column c) receive minimum training to carry out the practices involved in project management process.
- *Software Implementation process*: the purpose of this process is performing the analysis, design, construction, integration and tests activities in a systematic way or new or modified software products according to the specified requirements [16].
  Interpretation of process coverage: not any area is fully covered by the curricula program. However, it was found that graduates of the Informatics engineering curricula program (column a) and the Software engineering curricula program (column b) are qualified to carry out the practices involved in the software implementation process. The graduates of the Computer Science curricula program (column d) get a high level of knowledge to carry out the practices involved in software implementation process. Finally, graduates of the Computer Science curricula program (column c) need to improve their training to carry out the practices involved in software implementation process.

The results of the analysis show that from the undergraduates of curricula programs of Mexican Universities in Computer Science and Informatics, three curricula programs: Informatics engineering (column a); Software engineering (column b) and Computer engineering (column d) have a high coverage of the ISO/IEC 29110 practices. This allows undergraduates to be integrated in VSE with the knowledge required to develop practices established in VSE using ISO 29110; the reason to understand why this standard is well accepted and easily implemented in Mexican SMEs and VSEs.

6. Conclusions

This paper presents the results of analyzing the coverage of ISO/IEC 29110 basic profile practices from the knowledge provided by four curricula programs of universities of México related to Computer Science and Informatics. The results show that most of the graduates acquired a high level of fundamentals and knowledge to be highly qualified regarding the ISO/IEC 29110 practices.

Same as in the papers with similar analysis performed to the Mexican model Moprosoft, the results of this paper does not pretend to determine whether or not graduates are able to perform a practice, but to analyze how helpful is the knowledge provided to the graduates during their training in the university.

Therefore, it can be highlighted that Software engineering curricula programs have a better coverage of practices involved in both processes, followed by Informatics engineering and Computer engineering curricula programs. Finally, the Computer Science curricula program with the lowest results has better level in the software implement process, but it needs to improve their training in project management process.

Besides, it is important to mention that the basic profile of the ISO/IEC 29110 series provides a standard that is well accepted in Mexican software industry because it provides basic practices throughout two key processes (project management and software implementation) that can be easily implemented because of the graduates' fundamentals and knowledge that guarantee the quality of products in these type of organizations.

References

1. Ministry of Economy: SMEs: fundamental Link to the growth of Mexico, vol. 2013 (2013)
2. Laporte C., O'Connor R.: Systems and software engineering standards for very small entities: accomplishments and overview. Computer, IEEE Computer society. 49(8) pp 84-87 (2016)
3. INEGI: Encuesta Nacional sobre Productividad y Competitividad de las Micro, Pequeñas y Medianas Empresas (ENAPROCE). (2015) http://www.inegi.org.mx/est/contenidos/proyectos/encuestas/establecimientos/otras/enaproce/default_t.aspx
4. Sanchez-Gordon, M.L., O'Connor, R. V., Colomo-Palacios, R. and Sanchez-Gordon, S.: A Learning Tool for the ISO/IEC 29110 Standard: Understanding the Project Management of Basic Profile. In Proceedings 16th International Conference on Software Process Improvement and Capability dEtermination (SPICE 2016). Springer-Verlag. (2016)
5. Sanchez-Gordon, M.L., O'Connor, R. V., Colomo-Palacios, R. and Herranz, E.: Bridging the Gap between SPI and SMEs in Educational Settings: A Learning Tool Supporting ISO/IEC 29110. 23nd European Conference on Systems, Software and Services Process Improvement (EuroSPI 2016), Springer-Verlag. (2016)
6. Moreno A., Sanchez-Segura M.A., Medina-Dominguez F.: Analysis of Coverage of CMMI practices in software engineering curricula, in SEPG Europe 2012 Conference proceedings, Special Report CMU/SEI-2012-SR-005. pp. 42-77September 2012 (2012)
7. Laporte, C.Y. and O'Connor, R.V.: Software Process Improvement in Graduate Software Engineering Programs, proceedings 1st International Workshop Software Process Education, Training and Professionalism (SPEPT 2015), pp. 18 - 24, CEUR Workshop Proceedings Vol. 1368, Gothenburg, Sweden, 2015
8. Laporte, C.Y., O'Connor, R.V., Garcia Paucar, L. and Gerancon B.: An Innovative Approach in Developing Standard Professionals by Involving Software Engineering Students in Implementing and Improving International Standards, Journal of the Society for Standards Professionals, Vol. 67, No. 2 (2015)
9. Laporte, C.Y. and O'Connor, R.: Software Process Improvement in Industry in a Graduate Software Engineering Curriculum, Software Quality Professional, Vol. 18, No. 3, (2016)

10. Muñoz Mirna, Peña Pérez-Negrón Adriana, Mejía Jezreel and Graciela Lara Lopez: Analysis of Coverage of Moprosoft practices in curricula programs related to computer science and informatics in the 4th International conference of software process improvement CIMPS 2016, pp 35-45. (2016)
11. M. Muñoz, A. Peña, J. Mejia and G. Lara: Coverage of the University Curricula for the Software Engineering Industry in Mexico, IEEE Latin America Transactions, Vol.14, Pag.2383-2389. (2016)
12. Mirna Muñoz, Adriana Peña Pérez Negrón, Jezreel Mejia and Graciela Lara López: Actual State of the Coverage of Mexican Software Industry Requested Knowledge Regarding the Project Management Best Practices, Computer Science and Information Systems (ComSIS), Vol.13, Pag.849-873. (2016)
13. NYCE: Empresas Certificadas en la Norma ISO/IEC 29110 https://www.nyce.org.mx/wp-content/uploads/2016/12/PADRON-DE-EMPRESAS-CERTIFICADAS_ISO29110_14-12-2016.pdf . (2016)
14. Claude Y. Laporte: Challenges and realities of ISO/IEC 29110 series for very small entities, seminary Zacatecas. (2017)
15. ISO/IEC: Software engineering- Lyfecycle profiles for very small entities (VSEs). Part 5-1-2: Manageent and engineering guide: generic profile group: Basic profile. ISO/IEC TR 29110-5-1-2:2011(E). (2011)
16. Claude Y. Laporte: The Generic Profiles for VSEs Developing Software. (2011). http://profs.etsmtl.ca/claporte/english/VSE/

A Means-Ends Design of SCRUM+: an agile-disciplined balanced SCRUM enhanced with the ISO/IEC 29110 Standard

Sergio Galván-Cruz[1] , Manuel Mora[2], Rory O'Connor[3]

[1] Autonomous University of Aguascalientes, Electronic Systems, Av. Universidad, 940, Aguascalientes, Mexico, checogc23@gmail.com
[2] Autonomous University of Aguascalientes, Information Systems, Av. Universidad, 940, Aguascalientes, Mexico, dr.manuel.mora.uaa@gmail.com
[3] Dublin City University, School of Computing, Glasnevin, Dublin 9, Ireland, rory.oconnor@dcu.ie

Abstract. Agile systems development methodologies (ASDMs) have gained high acceptance in very small entities (VSEs) of software development seeking quality at minimal effort. SCRUM and XP in industrial settings and UPEDU in academic ones are main of them. Similarly, Software Process Improvement (SPI) initiatives promote the utilization of process frameworks and standards. However, despite both worlds (i.e. ASDMs and SPI) pursue a shared end of high-quality software, both are separated by different underlying approaches. We consider that ASDMs can get benefits from SPI through controlled enhancements (i.e. an agile-discipline balance) without elimination of agility. Thus, in this research, we report the design of SCRUM+, an enhanced SCRUM with recommendations on roles, activities-tasks and artifacts from the SPI standard ISO/IEC 29110. SCRUM+ was designed by using a Means-Ends analysis. Our final aim is to provide such an enhanced SCRUM methodology via an Electronic Process Guide (EPG) to help practitioners for a better use of agile approaches with SPI added recommendations that be found theoretically robust and potentially useful regarding SCRUM from a panel of experts and SCRUM practitioners.

Keywords: SCRUM+, SCRUM, ISO/IEC 29110, SPI, enhanced methodology

1 Introduction

Software process models and standards (SPMSs) such as CMMI-DEV and the ISO/IEC 12207 [1, 2], have been developed by international associations for helping to software development organizations to meet the current demands for quality process and product improvements [3, 4]. SPMSs are important for software

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_2

development organizations because their correct implementation has generated relevant benefits such as: process cost reduction, critical software failure reduction, quality product increment, team productivity increment, and customer satisfaction increment among others [5, 6]. Such SPMS are core initiatives (i.e. process frameworks besides best practices and tools [7]) into the Software Process Improvement (SPI) approach. According to [7] SPI refers to "*a systematic approach to increase the efficiency and effectiveness of a software development organization and to enhance software products*". However, according to [4] these SPI initiatives "*were not written for small projects, small development organizations, or companies with between 1 and 25 employees, and are consequently difficult to apply in such settings*". Thus, VSEs (whole business or project teams) while represent a high percentage of software business in the world [8], are unserved potential users by such normal software process standards and models [9, 10]. Given this problematic situation, a new ISO/IEC 29110 software process lifecycle standard [11] was elaborated specifically to VSEs.

In this same context of VSEs, it has been also identified the preference for using Agile Software Development Methodologies (ASDMs) such as: SCRUM and XP in industrial settings [12, 13] and UPEDU in academic ones [14]. Furthermore, ASDMs literature claims [12, 13] similar overall benefits achieved for opposite software development methodologies (i.e. rigor-disciplined ones framed on SPI process frameworks) with the additional advantages provided only by the agile approach [15]. Thus, the ASDMs should be successfully used by VSEs. However, while there are evidences of a high rate of utilization of these agile methodologies [12, 13] by VSEs, it has been also reported that there are some contextual prerequisites for a successful utilization [16, 17]. We consider that these contextual prerequisites for successful adoption of ASDMs refers to the adherence to best scholastic practices provided by SPMS as the SPI approach promotes [16, 17]. Thus, the practitioners of ASDMs need to enhance their ASDM with some recommendations from rigor-disciplined development methods instead of using ASDMs directly [18, 19].

In this research, we report the design of SCRUM+: an enhanced SCRUM [20, 21] with some recommendations from the ISO/IEC 29110 standard. This new standard [11] has been released for VSEs but it was designed independently of the ASDM approach. Hence, thus we pursue to present a more robust balanced agility-disciplined SCRUM method aligned to the recommendations from several literatures [18, 19]. In practical perspective, it provides a dual overall benefit: for SCRUM practitioners to count with a more robust disciplined process enhanced with some critical recommendations from the ISO/IEC 29110 standards, and for the ISO/IEC 29110 community to count with a specific adaptation of SCRUM with a greater coverage of the expected practices to be conducted in the ISO/IEC 29110 standard [11, 22] than SCRUM actually covers [20, 21].

The remainder of this paper continues as follows: the research process is reported in the section 2; the theoretical bases on SCRUM, the ISO/IEC 29110 standard, and the Agility-Discipline debate are reported in the section 3; the application of the Means-Ends method for designing SCRUM+ is presented in section 4; finally, limitations, recommendations and conclusions of this research are reported in section 5.

2 Research Process

2.1 Research Goals and Design Restrictions.

The main overall research goal is: to design an enhanced and agile-disciplined balanced SCRUM+ methodology based on the original SCRUM methodology and best practices provided by the ISO/IEC 29110 (Entry Profile, Project Management process) [11, 22]. Two specific and critical design restrictions are: 1) SCRUM+ must be still perceived as an agile method by practitioners (i.e. it means that SCRUM+ must not lose its agile essence); and 2) SCRUM+ must reach at least a high coverage level with the ISO/IEC 29110 standard (Entry Profile, Project Management process).

2.2 Research Methods and Materials.

The Means-Ends analysis technique [23] was initially elaborated in the early Artificial Intelligence research stream in the 50's [24] as a Problem-Solving technique. According to Newell and Simon [25]: "*Problem solving can be viewed, then, as finding one of the few paths that leads from a problem's initial state to its goal state through some space of possible intermediate states*".

In Means-Ends Analysis technique [25] a problem is a situation faced by a person to reach a desired state (named *Goal State*) from an initial departure point (named *Initial State*) and it is not known in advance the set of actions (named *Operators*) and the sequence of application (named *Path*) on objects (named *Operands*) that must be applied. Thus, to find a *Solution* means to find a sequence of *Operators* applied to *Operands* to transform the *Initial State* in the *Goal State*. A *Solution* can be *Optimal* or *Satisfactory*. For many problems, there are not known or practically feasible algorithms (i.e. a predefined set of actions to be followed for transforming an *Initial State* in a *Goal State*) to be applied for reaching to the *Goal State*. In these cases, the concept of Heuristics is applied [24]. Heuristics are recommendations and clues gained through the experience in similar or related problems that are suggested to be applied (i.e. Heuristics on what *Operators* apply on the *Operands* given a current *State* and the expected *Goal State*). This process, according to [23] has two principal features: 1) Reduction of Differences which is a preference of problem solvers to use the *Operators* that produce *States* most similar to the *Goal State*; and 2) Sub-Goaling which happens when a problem can be divided in sub-Problems and thus its final *Solution* can be reached when the *Solution* for all of the Sub-Goals is reached under necessity of all Sub-Goals (i.e. connected by AND logical operator) or at least one *Solution* is reached (i.e. connected by OR logical operator). Hence, thus, we consider that the Means-Ends Analysis technique provides a systematic well-tested method that can be applied for the systematic design of SCRUM+.

In this research, we used the following materials: 1) the ISO/IEC 29110 Entry Profile document [22]; 2) the official guides for SCRUM [20, 21]; 3) a SCRUM book [26]; 4) a SCRUM EPG (Electronic Process Guide) [27]; and 5) a coverage analysis of SCRUM regarding the ISO/IEC 29110 Entry Profile Project Management process [28].

3 Theoretical Background

3.1 SCRUM Methodology.

SCRUM has been reported as the most used agile methodology [13]. According to [29] "*SCRUM has a project management emphasis. SCRUM has been applied mainly to software projects, but a number of non-software projects have also been managed with SCRUM---the principles are applicable to any projec*t". Moreover, according to [21] "*SCRUM is a framework for developing and sustaining complex products*".The model of SCRUM was designed for optimizing the flexibility, creativity and productivity of well-trained teams.

SCRUM can be structured in three Roles, seven Activities (with 24 tasks) and five Artifacts. The three roles (reported in an IDEF0 diagram [30] as Mechanisms) are: 1) "Product Owner", 2) "Scrum Master", and 3) "Development Team". The "Product Owner" is responsible for the product backlog (its content, availability and ordering). The "Scrum Master" can be considered the SCRUM expert and project leader that will interact with the other roles for leading and guiding them toward the end goal. The "Development Team" "consists of professionals who do the work of delivering a potentially releasable Increment of DONE product at the end of each Sprint" [21].

The seven SCRUM Activities are: 1) "Planning Pre-Game", 2) "Systems Architecture Pre-Game", 3) "Sprint Planning Game", 4) "Daily SCRUM Game", 5) "Sprint Increment Development Game", 6) "Sprint Review Game", and 7) "Sprint Retrospective Game". Some literature [31, 26] adds an explicit final activity of 8) "Project Closure Post-Game". In this research, we have focused on the Project Management (PM) activities (i.e. the Activities 1, 3, 4, 6, 7 and 8). The activities 2 and 5 corresponds to Software Implementation (SwI) process.

The five SCRUM Artifacts are: 1) "User Need List", 2) "Product Backlog" (it includes the "User Stories"), 3) "Sprint Backlog" (it includes the "Sprint Burndown Chart"), 4) "Increment", and 5) "Acceptance Criteria". The specific Artifacts related to Project Management Process are 1, 2, 3, and 5. The 1) "User Need List" is the open list of needs and requirements expressed by the Customer. The 2) "Product Backlog" is "*an ordered list of everything that might be needed in the product and is the single source of requirements for any changes to be made to the product*" [21]. An "User Story" is "*a card that describes an increment of value to the customer. The user story is written for the developer in order to express the increment of value*" [26]. The 3) "Sprint Backlog" is "*the set of Product Backlog items selected for the Sprint, plus a plan for delivering the product Increment and realizing the Sprint Goal*" [21]. This Artifact includes the "Sprint Burndown Chart" which is a chart which "*shows the amount of work remaining across time*" and permits to visualize "*the correlation between the amount of work remaining at any point in time and the progress of the project team(s) in reducing this work*" [32]. The 4) "Increment" is *the sum of all the Product Backlog items completed during a Sprint and the value of the increments of* *all previous Sprints*" [21]. The "Acceptance Criteria" is "*essentially a clarification of the story. It gives the developer a set of steps that must be completed before the story can be considered done. The acceptance criteria are created by the product owner with the help of the customer. It sets the expectation of the user story*" [26].

3.2 The ISO/IEC 29110 Standard – Entry Profile

The ISO/IEC 29110 standard (Entry Profile) [9] provides a lightweight process model developed for organizations classified as very small entities (VSEs employs from 1 to 25 people). According to [31] standard emerged for the needs identified in VSEs on: 1) clear and detailed guidance with templates and examples; 2) a lightweight and easy-to-understand standards; and 3) standards with minimum cost, time, and resources for their implementation. This ISO/IEC 29110 standard has three Roles, two Process Categories, and fourteen Artifacts. The three roles are: "Customer", "Project Manager", and "Work Team". The two Process Categories are: Project Management (PM) and Software Implementation (SI).

Project Management aims to establish and carry out the tasks of the software implementation, which will fulfill the objectives of the project according to quality, time and expected costs. PM includes four activities: ***Planning, Control, Execution and Closure***. **Software Implementation** aims to systematically analyze, design, construction, integration and testing of software products processed according to specified requirements. SI includes six activities: ***Initiation, Analysis, Design, Construction, Tests and Delivery.***

The fourteen Artifacts are: 1) Acceptance Record, 2) Change Request, 3) Meeting Record, 4) Progress Status Record, 5) Project Plan, 6) Project Repository, 7) Requirements Specifications, 8) Software, 9) Software Component, 10) Software Configuration, 11) Software Component Identification, 12) Statement of the Work, 13) Test Cases and Test Procedures, and 14) Test Report. The Activities and Artifacts of interest for this research are the corresponding to PM Process Category. These activities are: ***Planning, Control, Execution and Closure***; and these Artifacts are: 1) Acceptance Record, 2) Change Request, 3) Meeting Record, 4) Progress Status Record, 5) Project Plan, 6) Project Repository, 10) Software Configuration, and 12) Statement of the Work.

3.3 The Agility-Discipline Debate

According to several relevant literatures [16, 17], the direct application of ASDMs does not guarantee the proffered benefits of agility Project Management. Furthermore, from a disciplined Project Management approach [18, 19] there had been logical arguments on the need to robust the agile methods with some disciplined-oriented best practices. A summary of recommendations for having a balanced agility-disciplined Project Management approach (called also *ambidextrous* approach [34]) is as follows: 1) risks are not managed explicitly in agile methods; 2) a particular organizational culture is required for agile methods while that disciplined is less contingent to this factor; 3) agile methods can be considered chaotic by excessive flexibility and customization for teams trained in disciplined methods; 4) control and monitoring of project must be still exercised; 5) agile methods are more focused on small teams and small project (that can be large by evolution but not for an initially planned scope as a large project) and thus their scale up suffer from drawbacks; 6) new current software projects are more complex than past ones so both approaches (disciplined and agile one) are required. Thus, a call for elaborating balanced agility-disciplined Project Management methodologies is currently reported in the literature.

4. A Means-Ends Design of SCRUM+

4.1 The SCRUM Initial Status as the Core Input for the Means-Ends Analysis

To design SCRUM+ as a balanced agility-disciplined enhanced SCRUM methodology, we start from the results reported in [28] regarding a thoroughly analysis of the coverage of SCRUM, XP and UPEDU Project Management processes regarding the ISO/IEC 29110 (Entry Profile, Project Management process). According to [35] SCRUM, XP and UPEDU had respectively an overall coverage of moderate (79%), low (51%) and high level (93%) respectively. Hence, while the obvious selection of reporting UPEDU as a ready-to-use balanced agility-disciplined methodology and highly in congruence with the ISO/IEC 29110 Entry Profile concerning to the Project Management process, UPEDU is not a well-known agile methodology in industrial settings and it supports an agile approach based on a simplified rigor and discipline from its derivation from RUP (i.e. a strong disciplined development methodology). In the opposite case, to try to enhance XP which has a low compliance level implies the addition of many missed issues, and thus, the enhanced XP can be perceived theoretically far away of the agile approach by practitioners. Thus, in this research, it has been selected SCRUM that reached a moderate level (79%) and its enhancement toward next level (i.e. high) can produce a less conceptual disruption perception than the change required in XP from low to high coverage level (i.e. a suitable balanced agility-disciplined methodology).

4.2 Application of the Means-Ends Analysis Technique for Designing SCRUM+.

We propose six heuristic strategies to perform systematically transformations from SCRUM to SCRUM+. These six heuristic strategies are considering the design limits reported. The strategies 1 to 3 tried to eliminate only the items (i.e. a Role, an Activity, or an Artifact) that are evaluated in overall as NULL level, and the strategies 4 to 6 tried to eliminate both the overall of NULL and LOW levels regarding their compliance level with the ISO/IEC 29110 Entry Profile. These items (i.e. a Role, an Activity, or an Artifact) are not mentioned or are weakly reported in SCRUM regarding the ISO/IEC 29110 Entry Profile. The items were analyzed on the three components (Roles, Activities and Artifacts).

For example, the Strategy 1 was based on doing soft (minimal) changes from SCRUM to SCRUM+ by moving the NULL (✖✚) and LOW (⊙) status found in individual feature (i.e. cells) of each item (i.e. a specific Role, Activity or Artifact) whose status level is NULL (✖✚). Thus, NULL (✖✚) to LOW (⊙) and from LOW (⊙) to MODERATE (◒) are the changes to be applied. The other two status level of

MODERATE (◒) and HIGH (●) found in the cells were kept in the same status level. These changes were soft (minimal), and it pursued a soft evolution from the SCRUM to SCRUM+ with the minimum change as possible for every specific item in the three components (Roles, Activities and Artifacts). The Strategies 4, 5 and 6 were almost the same of 1, 2 and 3 ones respectively, with the unique difference that the changes applied in strategies 1, 2 and 3 were applied only on the items (i.e. a specific Role, Activity or Artifact) whose status level is NULL (✖+), while that in the strategies 4, 5 and 6 the changes are applied on items whose status level is NULL (✖+) and LOW (⦿).

All these qualitative assessments were finally mapped to a numerical scale from 0 to 3, and their average value multiplied by their corresponding weight assigned to the specific Role, Activity-Task or Artifact. Thus, the final scores can be from 0 to 100 points. The Table 1 reports the final levels reached by each strategy. The value of 100 points for the strategy 6, for instance, implies to add to the original SCRUM all identified missing attributes for Roles, Activities-Tasks and Artifacts from the ISO/IEC 29110 standard, but it naturally will produce a loss of the agile essence of SCRUM. Thus, the selected strategy to produce SCRUM+ faces a trade-off situation between getting an improved methodology and keeping its agility status.

Table 1. Results of the Six Means-Ends Strategies

	Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5	Strategy 6
Roles	78	89	100	78	89	100
Activities	93	97	100	93	97	100
Artifacts	73	80	80	83	90	100
Total	81	89	93	85	92	100

3.3 Solution: Selected Means-Ends Strategy

In order to select the final solution we defined three criteria: 1) the solution must have an overall level that it is between 80 and 89 points (i.e. a high coverage level); 2) the solution must have the minimal overall Euclidean Distance M3, which measures the distance to the origin (0,0) from a Solution Strategy mapped in a 2D plane of M1xM2, where the point (M1,M2) corresponds to M1 and M2 as the Euclidian Distance between the Solution Strategy, and the SCRUM solution and the ISO/IEC solution respectively; and 3) the Solution Strategy must have a Face Validity approbation through the visualization of the 3D scatter graph (see Fig. 1). The Face Validity test means that the sphere of the solution in the 3D-scatter graph be perceived as suitable for being not so far to both SCRUM and ISO/IEC 29110, and thus very near to the theoretically IDEAL solution elaborated.

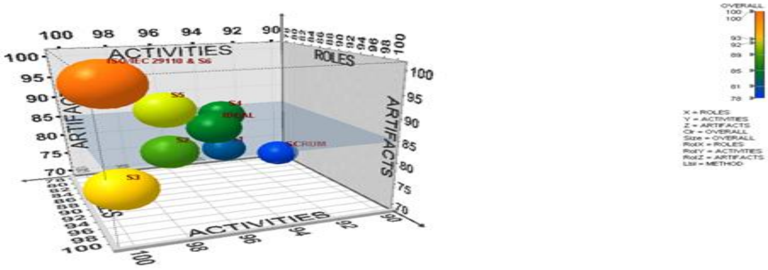


Fig. 1. Face Validity Test of Strategies with a 3D-scatter Graph.

By space limitations, we do not report the specific metrics calculated for the three criteria. However, in the Table 2 is reported the summarization of the results. Thus, we found the Strategy 2 as the unique solution that fitted the three criteria.

Table 2. Final Results from the Three Criteria

	Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5	Strategy 6
Criterion 1	OK	OK		OK		
Criterion 2		OK			OK	
Criterion 3		OK			OK	

Again, by space limitations, we do not report the whole transformations executed on SCRUM on Roles, Activities-Tasks and Artifacts, but in the Table 3, we illustrate representative examples of enhancements taken from the ISO/IEC 29110 Entry Profile.

Table 3. Example of the differences between SCRUM y SCRUM+ Roles, Activities-Tasks and Artifacts

SCRUM	SCRUM+
	Roles (Product Owner)
Represents Customer's interests	**Represents Customer's interests** --- *Accomplish a mandatory formalized project start and closure.*
Authorize and review project outcomes	**Authorize and review project outcomes** --- Verify that the main interests of the customer needs be specified in the project start and closure in formalized way.
Accept or reject final product	**Accept or reject final product** --- Accomplish mandatory signed closure document.
	Activities-Tasks (Sprint Review)
Review of Increment **Review of Project Plan**	**Review of Increment** --- It is important to ensure the project closure with a document as a contract. In this case, it is very important to write the evidence of the present increment. --- It is very important to update the repository because it can be consulted in the future for auditing purposes. In this case, the last increment should be considered. **Review of Project Plan** --- It is important ensure the closure project with a document as a contract. In this case, is very important to write the evidence according with

	the projections to be done in the initial project plan. --- It is very important to update the repository because it can be consulted in the future for auditing purposes. In this case, the project plan should be considered.
Update Product Backlog	**Update Product Backlog** --- It is important to ensure the project closure with a document as a contract. In this case, is very important to write the evidence according with the projections to be done. --- It is very important to update the repository because it can be consulted in the future for auditing purposes. In this case, the product backlog should be considered.
Artifacts (Increment)	
Set of User Stories implemented (DONE) in the last Sprint.	**Set of User Stories implemented (DONE) in the last Sprint.** ---Determine how much was the cost of possible changes on the user stories done. Check against the initial plan.
Value provided to previous Increments from previous Sprints.	**Value provided to previous Increments from previous Sprints.** ---Determine the cost of possible changes on the previous increment from the previous sprint.
Review of potential adjustments to the Sprint Backlog	**Review of potential adjustments to the Sprint Backlog** ---Determine how much was the cost of possible changes on the Sprint Backlog done. Check against the initial plan.

5. Conclusions

This research was pursued with the objective to design systematically an enhanced balanced agility-disciplined SCRUM methodology from the original SCRUM and the ISO/IEC 29110 Entry Profile. Motivation is based on the recurrent literature on the need to strengthen the agile methods (in particular, the Project Management processes) with a balanced approach. A Mean-Ends Analysis technique was used for systematically produce six solutions. Three criteria were fixed for selecting the best one from these six solutions. A final solution SCRUM+ with an overall coverage of 89% was reached as a Means-Ends transformation from the original SCRUM with a 79% level of coverage.

We consider that as any conceptual research, limitations on the reproducibility and internal and external validity of these results can be reported. On reproducibility of results while we use the most original source materials the variability of expertise and self-interpretation of the designers can introduce variations. On internal validity, we consider that the utilization of different source material can introduce variations. On external validity, we conducted an initial conceptual validation from a panel of experts (which is not reported here by space limitations) with suitable initial results but a more robust empirical validation with SCRUM practitioners is planned for next step of this research. As main recommendations, we can report: 1) to conduct the empirical validation with SCRUM practitioners through a survey study; 2) to conduct experiments to compare the utilization of SCRUM vs SCRUM+; 3) to elaborate an Electronic Process Guide of SCRUM+ and promote their utilization in a VSE to study empirically via a Case Study their impacts and limitations.

Finally, we can conclude that balanced agility-disciplined Project Management methodologies are required and that the Means-Ends Analysis technique provides a robust method for systematically produce potential solutions such as SCRUM+ which was transformed from the original SCRUM.

References

1. SEI. 2010. CMMI for Development, Version 1.3. Technical Report CMU/SEI-2010-TR-033. Software Engineering Institute (SEI): Pittsburgh, PA.
2. ISO/IEC. ISO/IEC 12207:2008, Systems and Software Engineering – Software Life Cycle Processes, 2008, ISO organization, Geneva, Switzerland.
3. Succi, G., Valerio, A., Vernazza, T., Succi, G.: Compatibility, standards, and software production, vol. 6, no. 4, pp. 140-146. Standard View (1998)
4. Laporte, C., Alexandre, S., O'Connor, R.: A Software Engineering Lifecycle Standard for very Small Enterprises, pp. 129--141. Springer-Verlag Berlin Heidelberg (2008)
5. SEI: CMMI for development, version 1.2: CMMI-DEV, V1.2. Software engineering institute. CMU/SEI-2006-TR-008, ESC-TR-2006-008. Retrieved from www.sei.cmu.edu (2006)
6. Clarke, P., O'Connor, R.: The situational factors that affect the software development process: Towards a comprehensive reference framework, vol. 54, no. 5, pp. 433--447. Information and Software Technology (2012)
7. Unterkalmsteiner, M., Gorschek, T., Islam, A. M., Cheng, C. K., Permadi, R. B., Feldt, R.: Evaluation and measurement of software process improvement—a systematic literature review. IEEE Transactions on Software Engineering, vol. 38, no. 2, pp. 398--424. IEEE (2012)
8. Organization for Economic Co-Operation and Development (OECD).: SME and Entrepreneurship. (Outlook ed.) (2005).
9. O'Connor, R., Laporte, C.: Towards the Provision of Assistance for Very Small Entities in Deploying Software Lifecycle Standards. In: Proceedings of the 11th International Conference on Product Focused Software Development and Process Improvement (2010).
10. O'Connor, R. V., Laporte, C. Y.: Software Project Management with ISO/IEC 29110. Proceedings of the 19th European Conference on Systems, Software and Services Process Improvement (EuroSPI 2012), vol. 301. Springer-Verlag (2012)
11. ISO.: Software Engineering-Lifecycle profiles for Very Small Entities (VSEs)-. ISO/IEC TR 29110-5-1-1. ISO, Geneva, Switzerland (2012)
12. Dybå, T., Dingsøyr, T.: Empirical studies of agile software development: A systematic review. Information and software technology, vol. 50, no. 9, pp. 833--859 (2008)
13. West, D., Grant, T., Gerush, M., D'silva, D.: Agile development: Mainstream adoption has changed agility, vol. 2, no. 1, pp. 41. Forrester Research (2010).
14. Robillard, P. N., Kruchten, P., d'Astous, P. Yoopeedoo (UPEDU).: A process for teaching software process. In: Proceedings of the 14th Conference on Software Engineering Education and Training, pp. 18--26. IEEE (2001)
15. Sidky, A., Arthur, J., Bohner, S.: A disciplined approach to adopting agile practices: the agile adoption framework. Innovations in systems and software engineering, vol. 3, no. 3, pp. 203--216 (2007)
16. Misra, S. C., Kumar, V., Kumar, U.: Identifying some important success factors in adopting agile software development practices. In: Journal of Systems and Software, vol. 82, no. 11, pp.1869-1890 (2009)
17. Chow, T., Cao, D. B.: A survey study of critical success factors in agile software projects. In: Journal of systems and software, vol. 81, no. 6, pp. 961-971 (2008)
18. Boehm, B.: Get ready for agile methods, with care. In: Computer, vol. 35, no. 1, pp. 64--69 (2002)
19. Boehm, B., Turner, R.: Using risk to balance agile and plan-driven methods. In: Computer, vol. 36, no. 6, pp. 57--66 (2003)
20. Schwaber, K.: Scrum development process. In: Business Object Design and Implementation, pp. 117--134. Springer London (1997)

21. Sutherland, J., Schwaber, K.: The scrum guide. The definitive guide to scrum: The rules of the game. Scrum.org (2013)
22. ISO Software engineering — Lifecycle profiles for Very Small Entities (VSEs) — Part 5-1-1: Management and engineering guide: Generic profile group: Entry profile. ISO/IEC TR29110-5-1-1. ISO, Geneva, Switzerland (2012).
23. Anderson, J. R.: Problem solving and learning. American Psychologist, vol. 48, no. 1, pp. 35--44 (1993)
24. Newell, A., Shaw, J. C., Simon, H. A.. Chess-playing programs and the problem of complexity. In: IBM Journal of Research and Development, vol. 2, no. 4, pp. 320-335 (1958)
25. Newell, A., Simon, H. A.: Human problem solving, vol. 104, no. 9. Englewood Cliffs, NJ: Prentice-Hall (1972)
26. Blankenship, J., Bussa, M., Millett, S., Lewis, R., Foggon, D.: Pro Agile. NET Development with Scrum. Apress (2011)
27. Eclipse Process Framework Project, http://epf.eclipse.org/wikis/scrum/
28. Galvan-Cruz, S., Mora, M., O'Connor, R. V., Acosta, F., Álvarez, F.: An Objective Compliance Analysis of Project Management Process in Main Agile Methodologies with the ISO/IEC 29110 Entry Profile. In: International Journal of Information Technologies and Systems Approach (IJITSA), vol. 10, no. 11, pp. 75--106 (2017)
29. Highsmith, J. A.: Agile software development ecosystems, vol. 13. Addison-Wesley Professional (2002)
30. IDEF: IDEF Method Report - The Standard for Integration Definition for Function Modeling (IDEF0). Federal Information Processing Standards Publication 183. Retrieved from http://www.idef.com/pdf/idef0.pdf (1993).
31. Galal-Edeen, G.H., Riad, A.M., Seyam, M.S.: Agility Versus Discipline: Is Reconciliation Possible?, pp. 331--337 (2007)
32. Schwaber, K: What is Scrum?, Scrum: Its about common sense. Internet document (2007)
33. Laporte, C. Y, Fanmuy, G., Ptack, K.: The Development of Systems Engineering International Standards and Support Tools for Very Small Enterprises. In: 22nd Annual International Symposium of the International Council on Systems Engineering, pp. 9--12 (2012)
34. Vinekar, V., Slinkman, C. W., Nerur, S.: Can agile and traditional systems development approaches coexist? An ambidextrous view. Information systems management, vol. 23, no. 3, pp. 31--42 (2006)
35. Public Site of the ISO Working Group Mandated to Develop ISO/IEC 29110 Standards and Guides for Very Small Entities involved in the Development or Maintenance of Systems and/or Software, http://profs.etsmtl.ca/claporte/english/VSE/ (2016)

Formalizing a Cost Construct Model related to the Software Requirements Elicitation Techniques

Iván Quintanilla[1] and Dante Carrizo[1]

[1] Departamento de Ingeniería Informática y Cs. de la Computación
Universidad de Atacama
Copiapó, Chile
ivan.quintanilla@alumnos.uda.cl, dante.carrizo@uda.cl

Abstract. Requirements elicitation uses several techniques helping to gather needs that software systems to be developed must satisfy. The selection of these techniques, in any moment of the software development, is made based on its context-dependent effectiveness. However, the cost of using these techniques is generally not considered in this decision. This work aims to formalize a cost construct to determine the resources needed to use a specific elicitation technique. For this, authors develop a method based on the fragmentation procedure of each technique in unitary actions or operations. This method has allowed us to establish the cost of 16 techniques elicitation techniques. This cost model proposed seeks to complement the selection criteria of requirements elicitation techniques.

Keywords: Requirements elicitation, elicitation techniques, cost construct, cost model, fragments.

1 Introduction

Requirements engineering aims to collect the needs of stakeholders to define the functions and constraints of the system to be built, which is called as software requirements. This stage includes activities such as: elicitation, analysis, specification, validation and requirements management [1]. Requirements elicitation (RE) is the stage where the developer (or development team), in his / her role of elicitor, interacts with the stakeholder to obtain information regarding his or her problem and / or needs in order to specify the functions and constraints of software system to be built. However, this is not a simple task, since stakeholders can omit or hide information about what they do. This is because they consider that, from their point of view, it is not important or because it constitutes sensitive information that they should not reveal. In order to carry out this task, certain means can be used to help extract information from stakeholders, known as software requirements elicitation techniques.

However, knowing which techniques should be used in a particular software project requires defining criteria for their selection. In scientific literature, there are several proposals that include frameworks or methodologies to select the appropriate

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_3

techniques for a software project [2]. All of them use effectiveness as a criterion of decision, that is, they tend to choose those that allow to better extracting the needs of stakeholders [3]. However, these proposals generally do not take into account the cost incurred by development teams when using an elicitation technique, although it seems to be a relevant aspect for organizations with budget constraints.

For this reason, the objective of this work is to establish a construct associated to the requirements elicitation techniques that allows to describing the cost of its use. The cost of using an elicitation technique seeks to reflect the resources, both human and material, needed to use a technique. The study considered 16 techniques of elicitation for which its prescriptive procedure was formalized as a basis for fragmenting it into unitary actions that can be estimated.

The article continues as follows: Section 2 presents the related work that guide this research. Section 3 describes the methodology of work. Section 4 presents the development of the model. Section 5 corresponds to the discussion section, where the cost model was explained with the prototype technique. Finally, Chapter 6 presents the conclusions of the paper.

2 Related Work

In the literature, several studies that propose directions or guides for the selection of elicitation techniques can be found. Nuseibeh and Easterbrook [4] argue that each technique, in itself, has strengths and weaknesses, which makes them perform better in some situations than in others and in a particular domain of application, which will make it necessary to select the most appropriate according to the characteristics that each software project presents.

Maiden and Rugg [5] argue that, in general, practitioners do not know the wide range of elicitation techniques that exist, and even when they know them, they do not have adequate guidelines for their selection. Therefore, the authors propose a framework for the selection of elicitation techniques that defines a set of criteria, such as purpose of the requirements, types of knowledge to be captured, context of acquisition, among others. However, these criteria are related to the effectiveness that the techniques can have on the characteristics of the project.

Carrizo, Dieste and Juristo [6] propose a degree of adjustment or adequacy that the techniques have in front of a project, that is, how well they fit the software project, for which they define certain factors as elicitor, informant, and problem do-main, among others. These factors are used by the authors in [7], where they present a framework for the selection of software requirements elicitation techniques. This adequacy is related to the effectiveness of the techniques in the project.

Jiang and Eberlein [8] defined 21 project attributes, but their paper show only 7, such as project size, volatility of requirements, requirements volability, project category degree of safety criticality, time constraints and cost constraints. In the study they assessment of the requirements elicitation techniques with respect to software project attributes before mentioned based on the suitability using a Technique Suitability Model.

The cost estimation, principally in the software engineering, is not an easy task, for this there are a several clases of software cost estimation models and techniques, such as: parametric models, expertise-based techniques, learning-oriented techniques, dynamics-based models, regression-based models, and composite-bayesian tecniques. [9]. However, in order to make decisions about which technique to use at any given time, the availability of project resources should also be considered. For this, it is necessary to know the cost structure of applying each technique. Note that there are no studies in the related literature that aim to guide practitioners in this direction.

3 Research Methodology.

According to Rugg and others [10], fragmenting a technique makes it easier to recognize its strengths and weaknesses, so that in order to derive requirements, not only more than one elicitation technique can be used, but also a custom-built method can be created of the fragments of different techniques, for a particular situation.

This research is based precisely on this concept of fragmentation of elicitation techniques. Thus, to obtain the cost construct, we intend to propose a model with which to determine a cost order based on unit actions of each elicitation technique. To obtain such a model, the following steps, shown in Fig. 1, were developed.

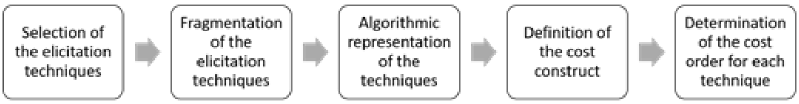


Fig. 1. Research methodology

- Selection of techniques: In order to generate the model, we selected 16 techniques of elicitation among the most popular and well known in software engineering [11].
- Fragmentation of techniques: For each technique, a prescription or use of the technique was formalized. For this, a bibliographic review was performed from where the application procedure and examples of use of the technique in the requirements elicitation were extracted. Once the prescription is defined, the proposed procedure is divided into fragments or unit operations.
- Algorithmic representation of techniques: The procedure of each technique, expressed as a set of fragments, is represented algorithmically, that is, the steps to follow to apply each technique are described as a pseudo-code algorithm.
- Definition of the cost construct: Once the procedure of each technique is expressed algorithmically, the cost construct associated to the techniques is defined, which associates a cost order based on the resources necessary for its use. To calculate this order, an algebraic equation is presented, that expresses the cost order as the sum of the human and material resources used in each of the fragments describing the technique.
- Determination of cost order: For each technique, the cost order is calculated using the model proposed.

4 Development of the Cost Model

This work considered a basis of sixteen elicitation techniques; however, for the purposes of this paper and for reasons of space, henceforth the prototype technique will be used as an example. In this section, the stages carried out to obtain the cost of elicitation techniques are presented. The first subsection shows the prescription of the technique, i.e. the procedure to use it and also its fragmentation and structure; the second subsection defines the cost model; and the third subsection describes the application of the model in order to calculate the cost of the prototype technique.

4.1 Prescription and structuring of elicitation techniques

The prescription of a technique corresponds to its use guide, that is, it describes a procedure of how the technique is applied to elicit software requirements. Prescription of the prototype technique is presented below.

Prototype technique.

This method is associated with the idea of developing a simplified version of the system to be built. These artifacts are called software or hardware prototypes. According to Naumann and Jenkins [12], a software prototype corresponds to an intentionally incomplete system to be modified, expanded, complemented or supplanted. Its use is varied, although in this case we are interested those that serve to obtain requirements. Several authors, Naumann and Jenkins [10], Seen [13], Asur and Hufnagel [14] and Stephens and Bates [15], have applied this technique. However, the application procedure, mainly the execution stage, is not formally defined. It is for this reason that an original set of steps for the execution of the technique is pro-posed, although the general procedure is based on Senn's [13] (See Table 1).

Table 1. Procedure of prototype technique.

Prototype Technique	
Ex ante Execution	
Step	**Actions**
1	**Identification of Known Requirements** The educator and stakeholders identify the known requirements and determine the purpose of the application prototype
2	**Development of a Model** • Explain the iterative method of the prototype and responsibilities to stakeholders who participate directly in the whole process. Stakeholders and educators jointly identify the data that is necessary for the system and specify the output that the application must produce. In the development of a prototype the following components are prepared: • The language for the dialogue or conversation between the user and the system • Screens and formats for data entry • Essential Processing Modules • System Output
3	**Determination of inputs and outputs** Incorporating in the input / output interface features representative of those that will be included in the final system allows for greater accuracy in the session.
4	**Logistics** • Agree on the date and time of the session and the place of the session. Prepare the room where the session will be held and cite the participants. The full duration of the session should not exceed 2 hours.
Execution	
5	**Development** • To perform the session, the following protocol is proposed: • At the beginning of the session, the educator should introduce himself / herself, state the purpose of the session, list the prototype functionalities, and describe the context of the prototype. • Allow the stakeholder to operate the prototype, allowing it to execute, one by one, the functions implemented in it. • In each function, the eductor must explain to the stakeholder what the role is doing, the stakeholder must execute the function. The stakeholder can ask questions to the engineer, regarding doubts with the execution of a function as data entry, steps to follow, etc. The eductor must answer the questions asked. • During the execution of a function, the stakeholder can make comments or suggestions regarding things that he considers to be missing or that should be modified. The eductor should take note of each of the suggestions or comments made by the stakeholder. • Once all of functions have been performed, the educator concludes the session by presenting a summary of the comments and suggestions made by the stakeholder, which were documented during the presentation of the prototype.
Ex post Execution	
6	**Record of captured information** Formalize captured information that can serve as input to the definition of requirements.

Fragmentation.

The prescription of the techniques presented in the previous section is done through a procedure and, as such, it is possible to represent it structurally through a pseudo code (See Table 2). In addition, this structuring will facilitate the determination of cost magnitude for the model to be proposed.

Table 2. Fragmentation of the procedure of prototype technique

```
    // Ex ante execution
/ ** Identification of known requirements ** /
REPEAT
     [Identify Known Requirements]
UNTIL (end of known requirements)
[Determine prototype purpose]

/ ** Development of a model ** /
[Prepare language for dialogue between actors] [Preparing screens and formats for data entry]
[Prepares essential processing modules] [Prepare system output]

/ ** Determination of inputs and outputs ** /
[Incorporate own characteristics of the final system]

Execution
/ ** Prototype review
Perform the experience with the system under real conditions
** /
[Decide to record session]
IF (recording session) THEN
     [Start recording] [Presentation of the analyst] [Declare purpose of the session] [List prototype
     functions] [Describe prototype context] [Allow the stakeholder to operate prototype]
     WHILE (There are functions)
          [Explain what the function does] [Lead stakeholder to execute function]
          IF (there are questions) THEN
               [Listen to stakeholder question] [Answer question to stakeholder]
          ENDIF
          IF (there are suggestions) THEN
               [Listen to stakeholder suggestion] [Answer suggestion to stakeholder] [Take note of
               suggestion]
          ENDIF
          IF (there are comments) THEN
               [Listen to stakeholder comment] [Reply to stakeholder] [Take note of comments]
          ENDIF
     ENDWHILE
ENDIF
            [Presentation of the eductor]
            [Declare purpose of the session] [Enumerar funciones del prototipo] [Describe prototype
            context] [Allow the stakeholder to operate prototype]
            WHILE (There are functions)
               [Explain what the function does] [Lead stakeholder to execute function]
               IF (there are questions) THEN
                    [Listen to stakeholder question] [Answer question to stakeholder]
               ENDIF
               IF (there are suggestions) THEN
                    [Listen to stakeholder suggestion] [Answer suggestion to stakeholder] [Take note of
                    suggestion]
                 ENDIF
                 IF (there are comments) THEN
                    [Listen to stakeholder comment] [Reply to stakeholder] [Take note of comments]
                 ENDIF
               ENDWHILE
               IF (there are suggestions) THEN
                 [Submit summary of suggestions taken]
               ENDIF
               IF (there are comments) THEN
                    [Present summary of comments taken]
               ENDIF
            ENDIF
               // Ex post execution
               / ** Record of captured information ** /
[Formalize the information obtained]
```

Cost model of elicitation techniques.

The computational cost or machine cost refers to the resources needed to execute an algorithm. These resources, as it is read above, involve a temporal resource (time) and a spatial resource (machine resources, memory), which are studied through algorithmic complexity (temporal and spatial complexity). The complexity analysis aims to quantify the physical measures: "execution time and memory space" and compare different algorithms that solve the same problem in order to determine their efficiency [16]. An algorithm is, in simple words, a finite set of steps that must be performed to solve a problem.

The selection of one or more techniques to elicit requirements is related to their effectiveness in a software project. There are several proposals to select techniques according to the context of the project, which is defined through certain attributes that describe it. The effectiveness of the elicitation techniques (T) can be seen as a set of paths that allow going from a state t0 to a state t1, where a certain number of requirements have been obtained for the software to be constructed. This can be seen in Fig. 2.

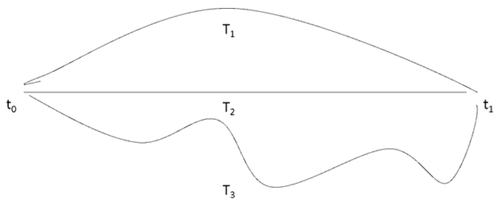


Fig. 2. Effectiveness of elicitation techniques

Equation (1) shows an analytical description of the effectiveness.

$$\text{Effectiveness} = \frac{\text{N° Requirements elicited}}{\textit{Time}} \quad (1)$$

The effectiveness may be similar to the temporal complexity in the analysis of algorithms. The above means that to obtain a certain amount of requirements elicited, one or several techniques allow me to arrive faster (in a shorter time) from t0 to t1, so they are said to be more effective than others, but, what will be the cost of taking that road faster?

The cost related to an elicitation technique corresponds to the resources needed to use this technique. So, following the analogy of the figure, the associated costs of elicitation techniques could be seen as tolls on a route. This means that although the road can be seen in the first instance as the fastest, it may have many tolls on the route so its cost will be higher than other roads. This situation is plotted in Fig. 3.

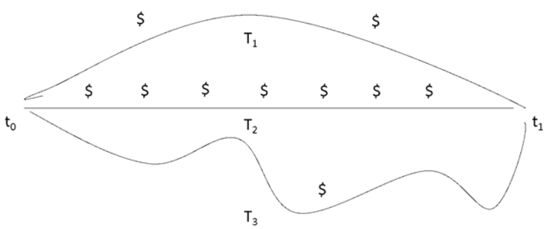


Fig. 3. Cost related to elicitation techniques

Equation (2) shows the analytical representation of the cost.

$$\text{Cost} = \frac{\text{N° Requirements elicited}}{\textit{Resources}} \tag{2}$$

The cost may be similar to spatial complexity in analyzing algorithms. This means that to obtain a certain number of requirements, a set of resources will be necessary to perform this task. In the case of elicitation techniques, resources include human and financial resources.

Cost Model.
The cost model is an algebraic equation for calculating the cost related to software requirements elicitation techniques. The cost model is shown in equation (3).

$$C_T = \sum_{i=1}^{n} Crh_i + Crm_i \tag{3}$$

Where:

- C_T: Cost related to the technique.
- Crh: Cost of human resources used in the technique.
- Crm: Cost of material resources used in the technique.
- i: Represents each of the fragments or unit operations of the technique.

The material resources represent the use of things (measured in hours-object) necessary to apply the technique as materials, offices, etc. The cost here is represented by economic cost segments (Low, LC; Medium, MC and High, HC). For example, if an office or classroom is to be leased for the time of duration of the elicitation session (td), then, as shown in equation (4), the usage cost of the classroom will be:

$$Use\ Cost = td * HC \tag{4}$$

A high cost means that the cost (per hour) of the material resource is greater than that of other material resources such as pencils, computers, minutes of call, etc. The cost model seeks to determine the cost in both human and material resources that must be incurred in each of the unit operations that are carried out when using an elicitation technique. This is outlined as a second criterion for the selection of techniques, i.e. as a second decision filter, since they can first be selected according to their effectiveness, and then those whose cost the company may incur. When obtaining the cost and the effectiveness, a cost-benefit relation can be obtained in favor of the selection of the elicitation techniques.

Applying the cost model of elicitation techniques.
In this section, the cost model is applied to the prototype technique, for which the cost of the unit operations that compose it will be calculated. Table 3 shows the calculation performed. The cost model consists of an algebraic equation that considers the human and material resources needed to use an elicitation technique. However, the determination of the cost order for each technique was performed under certain conditions and / or restrictions, such as the duration of the elicitation sessions. This duration will depend on what the elicitor and the stakeholder agree.

Another condition is that the cost of the techniques is for the use of these once only, that is, once a technique is used, it can be used again; another technique can be used, or can advance the elicitation process of software development.

In group techniques, the technique is conditioned to only one elicitor (EL), who re-presents the hour-man value of the elicitor. However, in these group techniques, there can be more than one stakeholder, and they can be of different types: STKg, if he or she is a manager; STKj, if he or she is a head of unit; and STKu, if he or she is a user. These denominations represent the hour-man value of each of the stakeholder types. For individual techniques, the STK variable represents the hour-man value of the stakeholder when it is an individual technique, that is, when there is only one stakeholder.

Unit operations represent actions to carry out the procedure. These represent small tasks, which cannot be divided into smaller tasks. However, there are tasks that can be divided into smaller ones, but that for simplicity of the study they were decided to express as a unitary operation.

5 Discussion

Table 4 shows a summary of the cost and cost order obtained for each elicitation techniques. According to the table, it was obtained that, for the individual techniques, that is, those in which the interaction is between one elicitor and one stakeholder, the cost order is 1, that is, constant. This means that for these techniques, the cost will not depend on the size of the input, or the number of stakeholders. The cost will depend on other factors, such as the type of stakeholder selected, since if the stakeholder is a manager the effort required will be greater because the value man-hours is greater than a stakeholder-user. Therefore, the cost will be higher. Another factor that may influence the cost of a technique is the cost person-hours of the elicitor, and this will depend on the experience of the same.

On the other hand, the group techniques present an order of cost n, that is to say linear, which means that its cost will depend on the number of participating stakeholders. As the stakeholders number increases, the effort will increase, i.e. the number of man hours will increase. As the cost order considers human and material resources, in addition to the effort required using a technique, the cost will also depend on the cost of the material resources needed for its use.

Accordingly, based on number of stakeholders, material resources and elicitor’s experience, among elicitation techniques considered the interview (unstructured and structured) present the lowest cost (interaction 1-1 between stakeholder and elicitor, do not require many material resources and do not require much elicitor’s experience), followed by the questionnaire technique or the Delphi method (it do not require much material resources, but both techniques consider a group of stakeholders, so the effort required grows according to the stakeholders and require a medium level elicitor’s experience). On the other hand, use cases and laddering techniques could also be considered with the lowest cost, since they are of order 1, that is to say constant, and does not require many material resources for its execution, however in this case, the elicitor's experience influences, which increases the man-hours value of effort and, finally, the cost of the technique.

Table 3. Calculation of cost order of prototype technique.

	Human resources			Material resources			
Unit operation	**Person**	**Time**	**Worker (h*h)**	**Amount**	**Object**	**Time**	**Use Cost (h*o)**
Identify known requirements	EL	t	t*EL	1	Computer	t	t*MC
					Internet	t	t*MC
Determine purpose of prototype	EL	t	t*EL	1	Computer	t	t*MC
					Internet	t	t*MC
Prepare language for dialogue between actors	EL	t	t*EL	1	Computer	t	t*MC
					Internet	t	t*MC
Prepare screens and formats for data entry	EL	t	t*EL	1	Computer	t	t*MC
					Internet	t	t*MC
Prepares essential processing modules	EL	t	t*EL	1	Computer	t	t*MC
					Internet	t	t*MC
Prepare system output	EL	t	t*EL	1	Computer	t	t*MC
					Internet	t	t*MC
Incorporate characteristics of the final system	EL	t	t*EL	1	Computer	t	t*MC
Decide to record the session	EL	t	t*EL	1	Computer	t	t*MC
					Internet	t	t*MC
Start recording				1	Notebook (Logbook)	td	td*LC
Presentation of the analyst				1	Pencil	td	td*LC
Declare purpose of the session				1	Clock	td	td*LC
List prototype functions				1	Recording device	td	td*MC
Describe prototype context				1	Recording memory	td	td*MC
Allow stakeholder to operate prototype				1	Place	td	td*HC
Explain what the function does				1	Computer	td	td*MC
Lead stakeholder to execute function							
Listen to stakeholder question	EL + STK	td	td*(EL + STK)				
Answer question to stakeholder							
Listen to stakeholder suggestion							
Answer suggestion to stakeholder							
Take note of suggestion							
Listen to stakeholder comment							
Answer to stakeholder comment							
Take note of comments							
Present summary of suggestions taken							
Present summary of comments taken							
Formalize the information obtained	EL + STK	t	t*EL	1	Computer	td	t*MC
		Total HR	**9t*EL + td*(EL + STK)**			**Total MR**	**16t*MC + 3td*LC + 2td*MC + td*HC**
		Cost	**9t*EL + td*(EL + STK) + 16t*MC + 3td*LC + 2td*MC + td*HC**				
		Cost Order	**(1) CONSTANT**				

Table 4. Summary cost order and cost order for eliciting techniques.

TÉCHNIQUE	COST	COST ORDER
Open-Ended Interview	23t*EL + td*(EL+STK) + 11t*LC + 30t*MC + 3t*LC + 2td*MC + td*HC	(1) CONSTANT
Structured Interview	22t*EL + 3t*(EL+STK) + td*(EL+STK) + 11t*LC + 30t*MC + 2td*MC + 3td*LC + td*HC	(1) CONSTANT
Observation of usual tasks	5t*EL + 4t*(EL + n*STKg + n*STKj + n*STKu) + 2t*(EL+CLI) + td(EL + n*STKg + n*STKj + n*STKu) + 9t*LC + 14t*MC + 3td*LC + 2td*MC + td*HC	(n) LINEAR
Card Sorting	7t*EL + t*(EL+STK) + td*(EL+STK) + 3t*LC + 14t*MC + 3td*LC + 2td*MC + td*HC	(1) CONSTANT
Questionnaire	8t*EL + t*(EL + n*STKg + n*STKj + n*STKu) + 16t*MC + 3td*LC + 2td*MC	(n) LINEAR
Protocol Analysis	9*t*EL + t*(EL+STK) + td*(EL+STK) + 8t*LC + 14t*MC + 3td*LC + 2td*MC + t*HC	(1) CONSTANT
Repertory Grid	5t*EL + 4t*(EL+STK) + td*(EL+STK) + 15t*LC + 8t*MC + 3td*LC + 2td*MC + td*HC	(1) CONSTANT
Brainstorming	5t*EL + 3t*(EL + n*STKg + n*STKj + n*STKu) + td*(EL + n*STKg + n*STKj + n*STKu) + 6t*LC + 10t*MC + t*HC+ 3td*LC + 2td*MC + td*HC	(n) LINEAR
Nominal Group Technique	7t*EL + td * (EL + n*STKg + n*STKj + n*STKu) + t*(EL + n*STKg + n*STKj + n*STKu) + 3t*LC + 11t*MC + t*HC + 3td*LC + 2td*MC + td*HC	(n) LINEAR
Focus Group	11t*EL + 8t*(EL + n*STKg + n*STKj + n*STKu)+ td*(EL + n*STKg + n*STKj + n*STKu) + 8t*LC + 23t*MC + 7td*LC + 2td*MC + td*HC	(n) LINEAR
Delphi Method	11t*EL + t*(EL + n*STKg + n*STKj + n*STKu) + td*(EL + n*STKg + n*STKj + n*STKu) + 20t*MC + 2td*LC + 4td*MC	(n) LINEAR
Laddering	4t*EL + 3t (EL + STK) + td*(EL + STK) + 10t*LC + 6t*MC + t*HC + 2td*MC + 2td*LC	(1) CONSTANT
Participant Observation	t*EL + td*(EL + n*STKg + n*STKj + n*STKu) + t*MC + 3td*LC + 2td*MC + td*HC	(n) LINEAR
Prototypes	9t*EL + td*(EL + STK) + 16t*MC + 3td*LC + 2td*MC + td*HC	(1) CONSTANT
Joint Application Development	12t*EL + td*(EL + n*STKg + n*STKj + n*STKu) + 24t*MC + 3td*LC+ 2td*MC + td*HC	(n) LINEAR
Use Cases	3t*EL + 3t*(EL + STK) + td*(EL + STK) + 10t*LC + 4t*MC + 3td*LC + 2td*MC * td*HC	(1) CONSTANT

At the other extreme are the group techniques that prove to be the most costly. This is the case of focus group, since it requires a specific group of stakeholders, which could have a high value of person-hours, increasing the effort required to perform the technique. In addition, many material resources are required both for the group, in general, and for each stakeholders. In these techniques, the higher cost will depend on the stakeholders selected. Alongside this, the JAD technique includes the role of the executive sponsor as a stakeholder, which can increase the value of person-hours and ultimately increase the cost. In the case of techniques Observation of habitual tasks and Participant observation, the role of the client appears which, as in the previous case, can increase the value of person-hours and, therefore, the final cost of the technique.

In summary, the techniques that present the least cost are the Interviews, both Unstructured and Structured. The most costly are JAD, the Observation of habitual tasks and Participant observation. The final cost will depend on the characteristics of the project and the cost incurred both in human resources and materials needed to use a technique.

5 Conclusions

Software engineers need to select the most appropriate requirements elicitation technique at each point in software development. This should not only contribute to the quality of the software product that meets the needs of the stakeholders, but also must allow adjusting to budgetary constraints of the project. This research work is oriented in this direction, since, until now, elicitation techniques were selected considering only their effectiveness, without considering the cost involved.

The proposed cost model makes it possible to calculate the cost of using a requirements education technique. Thus, cost pretends to be a relevant criterion or filter for the selection of techniques, so that once the effectiveness and cost of a technique have been determined, a cost-benefit relationship can be established for the techniques in a particular software development project.

In this work, the cost construct does not calculate the final cost incurred in using an elicitation technique, but instead gives an order of magnitude of cost, that is, the cost is expressed as a mathematical equation in order to have a notion of the cost when using the technique. In addition, the cost model considers only material and human resources as a first approach in order to determine the cost of elicitation techniques, i.e. how much it ultimately costs to use an elicitation technique.

The work allows to associate cost reasons with sixteen well-known techniques, line of investigation unpublished in the software engineering. As future work, we hope to study the development of a unified model that considers both the effectiveness and the cost of elicitation techniques in order to support the development teams in their selection.

References

[1] Abran, A., Moore, J. W.: SWEBOK: Guide to the Software Engineering Body of Knowledge. ed. 2004, IEEE Computer Society, California (2004).

[2] Davis, A., Dieste, O., Hickey, A., Juristo, N., Moreno, A. M.: Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a

Systematic Review. Proc. 14th IEEE International Conference Requirements Engineering, 179 – 188 (2006).
[3] Carrizo, D., Ortiz, C., Aguirre, L.: What Do Researchers Mean by The Right Requirements Elicitation Techniques. Ingeniare. Revista chilena de ingeniería, vol. 24, pp. 263-273, April 2016
[4] Nuseibeh, B., Easterbrook, S.: Requirements Engineering: A Roadmap. In A.C.W. Finkelstein, proceedings of the 22nd International Conference on Software Engineering, ICSE'00, pp. 35-46. IEEE Computer Society Press, Limerick (2000).
[5] Maiden, N., Rugg, G.: ACRE: Selecting Methods for Requirements Acquisition. Software Engineering Journal, 11(3), 183-192 (1996).
[6] Carrizo, D., Dieste, O., y Juristo, N.: Study of Elicitation Techniques Adequacy. Proceedings 11th Workshop on Requirements Engineering, WER, pp. 104-114, Barcelona (2008).
[7] Carrizo, D., Dieste, O., Juristo, N.: Systematizing Requirements Elicitation Technique Selection. Information and Software Technology, 56 (6), 644-669 (2014).
[8] Jiang, L., Eberlein, A., Far, B., Mousavi, M.: Selecting Requirements Engineering Techniques Based on Project Attributes – A Case of Study. 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems, 269-278, Tucson (2007).
[9] Boehm, B., Abts, C., & Chulani, S. (2000). Software development cost estimation approaches—A survey. Annals of software engineering, 10(1-4), 177-205.
[10] Rugg, G., Mcgeorge, P., Maiden, N.: Method Fragments. Expert Systems, 17(5), 248-257 (2000).
[11] Carrizo, D. "Comparación de efectividad de las técnicas de educción de requisitos software: visión novel y experta". Ingeniare. Revista chilena de ingeniería, 20 (3), 386-397.
[12] Naumann, J. D., Jenkins, A. M.: Prototyping: The New Paradigm for Systems Development. Mis Quarterly, 29-44 (1982).
[13] Seen, J.: Análisis y Diseño de Sistemas de Información. 2da Edición, McGraw-Hill (1992).
[14] Asur, S., Hufnagel, S.: Taxonomy of rapid-prototyping methods and tools. Proceedings, Fourth International Workshop on, IEEE, 42-56, Research Triangle Park (1993).
[15] Stephens, M. A., Bates, P. E.: Requirements Engineering by Prototyping: Experiences in Development of Estimating System. Information and Software Technology, 32(4), 253-257 (1990).
[16] Schaeffer, E.: Complejidad computacional de problemas y el análisis y diseño de algoritmos. Publicación en línea (2011).

WYDIWYN – What You Define, Is What You Need: Defining Agile/Traditional Mixed Methodologies

Andrés Felipe Bustamante[1], Rafael David Rincón [1],

[1] Universidad EAFIT, Medellín, Colombia,
{afbustamar, rrincon}@eafit.edu.co

Abstract. Defining software development methodologies may jointly include traditional and agile frameworks practices. Agile frameworks give teamwork structure and dynamics which are focused on delivering value in an incremental way. Traditional frameworks establish clarity in work products and project phases along all its lifecycle. The combination of frameworks can help to apply best practices of different models, but increasing the complexity in the methodology definition when having to start from contexts of a different nature. You will see in this article a carried out study that confirms the use of agile/traditional mixed methodologies in the industry, the difficulties that emerged and that are still seen in this approach, and the creation, setting and analysis of a guide that helps to uncomplex the definition of mixed methodologies.

Keywords: Software Process Improvement, Software Development Methodologies, Agile Frameworks, Traditional Frameworks, VSEs.

1 Introduction

Defining a methodology is searching to establish a group of procedures that help to carry out an objective [1] [2], but What happens if different kind of frameworks are combined in a methodology definition?

Software development projects can be supported by a large diversity of reference frameworks [3], that by the type of methodologies in which they are frequently used [4], we have categorized them throughout this paper in traditional frameworks [4] such as RUP [5], CMMI [6], ISO/IEC 9001 [7] or PSP/TSP [8], and agile frameworks [9] such as Scrum [10], Kanban [11] or LEAN [12].

Traditional frameworks are based on a best-of practice and forms ensemble which help to take into account vital factors in a project lifecycle and that must be defined from the start [13], but that can generate too extensive phases, thanks to the greater detail in documentation or required activities [14].

Agile frameworks, on the other hand, propose an incremental evolutionary process, where the biggest value-generating activities for the customer are prioritized to be carried out and delivered as soon as possible, but they can leave out important factors in project planning, due to these frameworks not proposing the dimensions presented by traditional frameworks [14].

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_4

Agile and traditional frameworks integration is a reality which is being contemplated in the industry [2] [1] and there are proposals that allow to analyze applied mix solutions [14] [15] [13]. However, this approach generates an additional complexity in requiring to identify, correlate and select phases, roles, tasks or work products among different frameworks, that if it is not properly carried out, it may result in process repetition, non-adequate role execution or deficient definition of work teams [16] [17] [18].

WYDIWYN -What You Define, Is What You Need- model proposed in this paper looks for helping to define a methodology focused on what the company needs, simplifying the inherent complexity when combining agile and traditional frameworks.

This paper is structured in the following way: a background study will be seen in section 2 about mixed agile/traditional methodologies; the difficulty that we are seeking to solve is detailed in section 3; the design and structure of the suggested solution is presented in section 4; and an analysis, conclusions and a proposal of future work is found in the end in section 5.

2 Background

We perform a study of works that take into account approaches for software development methodologies, including agile and traditional practices based on the established protocol in [19].

Previous studies [20] [21] were adopted as base and they were refined looking to identify solution initiatives that have been approached in the last years, and which contemplate agile and traditional frameworks best practices along with advantages, disadvantages and comparatives. Applying the defined protocol, a search for various resources was made and the articles showed in table 1 were selected.

Table 1. Selected works. Source: Own elaboration

Work	Year	Work	Year	Work	Year	Work	Year	Work	Year
[17]	2012	[1]	2014	[22]	2013	[18]	2015	[21]	2016
[13]	2012	[2]	2014	[14]	2013	[23]	2015	[24]	2016
[25]	2013	[15]	2013	[26]	2014	[20]	2015	[27]	2014
[28]	2014	[29]	2011	[30]	2010	[16]	2013		

General ideas about the traditional reference frameworks are managed to be abstracted from the selected works [2] [13] [14] [30] [18] [23], where the projects are laid out as predictable and linear, with clearly-defined limits [2] and with tangible results in planning and design phases [25].

It was identified from the agile frameworks that they are characterized for including worries about the people development and team empowerment [1], at the same time, they are focused on the reduction and risk control, through software development and delivery in shorts periods of time [14].

Agile/traditional frameworks integration is contemplated from recommended best practices [13] [2] [15] [17] [29], directly bringing up mixed models and applying why it is useful to include good practices of agile and traditional frameworks in a joint way [28] [16]. Quick feedback, better tolerance to change, risk management [14] [1], good work structure, clear and accurate artifacts [2], contemplation of key necessary activities and practices to completely cover the lifecycle of the projects [13] are found among the main advantages.

Integrating different types of frameworks brings with it new challenges and problems that increase the complexity of a project. Contemplating for combined frameworks stressed differences in the software development lifecycle [16] [21], roles definitions [16] [17] [18], related activities that end up being redundant, complementary or that do not generate value according to a determined context [20] [13] [17] [14], work products that are redundant or that achieve similar objectives [16] and limitants in team sizes [1] [27].

As a conclusion of the study, agile and traditional frameworks integration is part of the current software industry [13] [2] [15] [17] [29] and it brings with it great advantages [13] [2] [15] [17] [29], but new challenges at the same time that end up in an extra complexity for the methodological definition [20] [16] [21] [17].

3 Difficulty

When defining a methodology with an agile/traditional mixed scheme is proposed by a company, there is an additional complexity in the starting point since a deep analysis about topics that have not been finished to lay out is needed [20].

There are difficulties in the activity integration of different frameworks, despite the achieved approaches in mixed agile and traditional proposals, increasing the complexity [13] [17] [14]. The resulting complexity is summarized in table 2.

Table 2. Agile and traditional frameworks integration complexity. Source: Own elaboration

ID	Description	Inherent complexity	Associated difficulty
CM P1	Lifecycle homologation	Need to identify and fit in a similar way, phases between frameworks.	Risk of execution of tasks in wrong phases, generating repetition of work or lack of ability to control risks at the right time [16] [21].
CM P2	Roles homologation	Need to identify and fit in a similar way, roles between frameworks.	Risk of misrepresenting responsibilities in the execution of tasks, affecting the quality of the product performed [16] [17] [18].
CM P3	Tasks correlation	Need to identify and correlate tasks characteristics between frameworks.	Risk of performing unproductive or unnecessary tasks, if the correlation and objective of activities between different frameworks is not clear [20] [13] [17] [14].
CM P4	Work products correlation and redefinition	Need to identify correlates and redefine similar goal work products between frameworks.	Risk of generating redundant or disjointed work products [16], if their end and correlation between frames are not known.
CM P5	Team size homologation	Need to identify and fit in a similar way, team size between frameworks.	Risk of unproductive times or deterioration of quality of deliverables, if the guidelines of suitable size of work equipment are not followed [1] [16].

The difficulty encompassed in this work is summarized in answering how can software development practices of agile and traditional frameworks be simplified. We work with the hypothesis that, having a multimodel structure of agile and traditional practice integration, we manage to reduce the complexity framed in table 2.

4 WYDIWYN -What You Define, Is What You Need-: Defining Agile/Traditional Mixed Methodologies

The definition of a methodology seeks to establish a group of procedures that help to achieve an objective [1] [2].

We identified in the previous sections advantages of agile/traditional mixed methodologies definition, such as quick feedback, better tolerance to change and risk management [14] [1], combined with good work structure, clear and accurate artifacts [2] and contemplation of key activities of sofware development lifecycle [13].

We saw how the inherent complexity to framework combination can lead, among other aspects, to repeated processes and work, lack of risk control, non-adequate role execution of tasks or deficient definition of work teams. The wrong definition of a methodology as a result of this complexity, guides us to institutionalize a way of work that does not reflect what the company needs.

There are proposals in diverse areas of knowledge of multimodel schemes in order to facilitate the integration of different frameworks [31] [32], where structuring and homologation techniques are laid out, and advantages when proposing integrating models [20] are identified.

The WYDIWYN -What You Define, Is What You Need- model proposed in this work makes use of structuring and homologation techniques and it has the objective of providing a guide for definition of agile/traditional mixed methodologies, focused on what a company needs, simplifying the identification, correlation and selection of phases, roles, tasks and work products among different frameworks.

The guide conception requires defining a structure, a visual model and a documentation and specification of tasks and work products.

4.1 Structure and visual model

Multimodel scheme approaches highlight advantages such as the inclusion capacity of different framework features, the incremental enrichment of the multimodel structure and the visual representation of activities and correlations in order to facilitate the understanding [31] [33]. It is important to lay out a scheme that allows to easily identifying these correlations, together with delivering different views of the information according to the aspects that are desired to be validated in any given moment.

Three important characteristics are determined that will help to define the guide, based on the complexity identified in the last section. First, the need of role (CMP1), lifecycle (CMP2) and team size (CMP5) homologation, leads us require a central axis in which the different frameworks can find a way of comparing themselves within.

Second, the need of task (CMP3) and deliverables (CMP4) correlation from the made homologation, it leads to require a group of variables that allow to structure a framework from different perspectives. Third, the need of an analysis of different correlated and homologated frameworks, leads to require a visual representation that allows interacting and validating the interested frameworks from the defined variables. These three characteristics can be appreciated in table 3, in the light of the defined complexity and how they are related.

Table 3. Structural and visual characteristics of the guide. Source: Own elaboration

ID	Name	Justification	Including
CT1	Central axis	Based on the identified homologation complexity (CMP1, CMP2, CMP5).	(Base1) Lifecycle (Base2) Roles (Base3) Team size
CT2	Structural variables	Based on the complexity of identifying, correlating and redefining tasks and work products (CMP1, CMP2, CMP3, CMP4).	(Var1) Lifecycle phases (Var2) Roles (Var3) Tasks (Var4) Work products
CT3	Visual representation	Based on the central axis and structural variables (CT1, CT2)	(View1) Central axis view (View2) Phases view (View3) Roles view (View4) Tasks view (View5) Work products view

Structure: The ISO/IEC 29110 standard [34] is the based used for the CT1 characteristic, due to it allowing to model agile and traditional frameworks, providing profiles that take into account the lifecycle and roles of software development, and it is focused on small work entities the same as the agile frameworks [34] [18] [24]. Additionally, there are currently definitions of agile/traditional mixed definitions that have managed to be executed from this standard [16] and it allows comparing the methodologies with international standards, and even being certified [34].

A diagram is approached for the following characteristic CT2, in figure 1, where is evidenced how all the structuring variables will interrelate with the central axis of characteristic CT1, for the integrations of agile and traditional frameworks.

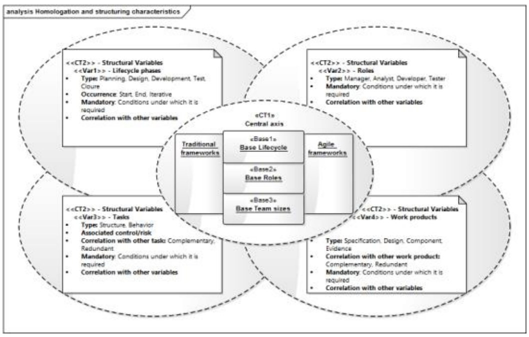


Fig. 1. Homologation and structuring characteristics. Source: Own elaboration

The figure presents an overlapped central axis that looks for lifecycle, roles and the team size homologation, addressing the CMP1, CMP2 and CMP5 complexities. The structuring CT2 variables are featured around and behind the central axis, that star from the need of simplifying the analysis to include phases, roles, tasks or work products in a methodology. This representation indicates that the structure variables must be expressed in terms of the central axis relation. Additionally, the variables present a categorization, specification of obligatory nature and a series of attributes that allow them to be catalogued and easily correlated covering the CMP3 and CMP4 complexities. Finally, the graphic overlaps the area spectrum of each structuring variable with the others, indicating that there are an intersection and a direct correlation among them and they must lay out bonds not only with the main axis, but also among the variables.

Some bases for the progress of the guide are initially defined, produced from an ISO/IEC 29110 [34] basic profile, the requirements and software development process areas of CMMI-DEV [6] and the SCRUM [10] definition. The lifecycle base phases are project planning, implementation and evaluation of the project plan, requirements analysis, software design, software development, testing and integration, product delivery, closure and self-assessment. The base roles are product receiver, team leader, requirements analyst, software designer, software developer and tester. The base size for a work team is between 6 and 10 people. These base characteristics can be redefined or increased as frameworks are added, when a direct relationship with the framework that is desired to be included is not possible to be found.

Guide visual model: Once the variable interrelation is laid out, it is required to create a visual mode of the guide addressing the CT3 characteristic. Rios suggests a multimodel structure[33] in the Enterprise Architect tool from Sparx Systems, creating hierarchical panels, visual schemes and correlation among components that "allow to see in a simple way all the elements that take part of a model" [33]. In turn, in this model is included a homologation and correlation of traditional frameworks such as ISO/IEC 9901, CMMI-SVC, CMMI-ACQ, CMMI-DEV and COBIT [33]. This approach offers a very interesting abstraction level, considering that it generates documentary and visual correlations inside the same model and inside diverse models as well.

This structure, laid out by Rios [33] was abstracted and extended in order to include the findings of this work and to be able to homologate both agile and traditional frameworks, representing the scheme in figure 1.

The important issue is to be able to hierarchically represent the frameworks, to connect components and to have different views of the information, agreed with the laid out structuring variables, allowing to correlate the lifecycle, roles, tasks and work products of frameworks, as it is shown in figure 2.

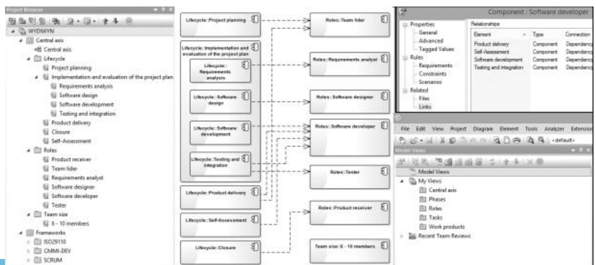


Fig. 2. Hierarchically structuring, correlations and views. Source: Own elaboration

It is important to highlight how we can have complete hierarchized structures of the different frameworks and of the central axis of the guide, allowing to explore in an organized way the different components and their substructures. We can appreciate how different types of relationships exist and that those can be applied to components or documents, allowing to search and navigate between correlations. Enterprise Architect delivers a group of tools to have access to preconfigured views of the information as well as personalized views, all of this starting from the created hierarchies and the correlations added to the components.

4.2 Documentation and tasks and work products specification

It is important to clearly define what resulting artifacts of each activity are expected as well as the required structure and its relation to other frameworks. Figure 3 shows how the ISO/IEC 29110-2 standard [34] presents a section of reference and taxonomy frameworks that explains how to establish tasks and work products, showing the information characteristics required and permitting to compare different frameworks.

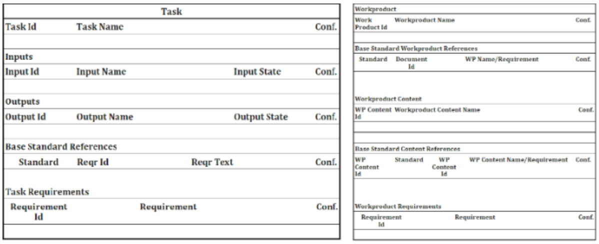


Fig. 3. Tasks and work products documentation. Source: ISO/IEC 29110 [34]

This way of documenting allows having clarity of the purpose, the entries, the exits and the correlations among tasks and work products. Moreover, it integrates the capacity of relating these artifacts with international standards, which lets comparing a methodology with international norms, offering more possibilities to be measured and certified by external agents [34].

For an easy generation of this documentation, the task attributes and the work products are entered from their creation in Enterprise Architect. The best aspect of this way of implementing is that Enterprise Architect allows creating documentation templates that can be exported in formats such as RTF, PDF and HTML, which lets us generate a complete documentation of a methodology including all the defined characteristics: phases, roles, tasks, work products, correlations and alignments with international standards.

As a result, we have a structure based on the engineering profiles of ISO/IEC 29110-5 [34]; a tool that has a catalog of correlated frameworks from lifecycle phases, roles, tasks and work products; and the possibility of generating documentation of a methodology from the same tool, based on selected frameworks and activities, complying with the framework and taxonomies of ISO/IEC 29110-2 [34] and generating a specification of a custom profile with tasks and work products.

Advantages of this aproach are that software development teams can use the tools of the model to define and document their own methodology and accelerate the adoption of agile and traditional frameworks. They can performe methodical replacements between activities that seek the same goal, and correlating activities of different frameworks. They must choose the frameworks to be applied, identify the correlations between the frameworks activities and select the activities to be carried out.

Community members will have the possibility to add frameworks to the tool, mapping them to the central axis of the model, correlating the model variables and generating the required documentation of tasks and work products.

5 Conclusions and future work

This work has presented a study that allows corroborating the presence of methodologies that combine agile and traditional frameworks in the industry. Great advantages of combining agile and traditional practices were identified, at the same time of new challenges that are born due to the additional complexity in the definition of the methodologies that mix frameworks of different types. It was evidenced that this complexity can take us, among other aspects, to repetitive processes, repetitive work, lack of risk control, task execution through non-adequate roles or deficient definition of work teams.

The WYDIWYN -What You Define, Is What You Need- model was approached, focused on helping to define methodologies oriented to what a company needs, simplifying the identification, correlation and selection of phases, roles, tasks and work products among different frameworks. The model was subdivided in three big characteristics: a central axis that allows comparing different frameworks, some structuring variables that permit to correlate different frameworks, and a visual model that lets interact, catalog and personalize the way to explore it. A base structure for the definition of these characteristics was laid out, also its representation in Enterprise Architect as the chosen tool to implement the guide and its documentation from the taxonomy of the ISO/IEC 29110 standard, that allow the methodologies to be certified and compared with international standards.

Main advantage of WYDIWYN is having a centralized scheme that helps to compare, correlate and discriminate activities of different frameworks. It also highlights the fact of having a tool that allows incremental evolving the model, and to create personalized views of the information. Finally, the ability to generate documentation from templates and compare methodologies with international standards, bring a value for the company, letting it be able to define processes that can be internationally certified and compared.

As an opportunity for improvement, a low focus on risk and control activities it is identified, which although they are considered, could been raised as a model relevant structural part. In addition, the dependency of a licensed tool, as Enterprise Architect, may adversely affect the guide acceptance and willingness of the community members to contribute.

As future work it is considered to add iteratively agile and traditional frameworks, while analyzing and refining central axis phases, roles and team size. We plan to evaluate a first phase of the refined result in some software development companies, waiting for a greater maturity of the current prototype. A model contribution explanatory guide will be created, so that software development community members can complement it. In addition, a centralized and versioned model for consultation and contribution will be laid.

References

1. H. A. Hornstein, «The integration of project management and organizational change management is now a necessity,» International Journal of Project Management, vol. 33, nº 2, pp. 291-298, 2014.
2. M. Špunda, «Mixed Agile/Traditional Project Management Methodology – Reality or Illusion?,» Procedia - Social and Behavioral Sciences, vol. 119, pp. 939-948, 2014.
3. F. Pino, F. García y M. Piattini, «Software process improvement in small and medium software enterprises: a systematic review,» Software Quality Journal, vol. 16, nº 2, pp. 237-261, 2008.
4. S. Nerur, R. Mahapatra y G. Mangalaraj, «Challenges of migrating to agile methodologies,» Communications of the ACM, vol. 48, nº 5, pp. 72-78, 2005.
5. Rational Software, Rational Unified Process, Best Practices for Software Development Teams, Cupertino: Rational Software White Paper, 1998.
6. C. P. Team, CMMI for Development, Version 1.3, Software Engineering Institute , 2010.
7. I. c. office, ISO 9001:2008 Quality management systems -- Requirements, Suiza, 2008.
8. J. Singh, M. Sharma, S. Srivastava y B. Bhusan, «TSP (Team Software Process),» International Journal of Innovative Research and Development, vol. 2, nº 5, 2013.
9. A. Cockburn y J. Highsmith, «Agile software development: The people factor,» Computer, vol. 34, nº 11, pp. 131-133, 2001.
10. K. Schwaber, Agile project management with Scrum, Microsoft Press, 2004.
11. R. Polk, «Agile and Kanban in Coordination.,» de Agile Conference, 2011.
12. A. Janes y G. Succi, Lean Software Development in Action, Springer-Verlag Berlin Heidelberg, 2014.
13. C. J. Torrecilla Salinas, M. J. Escalona y M. Mejías, «A scrum-based approach to CMMI maturity level 2 in web development environments,» de Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services, New York, NY, USA, 2012.
14. N. N. Tuan y H. Q. Thang, «Combining maturity with agility: lessons learnt from a case study,» Proceedings of the Fourth Symposium on Information and Communication Technology, pp. 267-274, 2013.
15. P. R. Pinheiro, T. C. Sampaio Machado y I. Tamanini, «Dealing the Selection of Project Management through Hybrid Model of Verbal Decision Analysis,» Procedia Computer Science, vol. 17, pp. 332-339, 2013.
16. A. C. Pasini, S. Esponda, M. Boracchia y P. M. Pesado, «Q-Scrum: una fusión de Scrum y el estándar ISO/IEC 29110,» de XVIII Congreso Argentino de Ciencias de la Computación, 2013.
17. P. Monteiro, P. Borges, R. J. Machado y P. Ribeiro, «A reduced set of RUP roles to small software development teams,» Proceedings of the International Conference on Software and System Process, pp. 190-199, 2012.
18. S. Galvan, M. Mora, R. O'Connor, F. Acosta y F. Alvarez, «A Compliance Analysis of Agile Methodologies with the ISO/IEC 29110 Project Management Process,» de Conference on ENTERprise Information Systems / International Conference on Project, 2015.

19. J. Biolchini, P. G. Mian, A. C. Natali y G. H. Travassos, «Systematic review in software engineering,» System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES, vol. 679, nº 05, p. 45, 2005.
20. A. F. Bustamante, J. A. Hincapié y G. Gasca-Hurtado, «Structure of a Multi-model Catalog for Software Projects Management Including Agile and Traditional Practices,» de Trends and Applications in Software Engineering, Proceedings of the 4th International Conference on Software Process Improvement CIMPS'2015, vol. 405, Sinaloa, Mazatlán: Springer International Publishing, 2015, pp. 87-97.
21. J. A. Hincapié, G. P. Gasca-Hurtado y A. F. Bustamante, «Multimodel catalogue heuristics for software project managemet,» de 2016 11th Iberian Conference on Information Systems and Technologies (CISTI), Las Palmas, Spain, 2016.
22. A. W. Brown, S. Ambler y W. Royce, «Agility at scale: economic governance, measured improvement, and disciplined delivery,» Proceedings of the 2013 International Conference on Software Engineering, pp. 873-881, 2013.
23. R. Mantovani Fontana, V. Meyer Jr, S. Reinehr y A. Malucelli, «Progressive Outcomes: A framework for maturing in agile software development,» Journal of Systems and Software, vol. 102, pp. 88-108, 2015.
24. J. D. Yepes González, C. J. Pardo Calvache y O. S. Gómez, «Estado del arte de la utilización de metodologías ágiles y otros modelos en pymes de software,» de VII Taller Internacional de Calidad en las Tecnologías de la Información y las Comunicaciones", La Habana, Cuba, 2016.
25. K. M. Whitney y C. B. Daniels, «The Root Cause of Failure in Complex IT Projects: Complexity Itself,» Procedia Computer Science, vol. 20, pp. 325-330, 2013.
26. P.-W. Ng, «Theory based software engineering with the SEMAT kernel: preliminary investigation and experiences,» Proceedings of the 3rd SEMAT Workshop on General Theories of Software Engineering, pp. 13-20, 2014.
27. S. Galván-Cruz, M. Mora, R. O'Connor, F. Acosta-Escalante y F. Alvarez, «On Project Management Process in Agile Systems Development Methodologies and the ISO/IEC 29110 Standard (Entry Profile),» de International Conference on Informatics and Computing (CNCIIC-ANIEI), Mexico, 2014.
28. A. Friis Sommer, I. Dukovska-Popovska y K. Steger-Jensen, «Barriers towards integrated product development — Challenges from a holistic project management perspective,» International Journal of Project Management, vol. 32, nº 6, pp. 970-982, 2014.
29. L. Buglione, «Light maturity models (LMM): an Agile application,» Proceedings of the 12th International Conference on Product Focused Software Development and Process Improvement, pp. 57-61, 2011.
30. M. Van Hilst y E. B. Fernandez, «A pattern system of underlying theories for process improvement,» Proceedings of the 17th Conference on Pattern Languages of Programs, nº 8, 2010.
31. G. Verbong, R. Mourik y R. Raven, «Towards integration of methodologies for assessing and promoting the societal embedding of energy innovations,» de ASRELEO Conference, 2006.
32. J. Hea, C.-K. Baoa, T.-F. Shub, X.-X. Yuna, D. Jiangc y L. Brwond, «Framework for integration of urban planning, strategic environmental assessment and ecological planning for urban sustainability within the context of China,» Environmental Impact Assessment Review, A Review of Practice and Prospects for SEA in China, vol. 31, nº 6, pp. 549-560, 2011.
33. J. P. Rios Álvarez, Modelo para la armonización bajo la estrategia de multisourcing apoyado en las buenas prácticas de la industria del software, Medellín: Universidad EAFIT, 2016.
34. B. S. Institution, PD ISO/IEC TR 29110-1:2016, Systems and software engineering — Lifecycle profiles for Very Small Entities (VSEs), UK: BSI Standards Publication, 2016.

Project Portfolio Management in Small Context in Software Industry: A Systematic Literature Review

Juan Linares[1], Karin Melendez [2], Luis Flores[2], Abraham Dávila[2]

[1]Escuela de PostGrado, Universidad Nacional Mayo de San Marcos, Lima, Perú
[2]Departamento de Ingeniería, Pontificia Universidad Católica del Perú, Lima 32, Lima, Perú

Juan.linares@unmsm.edu.pe, {kmelendez, luis.flores, abraham.davila}@pucp.edu.pe

Abstract. In recent years, the software industry in Latin America, mostly consisted of small companies, has shown a notable increase in its operations. However, the practices they use do not allow them to achieve sustainable growth over time. In particular, one aspect that stands out is the disorder that presents inside the company when managing multiple projects with a very small human team. This paper seeks to identify experiences of application of models related to portfolio management in small contexts in the software industry. The study is carried out through a systematic literature review in relevant digital repositories. We have found 60 primary studies obtained in the search in the considered databases. Primary studies show that project portfolio management models present difficulties in their adoption in small contexts. In addition, there are few studies that refer to experiences of application of portfolio management models in small context in the software industry and point out that a relevant aspect for non-adoption is the lack of understanding of the models in their original proposal.

Keywords: Project portfolio management, project management, software development, small entity.

1 Introduction

Project portfolio management (PPM) is an emerging aspect of business management, focusing on how to select, prioritize, integrate, manage and control projects and their resources in the context of multiple projects that exist in organizations [1]. The Project Management Institute (PMI) [9] also consider the identification, prioritization, authorization, administration and control of programs, projects and other activities or work, allowing the achievement of the organizational strategic objectives. In particular, small organizations have the challenge of managing multiple small-scale projects [2], with limited resources [3], and therefore need to have simple but useful tools for efficient project management [4].

PPM is relevant in small software organizations, when software staff members have multiple roles and responsibilities, and they function simultaneously in different types of activities [5]. The research about PPM in a small company is incipient, even more,

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_5

considering that the nature of this type of project portfolio is different from the traditional PPM [5].

In this context, there have been some initiatives in Information Technology (IT) fields that incorporate the PPM for small organizations. These include: Competisoft [6], MPS.BR [7], Moprosoft [8] and adaptations of the PMI Portfolio Management Standard [9]. However, these initiatives [6] [7] [8] are geared towards various software processes, one of which is the PPM.

From the literature, we found some studies such as [10] and [15] that justify the need of PPM for small companies, also surveys to determine problems in PPM in small companies [11] and their impacts in the implementation of PPM [12]; however, no practical experience is reported. Other studies consider the portfolio processes in a separately way and not as an integral PPM. Some of these studies include: project selection and resource allocation [13], a project selection model in small and medium enterprises [14] an approach for project selection [15]. In addition, considering the scarce resources of a small business, in [16] a multi-criteria heuristic approach is proposed to improve the allocation of resources in multi-project scheduling, and in [17] a skills-based framework for optimal selection of software projects and the allocation of resources is described.

In other industries, different than software industry, approaches have been proposed for the management of multiple small projects [18], and portfolio management approaches [19], but have not yet been fully implemented. In the software industry it is emphasized that companies that use agile methods need to have agile project portfolios [20], because the management of these portfolios should be adapted to the agile processes [21].

After all, studying PPM in small contexts is a need that is concretized when organizations grow in number of projects and not necessarily in resources. This paper presents a systematic literature review (SLR) on experiences of PPM in small software organizations (contexts); especially in context of recurrent software maintenance or a lot of simultaneous modification of a product in agreement to customer needs; and performing every maintenance request as a different project. Section 2, presents a background; Section 3 presents the RSL protocol; Section 4 shows the results and; Section 5, includes a discussions and future work.

Project portfolio management (PPM) is an emerging aspect of business management, focusing on how to select, prioritize, integrate, manage and control projects and their resources in the context of multiple projects that exist in organizations [1]. The Project Management Institute (PMI) [9] also consider the identification, prioritization, authorization, administration and control of programs, projects and other activities or work, allowing the achievement of the organizational strategic objectives. In particular, small organizations have the challenge of managing multiple small-scale projects [2], with limited resources [3], and therefore need to have simple but useful tools for efficient project management [4].

PPM is relevant in small software organizations, when software staff members have multiple roles and responsibilities, and they function simultaneously in different types of activities [5]. The research about PPM in a small company is incipient, even more, considering that the nature of this type of portfolio is different from the traditional portfolio management [5].

In this context, there have been some initiatives in Information Technology (IT) fields that incorporate the PPM for small organizations. These include: Competisoft [6], MPS.BR [7], Moprosoft [8] and adaptations of the PMI Portfolio Management Standard [9]. However, these initiatives [6] [7] [8] are geared towards various software processes, one of which is the PPM.

From the literature, we found some studies such as [10] and [15] that justify the need of PPM for small companies, also surveys to determine problems in PPM in small companies [11] and their impacts in the implementation of PPM [12]; however, no practical experience is reported. Other studies consider the portfolio processes in a separately way and not as an integral PPM. Some of these studies include: project selection and resource allocation [13], a project selection model in small and medium enterprises [14] an approach for project selection [15]. In addition, considering the scarce resources of a small business, in [16] a multi-criteria heuristic approach is proposed to improve the allocation of resources in multi-project scheduling, and in [17] a skills-based framework for optimal selection of software projects and the allocation of resources is described.

In other industries, different than software industry, approaches have been proposed for the management of multiple small projects [18], and portfolio management approaches [19], but have not yet been fully implemented. In the software industry it is emphasized that companies that use agile methods need to have agile portfolios [20], because the management of these portfolios should be adapted to the agility of agile processes [21].

After all, studying PPM in small contexts is a need that is concretized when organizations grow in number of projects and not necessarily in resources. This paper presents a systematic literature review (SLR) on experiences of PPM in small software organizations (contexts); especially in context of recurrent software maintenance or a lot of simultaneous modification of a product in agreement to customer needs; and performing every maintenance request as a different project. Section 2, presents a background; Section 3 presents the RSL protocol; Section 4 shows the results and; Section 5, includes a discussions and future work.

2 Background

In a preliminary literature review, we found studies that justify the PPM in small companies [10], [5], as well as studies in project selection and resource allocation in multiple concurrent projects [22], based mainly on mathematical models [14]. However, there is a lack of empirical evaluation in PPM, due in part to the cost involved in the formal adoption of this process in small organizations [14].

From the previous studies, we can understand the need for small companies (or organizations in general) that need to manage their project portfolio and that this is a particular situation, especially because of the lack of resources that small organizations have. In this sense, it is really necessary to identify the experiences of PPM in small organizations (or in small contexts), as this will allow identifying problems, practices and solutions in PPM that serve as a basis for defining new proposals for these contexts.

3 Research Protocol

The research protocol used in this study is an SLR based on the guidelines and lessons learned proposed by Kitchenham [36]. The processes of an SLR that consists of three phases: planning, conducting and documenting. The first and second phases are presented in detail below and the third phase is reported in the rest of the paper.

3.1 Planning SLR

Establish a research question. The research objective was to analyze the experiences of the PPM application in small organizations in the software industry. Research questions were established as:

- PI-1. What are the problems in the PPM?
- PI-2. What proposals, models or standards have been developed for PPM?
- PI-3. What are the requirements for the implementation of a PPM model?
- PI-4. What are the processes that usually begin in a PPM implementation?
- PI-5. What are the difficulties for the implementation of a PPM model?
- PI-6. What benefits are reported from the implementation of a PPM model?

Development of a SLR protocol. A set of activities was performed that facilitated the search in databases and article selection, according to the guidelines proposed by [25]. (1) Search string. In order to define the search string, we chose to use the PICOC strategy (from: Population, Intervention, Comparison, Output, Context), which is an adapted strategy from the medicine world to Software Engineering [25], [26]. Table 1 shows these keywords in accordance with the PICOC criteria. Table 2 shows the search strings defined from the PICOC strategy.

Table 1. Search string defined by PICOC strategy.

PICOC Element	Description
Population	PPM
Intervention	Experiences of application
Comparison	Not Apply
Outcome:	Empirical studies
Context:	Small contexts

The search in research repositories was complemented by a search in specialized repositories obtaining the ISO/IEC TR 29110-5-1-3 (a technical report of intermediate profile) referring to PPM [24] and the PMBOK (project management body of knowledge) document from PMI [9].

Selection criteria. The exclusion (EC) and inclusion (IC) criteria considered were:

- IC.1. Items that belong to indexed digital repositories will be accepted.

- IC.2. Articles from scientific journals or conference presentations will be accepted only if they are from specialized organizations on the subject of PPM
- IC.3. We will accept articles on experiences of application of PPM.
- IC.4. Articles written in English will be accepted.
- EC.1. Repetitive studies, only the last version of the article will be considered.
- EC.2. Preliminary studies, only the version of the article that is more complete will be considered.
- EC.3. Books, poster types, secondary and tertiary studies.
- EC.4. Studies whose title is not related to the issue of developing SLR.

Table 2. Search string

Element	Description
Population	(Project AND ("Portfolio Management" OR "PPM" OR "Portfolio" OR multi* OR simultan* OR interdependent* OR opm3 OR "Management of Portfolios" OR mop OR p3m3))
Intervention	(experi* OR empiri* OR practic* OR implementation OR adopt* OR adaptation* OR approach* OR perform* OR simulate* OR "Lesson Learned" OR stud*)
Comparison	Not Apply
Outcome	("Case study" OR "Case studies" OR "action research" OR survey* OR ethnographic)
Context	(vse OR sme OR ((small OR medium) AND (enterpri* OR indust* OR contex* OR organiza* OR busin* OR compan* OR Enti* OR setti*)))

Databases used. The articles obtained correspond to sources selected for their scientific relevance as:

- Scopus (http://www.scopus.com)
- EBSCO host (http://www.ebscohost.com/)
- ProQuest (http://www.proquest.com/)
- Elsevier ScienceDirect (www.sciencedirect.com)
- Thomson Reuters – Web of knowledge (www.webofknowledge.com)
- IEEE Xplore (http://ieeexplore.ieee.org/)
- ACM Digital Library (http://portal.acm.org)
- Emeral (http://www.emeraldinsight.com/)

Procedure for studies selection. The studies selection was performed using the defined inclusion and exclusion criteria. The stages of search and application of the criteria are presented in Table 3. After applying the selection procedure, a team of expert research validated the primary studies selection results.

Evaluation of quality of studies. A checklist was applied in order to assess studies quality, avoiding bias of the research and providing a quantitative comparison [25]. Table 4 presents the checklist adapted from [37]. Also, the authors [37] and [38] recommend assigning the following score: [Y]es = 1 [P]artially = 0.5 or [N]o = 0.

Table 3. Procedure for studies selection

Procedure	Criteria of Selection
Initial stage	EC.3, EC.4
First stage	EC.1, IC.2, IC.4, EC.2
Second stage	EC.4
Third stage	EC.3
Quarter stage	IC.3

Table 4. Quality checklist for studies

Id	Questions to evaluate the quality of the selected articles
Q1	Are the experiences of applying PPM clearly defined?
Q2	Is the research process clearly established?
Q3	Are work limitations clearly documented?
Q4	Do the findings of the study contribute to the academic community or industry?
Q5	Is the link between the objectives and conclusions of the study clearly established?

Strategy for data extraction. A data extraction form was designed to record the information that researcher obtains from primary studies, as noted in [26]. The form contains basic bibliographic information to identify the primary study, such as a unique identifier associated with the study, the year of publication, author of the article, title of the article, country where the study was conducted, and whether the article comes from a journal or conference [37]. Every additional item of the form was designed to respond to the research questions formulated, in order to reduce the likelihood of research bias [25]. The strategy for data extraction was based on the collection of data directly by the researcher in the extraction form designed, and then validated by the research team [27].

Data synthesis strategy. To answer the research questions we choose to conduct a narrative synthesis of primary studies [25], and the summary of data was addressed with research questions [28].

Validating of SLR protocol. The SLR protocol was reviewed in two stages, by one research in the first stage and by two experienced researcher in the second.

3.4 Conducting SLR

Obtaining results from the search. The search strings were executed in the selected databases. In order to obtain greater flexibility in the search strings, advanced search options were used for each database. Search refinement options were applied according to the defined inclusion criteria and to the syntax accepted by each digital database. Table 5 presents the variables that allow us to explain the conformation of the search strings for the databases.

Table 5. Search string elements

Element	Description
P = Population	(Project AND ("Portfolio Management" OR "PPM" OR "Portfolio" OR multi* OR simultan* OR interdependent* OR opm3 OR "Management of Portfolios" OR mop OR p3m3))
I = Intervention	(experi* OR empiri* OR practic* OR implementation OR adopt* OR adaptation* OR approach* OR perform* OR simulate* OR "Lesson Learned" OR stud*)
O= Outcome	("Case study" OR "Case studies" or "action research" OR survey* OR ethnographic)
C = Context	(VSE OR SME OR ((Small OR Medium) AND (Enterpri* OR Indust* OR Contex* OR organiza* OR busin* OR compan* OR Enti* OR Setti*))

Selecting primary studies. The studies found were consolidated into a single spreadsheet, which served to perform the selection procedure. The primary articles obtained in the previous process were evaluated following the checklist defined in Table 4. The result of this process is presented in Table 6.

Table 6. Extraction values for each iteration

Data Base	Selection Date	1st Stage	2nd stage	3rd Step	4th Stage	Primary Studies
Scopus	Ene-16	776	773	103	24	16
EBSCO	Ene-16	241	241	36	5	4
ProQuest	Ene-16	208	205	48	7	4
ScienceDirect	Ene-16	469	468	106	36	23
Web of Science	Ene-16	397	397	78	21	6
IEEE Xplore	Ene-16	546	540	85	8	3
ACM DL	Ene-16	744	735	53	5	1
Emerald	Ene-16	162	162	28	4	3
		3543	3521	537	110	60

Quality evaluation of the articles. The quality of the articles was evaluated and the results of which are shown in Appendix A.

Extracting relevant data. A data extraction form was used to record relevant information from each primary study that contributes to answering research questions [25]. Each primary study selected was read and its corresponding information was filled in the form. The phrase - "No information" was placed in the boxes for which no related information was found.

The data were collected in the source language of the article (see Appendix B); in this case these data extracted from the studies contributing to research questions PI-1, PI-2, PI-3, PI-4, PI-5 and PI-6.

Synthetize extracted data. This section provides an answer to the research questions raised and it is developed in Section 4.

4 Results Discussion

This section presents the answers to the research questions and their discussion.

4.1 Question PI-1. What are the problems in the PPM?

The project selection problem has been one of the strategic issues in project portfolio management [13] [83], since poor selection of projects may lead to mistaken decisions [27] regarding the strategic alignment of the enterprise. The study has found numerous techniques (see Table 7) ranging from weighted scoring [39] to complex mathematical solutions [29], [38]. Likewise, the problem of the allocation of human resources was scarce [33], so that the optimal distribution of resources must be taken into account [28] in order to use them efficiently [57], according to portfolio priorities of these concurrent projects [7]. Among others problems we can mention too many concurrent projects, excessive commitment in terms of workload and ineffective executive decision making [51].

4.2 Question PI-2. What proposals, models or standards have been developed for PPM?

It is noted that a greater number of articles do not require a specific management instrument as a method or a management framework. On the other hand of the methods founded, only two have their own name and are generally oriented toward the topic of project selection. Both, the methods and the frameworks founded are implemented in industries outside the field of information technology.

4.3 Question PI-3. What are the requirements for the implementation of the PPM model?

It was found (see Table 8) that the creativity and innovation capacity [15], the synergy of teamwork [9], and the process of human resources [2], are basic requirements for the implementation of a management tool in organizations [52] [51]. It is also important to have an operational integration plan that includes requirements [32], appointment of specific people to resolve conflicts in implementation [51]. On the other hand the leadership of these implementations has to be supported by the top management [9], as by their project managers [15] [26]. In this sense it can be said that people and processes are critical elements for their implementation [51].

Table 7. General list of problems in project portfolio management

Id	Description	Studies
P1	Financial resources shortage	S03, S08, S21, S28, S29, S30, S35, S39, S40, S41, S49, S50 S51, S58
P2	Assignment of human resources	S03, S08, S20,S23, S27, S28, S29,S38, S39, S40, S44, S45, S46, S50, S51, S57, S58
P3	Time shortage to meet expectations	S08, S21, S38
P4	Lack of training in managing multiple projects	S06, S07, S11, S12, S13, S19, S21, S41, S42,S48, S51,S56, S57,S51
P5	Inadequate technology and process brought from large projects	S03, S04, S11, S33, S47, S51, S53
P6	Lack of understanding of the nature of small projects	S06, S07, S11, S12, S01, S19, S21, S46, S51
P7	Lack of alignment of the projects with the objectives of the organization	S05, S09, S10, 13, S15, S16, S17, 26, S27, S37, S52, S55
P8	Little adaptation to emerging changes in the environment	S05, S09,S10,S13, S26,S36, S43, S54
P9	Visibility of projects in relation to day-to-day operations	S02, S09, S10, S25, S27, S34
P10	Lack of functional responsibility of the individuals	S09, S15, S16
P11	Poor communication in both vertical and lateral dimensions	S04, S09, S16, S17, S32, S37
P12	Maintain the motivation through multiple project teams	S02,S09, S60
P13	Inability to learn from previous projects, knowledge transfer	S09, S22,S25, , S59
P14	Loss of valuable information due to the temporary nature of the project	S09, S22, S25, S59
P15	The lack of timely information to enable intervention	S09, S22, S24, S25, S31, S32, S60

4.4 Question PI-4. What are the processes that usually begin the implementation of PPM?

Appendix C shows a general list of priority processes for the implementation of PPM models. The selection of projects (P1) and human resources management (P6) are the priority processes that are referenced in most of the reviewed articles; then in this order of priorities it is observed that financial resource management (P8) is also a priority process. It is observed that these priority processes are directly related to PPM.

4.5 Question PI-5. What are the difficulties for the PPM implementation?

Table 9 shows a general list of difficulties in the PPM implementation founded by the present study. The difficulty that most stands out for the implementation of a portfolio management tool is the lack of experience in small contexts by the human resources

(D3), as well as Lack of communication between senior management and operational areas (D1) Among other difficulties, can also be mentioned that a bad workload assignment makes negotiate priorities in day-to-day operations.

Table 8. General list of portfolio management implementation requirements

Id	Description	Studies
R1	Support from senior management	S01, S16, S30, S32, S33, S45, S46, S50, S51
R2	Implementation plan aligned with organizational goals and internal standards	S07, S16, S17, S18, S27, S51, S60
R3	The government should ensure the structuring of project management methodologies	S13, S21, S25, S30, S36, S54
R4	Establishment of project management offices (PMO)	S01, S13, S34
R5	Delegation of the authority of the Project Manager	S12, S28, S29, S49, S51, S52, S53, S56, S58
R6	Detailed knowledge of human resources processes, technology, management of multiple small projects	S02, S09, S12, S15, S26, S38, S40, S48, S51, S53, S55, S58
R7	Establishment of control levels in technical resources	S03, S07, S20, S29, S41, S50, S52, S55, S57
R8	Motivation of project team members, clearly defined roles	S11, S31, S42, S44, S51, S53, S56
R9	Experience in monitoring project portfolio management processes	S04,S08,S09, S28, S39, S51, S58
R10	Long-term balance between the integrity of the environment and the portfolio of projects	S22, S24, S35, S43, S47, S49, S59
R11	Integrate other methods for the development of new products	S05, S06, S10, S14, S17, S19, S31, S33

4.6 Question PI-6. What benefits are reported from PPM model implementation?

Appendix D shows a general listing of benefits after the implementation of PPM models for project portfolio founded by the present study. This list of implementation benefits will then be related to primary studies. The most common benefit after the implementation of a management tool is the better distribution of workload and improved communication (B7). There is also an improvement in the project selection and prioritization process (B2) after getting a more flexible structure.

5 Conclusion and future work

The project portfolio management, in spite of being important at the level of large industry has not been covered in its entirety in the field of small business. Sometimes PPM is relevant in small organizations only when the staff has multiple functions and responsibilities, and when they are simultaneously running different types of activities. In the information technology industry there are few articles that address the subject in depth, except for some that mention the selection of projects and the allocation of human resources.

- **Table 9.** General list of implementation difficulties in portfolio management

Id	Description	Studies
D1	Lack of knowledge of the organizational structure and its processes	S01, S10, S15 S22, S24, S26, S27, S34,S37, S46
D2	Inadequate scope definition	S04,S05,S17,S29, S31, S42, S48, S49, S51, S52,S60
D3	Lack of knowledge of resources and experience	S02,S08,S11,S12,S13,S14,S18, S23,S28,S44,S45, S50, S51,S54, S57, S58
D4	Balance between achieving goals and balancing the environment	S13, S17,S21, S30,S36,S43, S47, S49
D5	Too many concurrent projects, too much commitment in terms of workload	S08, S09, S20, S23, S42, S51
D6	Difficulty in defining responsibilities among stakeholders	S02, S21, S27, S38, S42, S51, S55, S56, S57
D7	The difficulty of forecasting trends in a market	S04,S05, S21, S39
D8	The difficulty of identifying technological possibilities	S06,S07, S19, S25, S33,S35,S41, S49, S59
D9	Lack of communication in the availability of information	S03, S14, S25,S31,S40,S55,S58
D10	Negotiate priority in day-to-day operations between projects that have common interfaces	S08,S14, S20, S46, S51
D11	Lack of communication between senior management and operational areas	S01,S03,S10,S16,S17,S22,S32, S39, S46, S52, S53, S55
D12	Poor distribution in the workload	S09, S20, S29, S23

In this sense, the main problem of PPM is the allocation of limited resources, about 54% of the articles describe this conflict, as well as shortages of financial resources. In addition, it was found that the time commitment to meet the expectations of stakeholders is critical; but in small organization they accept the projects because they have the subsistence risk, this covers 46%. The same percentage identified the problem of lack of knowledge and experience in the management of simultaneous projects in this type of organizations.

The most frequent proposals to deal with this management problem are four methods and five management frameworks, representing 30% and 38% of the total proposals respectively. We also found a methodology and three exploratory analysis studies using a survey tool. It should be noted that the studies found refer to specific topics such as: a) project selection and b) allocation of portfolio resources, still pending the study of the application of a specific method, model or framework for PPM in small organizations.

For the implementation of PPM proposals it is necessary to comply with certain requirements that favor it. One of these priority requirements for success is the existence of a plan for implementation that is aligned with internal organizational strategies and standards. 30% of the articles refer to this requirement. Another requirement that favors implementation also with 30% considered as a priority is to have detailed knowledge in human resources management for this type of organizations.

The processes with which PPM usually begins its implementation are project selection processes [20], and the process of human resource allocation [69], 46% of the primary articles of this systematic review begin the implementation process of PPM implementing in the first place these processes.

Among the difficulties in the implementation of portfolio management proposals is the lack of experience of team members in small contexts [44], with 30% of the articles referring to this difficulty, another difficulty is lack of knowledge in the organizational structure and its processes [20] with 23% of the articles referring to this difficulty for implementation.

After the implementation of the proposals described in the articles, the benefits were: (i) a better distribution of the workload favored by an improvement in internal communication with 46% of articles that refer to this benefit [70]; (ii) Improvement of the selection and prioritization of projects to be aligned to organizational objectives [44] and 38% of articles refer to this benefit; and (iii) decrease in conflict index in human resources management [69] and 23% of the articles refer to this benefit.

Acknowledgements

This work has been carried out within the ProCalProSer project funded by Innóvate Perú under Contract 210 - FINCYT-IA-2013 and by the Department of Engineering and the Software Engineering Research and Development Group (GIDIS) of the Pontifical Catholic University of Peru.

References

[1] M. Young and K. Conboy, "Contemporary project portfolio management: Reflections on the development of an Australian Competency Standard for Project Portfolio Management," 2013.
[2] A. Singh, "Resource Constrained Multi-Project Scheduling with Priority Rules & Analytic Hierarchy Process," Procedia Engineering, pp. 725 – 734, 2014.
[3] D. Verdugo and E. Salazar, Modelo de administración de proyectos en pymes. Chile, 2012, vol. 11.
[4] A. Mesquida, Un Modelo para Facilitar la Integración de Estándares de Gestión de TI en Entornos Maduros. Palma- España, 2012.
[5] J. Vähäniitty, K. Rautiainen, and C. Lassenius, "Small software organizations need explicit project portfolio management," IBM Journal of Research and Development, 2010.
[6] Competisoft, Mejora de procesos para pequeña y media industria del software: CYTED, 2008.
[7] Softex, Mejora de Procesos del Software Brasileno Guía General. Brasil: Sociedad SOFTEX, 2006.
[8] H. Oktaba and otros, "Modelo de procesos para la industria del software (Moprosoft)," Brasil, 2005.
[9] Project Management Institute, A Guide to the Project Management, 5th ed., USA:, 2013.

[10] J. Vähäniitty, "Do Small Software Companies Need Portfolio Management,Too?," Helsinki University of Technology, Finlandia, Tesis Grado 2006.

[11] UC Ihesiene, "A survey-based study of project management problems in small and medium scale enterprises (SMES) in Nigeria," European Scientific Journal, 2014.

[12] M. Kljajic Borstnar and A. Pucihar, "Impacts of the Implementation of a Project Management Information System - a Case Study of a Small R&; D Company," Organizacija, 2014.

[13] N. Archer and F. Ghasemzadeh, "An integrated framework for project portfolio selection," International Journal of Project Management, pp. 207-216, 1999.

[14] C. Lawson, P. Longhurst, and P. Ivey, "The Application of a New Research and Development Project Selection Model in SMEs," Technovation, pp. 242-250, 2006.

[15] I. Gabarret. "Why are small public incubators 'lagging behind'?," International Journal of Entrepreneurship and Small Business, 2014.

[16] A. Lova, C. Maroto, and P. Tormos, "A multicriteria heuristic method to improve resource allocation in," European Journal of Operational Research, pp. 408-424, 2000.

[17] O. Maj, Y. Ali, and Z. Fadi, "Skill-based framework for optimal software project selection and resource allocation", European Journal of Operational Research, pp. 308–318, 2014.

[18] A. Ilincuta and G. Jergeas, "A practical approach to managing multiple small projects 2003.

[19] A. Von Ahsen and M. Heesen, "Innovation portfolio management: a framework for SMEs in the automotive industry," International Journal of Technology Intelligence and Planning, 2009.

[20] K. Rautiainen, J. von Schantz, and J. Vähäniitty, "Supporting Scaling Agile with Portfolio Management: Case Paf.com," in Proceedings of the 44th Hawaii International Conference, 2011.

[21] C. Hörzb, J. Stettina, "Agile portfolio management: An empirical perspective on the practice in use," International Journal of Project Management, pp. 140-152, 2015.

[22] B.S., Eskerod, P. Blichfeldt, "Project portfolios—there's more to it than what management enacts," International Journal of Project Management, pp. 357–365, 2008.

[23] ISO/IEC 29110, Software engineering-Lifecycle profiles for Very Small Entities. Genova: International Organization for Standardization, 2011.

[24] B.A. Kitchenham and S. Charters, Guidelines for performing systematic literature reviews in software engineering, 2007.

[25] C. Mamédio da Costa Santos, "A Estrategia PICO para a construcao da pregunta de pesquisa en busca de evidÊncias", Latino-am Enfermagem, 2007.

[26] B. Williams and J. Carver, "Characterizing software architecture changes: A systematic review," Information and Software Technology, pp. 31–51, 2010.

[27] P. Brereton, "Lessons from applying the systematic literature review process within the software engineering domain," The Journal of Systems and Software, pp. 571–583, 2007.

[28] P.D. Allan, "A Portfolio Management Approach to Assessing Acquisition and Divestiture Candidates," Society of Petroleum Engineers, 2001.

[29] M. G. Oliveira and H. Rozenfeld, "Integrating technology roadmapping and portfolio management at the front-end of new product development," Technological Forecasting and Social Change, 2010.

[30] A. Frini and S. BenAmor, "A TOPSIS multi-criteria multi-period approach for selecting projects in sustainable development context," 2015.

[31] L. Dooley, G. Lupton, and D. O'Sullivan, "Multiple project management: a modern competitive necessity," Journal of Manufacturing Technology Management, 2005.

[32] M. Attalla, T. Hegazy, and E. Elbeltagi, "In-House Delivery of Multiple-Small Reconstruction Projects," Journal of Management in Engineering, 2004.

[33] M. Rungi, "Interdependency management in Project Portfolio Management: How to implement required procedures," Technology Management for Global Economic Growth (PICMET)., 2010.
[34] C. Beringer,and H. Gemuenden, "Establishing Project Portfolio Management: An Exploratory Analysis of the Influence of Internal Stakeholders' Interactions," Project Management Journal, 2012.
[35] J. Vähäniitty, "Towards Agile Product and Portfolio Management," Department of Computer Science and Engineering, Aalto University, Finland, DOCTORAL DISSERTATIONS 2011.
[36] B. Kitchenham and Brereton O., "Systematic literature reviews in software engineering – A systematic literature review," Information and Software Technology, vol. 51, pp. 7-15, 2009.
[37] B. Kitchenham, E. Mendes, G. H. Travassos, "A Systematic Review of Cross- vs. Within-Company Cost Estimation Studies", IEEE Trans on SE, vol. 33, no. 5, pp. 316-329, 2007.
[38] B. Rouhani and et al., "A systematic literature review on Enterprise Architecture Implementation Methodologies," Information and Software Technology, pp. 1–20, 2015.
[39] G. Widforss, The Project Office as Project Management Support in Complex Environments,Procedia Computer Science, vol. 64, pp. 764-770, 2015.
[40] Kumaraswamy, M. M. (2005). Constructing relationally integrated teams. Journal of Construction Engineering and Management .
[41] Canavesio, M. M. (2007). Enterprise modeling of a project-oriented fractal company for SMEs networking. Computers in Industry , 794-813.
[42] Small, M. H. (2006). Justifying investment in advanced manufacturing technology: a portfolio analysis. Industrial Management & Data Systems , 485-508.
[43] Hocova, P. (2009). Design and management of an innovative software enterprise: A case study of a spin-off from University. Management of Engineering & Technology, 2009. PICMET 2009. p2788
[44] Amoroso, D. (1998). Developing a model to understand reengineering project success. System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conference , 500-509.
[45] Deshpande, S. (2011). Teaching students software engineering practices for micro-teams. Frontiers in Education Conference (FIE), 2011, (págs. T1H-1). USA.
[46] Kirkham, L. (2014). Prioritisation of operations improvement projects in the European manufacturing industry. International Journal of Production Research , 5323-5345.
[47] Müller, R. (2016). A framework for governance of projects: Governmentality, governance structure and projectification. International Journal of Project Management , 957-969.
[48] Tavana, M. (2015). A fuzzy hybrid project portfolio selection method using Data Envelopment Analysis, TOPSIS and Integer Programming. Expert Systems with Applications , 8432-8444.
[49] Conforto, E. C. (2016). Agile project management and stage-gate model—A hybrid framework for technology-based companies. Journal of Engineering and Technology Management , 1-14.
[50] Da Silva, C. G. (2016). An improved visualization-based approach for project portfolio selection. Computers in Human Behavior .
[51] Terlizzi, M. A. (2016). Barriers to the use of an IT Project Management Methodology in a large financial institution. International Journal of Project Management, 467-479.
[52] Mitrega, M. (2015). Business relationship process management as company dynamic capability improving relationship portfolio. Industrial Marketing Management, 193-203.
[53] Ponsteen, A. (2015). Classification of Human- and Automated Resource Allocation Approaches in Multi-Project Management. Procedia - Social and Behavioral Sciences, 165-173.

[54] Nikkhou, Shima. (2016). Designing a Portfolio Management Maturity Model (Elena). Procedia - Social and Behavioral Sciences, 318-325.

[55] Li, Yongkui. (2015). Developing a city-level multi-project management information system for Chinese urbanization. International Journal of Project Management, 510-527. vol. 33, no. 3

[56] Padovani, Marisa. (2016). Integrated PPM Process: Scale Development and Validation. International Journal of Project Management, 627-642. vol. 34, no. 4

[57] Patanakul, Peerasit. (2015). Key attributes of effectiveness in managing project portfolio. International Journal of Project Management, 1084-1097. vol. 33, no.5

[58] Beşikci, Umut. (2015). Multi-mode resource constrained multi-project scheduling and resource portfolio problem. European Journal of Operational Research , 22-31. vol. 240, no.1

[59] Latipova, Alina T. (2015). On optimization of R&D Project Selection and SchedulingIFAC-PapersOnLine , 6-10. vol. 48, no.25

[60] Gökgöz, Fazıl. (2017). Portfolio optimization under lower partial moments in emerging electricity markets: Evidence from Turkey. Renewable and Sustainable Energy Reviews, 437-449. vol. 67

[61] Jugend, Daniel. (2016). Product portfolio management and performance: Evidence from a survey of innovative Brazilian companies. Journal of Business Research, 5095-5100. vol. 63, no.11

[62] Tolonen, Arto. (2015). Product portfolio management – Targets and key performance indicators for product portfolio renewal over life cycle. International Journal of Production Economics. 468-477.

[63] Santos, Vitor. (2016). A Review. International Journal of Production Economics. 1085-1094. vol. 10

[64] Comerio, y otros. (2015). Service portfolio management: A repository-based framework. 112-125.

[65] Arratia M.. (2016). Static R&D project portfolio selection in public organizations. 53-63.

[66] Kaiser, M., Arbi, F (2014). Successful project portfolio management beyond project selection techniques: Understanding the role of structural alignment. Project Managment, 126-139.

[67] Pajares Javier (2016). Technology Start-up Firms as a Portfolio of Projects: The Case of DIMA 3D. Procedia - Social and Behavioral Sciences. 59-66. vol. 226.

[68] Neves, Adriano José da Silva (2015). The Use of AHP for IT Project Priorization – A Case Study for Oil & Gas Company. Procedia Computer Science. 1097-1105. vol. 55.

[69] Muegge, S. (2005). An exploratory study of new product development at small university spin-offs. IEEE International Engineering Management Conference. 626-631. vol. II.

[70] Hajjdiab, H.. (2012). An industrial case study for Scrum adoption. Journal of Software. 237-242.

[71] Khalfan, M.M.A. (2007). Building trust in construction projects. Supply Chain Management. p385

[72] Licht, T.. (2006). Improving multiple project management using simulation. 4th International Industrial Simulation Conference 2006, ISC 2006. 191-198.

[73] Murtishaw, S. (2006). Methodological and practical considerations for developing multiproject baselines for electric power and cement industry projects for Global Change. 645-665. vol. 11, no.3

[74] Patterson, J. (2009). Plumbing the depths: Stories, e-portfolios: Pedagogy, ownership. 8th European Conference on eLearning 2009, ECEL 2009. 449-457.

[75] Ralston, F. (2010). Risk constrained contracting strategies of renewable portfolios. 2010 7th International Conference on the European Energy Market, EEM 2010.

[76] Reinsvold, C. (2008). Seeing the forest as well as the trees: Creating value with portfolio optimization. Proceedings - SPE Annual Technical Conference and Exhibition. 3955-3961. vol. 6

[77] Wörösch, M. (2012). Structuring requirements in a multi-project environment in the construction industry - Proceedings of the ASME Design Engineering Technical Conference. 443-454. vol. 5

[78] Paasivaara, M. (2009). Using Scrum in distributed agile development: a multiple case study. Proceedings - 2009 4th IEEE,ICGSE 2009. 195-204.

[79] Yu, Vincent F. (1999). Managing research collaborations as a portfolio of contracts: a risk reduction strategy by pharmaceutical firms. International Journal of Technology Management. 207-231.

[80] Prabhu, G. N. (2004). In-house delivery of multiple-small reconstruction projects. Journal of Management in Engineering. 25-31 vol. 18, no.3-4

[81] Street, Alexandre. (2009). Risk Constrained Portfolio Selection of Renewable Sources in Hydrothermal Electricity Markets. Ieee Transactions on Power Systems. 1136-1144 vol. 24, no.3

[82] Kent,David C. (2004). Understanding Construction Industry Experience and Attitudes toward Integrated Project Delivery.Journal of Construction Engineering and Management-Asce. p815

[83] Cooper R., Edgett S and Kleinschmindt E., "Portfolio management for new products", Perseus Books, New York, 1998.

Appendix:

https://drive.google.com/open?id=0B_K_llz5juqPQ0FSWkVkWDA0M0E

Practices for Addressing Environmental Sustainability through Requirements Processes

Gabriel Alberto García-Mireles and Héctor Antonio Villa-Martínez

Departamento de Matemáticas, Universidad de Sonora
Blvd. Encinas y Rosales s/n Col. Centro
83000 Hermosillo, Sonora, México
{mireles, hvilla}@mat.uson.mx

Abstract. Requirements activities performed in a software project are fundamental to address sustainable software systems and products. Although several techniques and frameworks have been proposed to describe sustainability requirements, there is a lack of process support for incorporating them in requirements centered processes. The objective of this paper is to understand the extent to which proposals to develop sustainability requirements can be mapped with the activities and tasks of the requirements processes that are part of ISO/IEC 12207. The set of reviewed papers, extracted from a previous systematic mapping study, was classified with respect to processes tasks. As a result, reviewed papers showed a trend to address tasks from the stakeholder requirements definition process, such as stakeholder identification, elicitation of requirements, and requirements analysis. However, activities and tasks within both the system requirements analysis process and software requirements analysis process need methodological support.

Keywords: Environmental sustainability, green software, software process, requirements process, sustainability requirements.

1 Introduction

Information Technology and, particularly, software can be a means to support the challenges that face sustainable development. It can be used to optimize industrial processes as well as to reduce energy and resource consumption. Green and sustainable software is defined as "software, whose impacts on economy, society, human beings, and environment that result from development, deployment, and usage of the software are minimal and/or which have a positive effect on sustainable development" [1]. Since this definition has a broad scope, in this work, the focus is on the environmental dimension of sustainability, also called green dimension [2], and the activities and tasks related to development, particularly to discover sustainability requirements.

Sustainability requirements contributes to sustainable development and they are deemed as nonfunctional requirements [3]. Thus, traditional methods in the requirements engineering field can be adapted to address them [4]. Requirements are considered, also, as "key leverage point" to support the identification of sustainability

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_6

goals [5]. Moreover, requirements engineers can bring out sustainability concerns during early stages of software development life cycle [6].

Although several researchers propose frameworks and methods for developing sustainability requirements, little is known about the way these proposals can support activities and tasks of the requirements processes described in the ISO/IEC 12207 [7]. Thus, this work focuses on mapping sustainability requirements proposals with stakeholder requirements definition process, the system requirements analysis process, and software requirements analysis process. Understanding the extent these proposals can be integrated within processes can help to develop guidelines targeted to practitioners.

The main contribution of this paper is the mapping of a set of techniques and practices for addressing sustainability requirements in the context of requirements processes. For practitioners, the list of techniques can be a starting point to improve their respective requirements processes. For researchers, the mapping output provides information about the extent to which proposals have addressed both sustainability requirements and processes' activities.

The remainder of this paper is structured as follows. Section 2 describes the relevant literature reviews about sustainability requirements and an overview of requirements processes included into ISO/IEC 12207. Section 3 depicts the main actions executed to map sustainability requirements proposals with requirements processes. Section 4 describes the main outcomes of the mapping, while Section 5 presents the discussion of these results. Finally, Section 6 presents conclusions and suggestions for further research.

2 Related Work

The first part of this section describes the approach used to classify practices, methods, techniques, or approaches described in current literature of sustainability requirements. The second part describes the structure of the requirements processes based on those described in ISO/IEC 12207.

2.1 Background about Sustainability Requirements Approaches

Two articles were found that review the extent to which software requirements professionals can use methods to address sustainability [8, 9]. The former provides an integrated view of the current approaches and techniques to address sustainability within requirements engineering (RE) practices. The RE activities considered were: feasibility study, stakeholder identification, requirements elicitation, analysis and negotiation, documentation, validation, and management.

In the second paper, authors [9] proposed a classification approach for environmental sustainability aware methods and frameworks. In this work [9], 16 papers, selected through a systematic mapping study (SMS), were classified by both requirements abstraction level and software requirements activities described in SWEBOK. These papers are focused on elicitation and analysis activities. However, none of these two review papers consider the way sustainability requirements methods can be mapped with software process models.

Standard processes models, such as ISO/IEC 12207 does not detail processes in terms of methods or procedures [7]. They should be determined by the software organization. Despite this fact, there is a lack of systematic approaches to represent and integrate methods that support quality attributes with current process models [10]. In the case of sustainability and software process, Lami et al. [11] proposed a sustainability engineering processes where an outcome is “techniques and methods for sustainability are applied”, but these techniques are not described.

2.2 Overview of the ISO/IEC 12207 Requirements Processes

ISO/IEC 12207 standard provides a set of life cycle processes, activities, and tasks for software development [7]. Given that software is part of a larger system, in this standard software is considered an integral part of a system. The standard does not prescribe a specific software life cycle model, development methodology, method, model, or technique. In this work, the focus is on the stakeholder requirements definition process, the system requirements analysis process, and the software requirements analysis process, that are called in this paper requirement processes. The description of these processes includes both activities and tasks.

The purpose of the stakeholder requirements definition process is to define the stakeholder requirements, considering the environment in which software will operate. The process has following activities: identify stakeholders (one task), identify requirements (5 tasks), evaluate requirements (1 task), agree requirements (three tasks), and record requirements (one task).

The systems requirements analysis process transforms stakeholder requirements into a set of system technical requirements. The main outcome is a set of functional and nonfunctional requirements that were analyzed for correctness and testability. It has two activities: specify system requirements (one task) and evaluate system requirements (one task). Finally, software requirements analysis process specifies the requirements for the software components of the system. The software requirements also are analyzed for correctness and testability and the interfaces are defined. It has only one activity which is named “software requirements analysis” (three tasks).

3 Method to Map Sustainability Requirements Techniques with Software Processes

The specific research question that this work addresses is: What methodological support exists to develop sustainability requirements in the context of tasks described into ISO/IEC 12207 requirements related processes?

To determine the extent to which sustainability requirements approaches support activities and tasks of requirements processes, 16 papers from a mapping study were obtained [9]. The search string included terms such as ‘sustainability’ and ‘requirements’ and it was executed in four databases (Scopus, Web of Science, IEEE and ACM) [9]. Inclusion criteria considered peer-reviewed papers published up to 2016 and written in English, addressing practices for developing sustainability requirements [9]. Extracting data from these papers were based on a template that contains the following fields: id, reference, tasks of the requirements processes, sustainability requirements (description and examples), main findings, issues and evidence of applying the method. When possible, verbatim data was extracted. The data extraction procedure was carried out by the first author and verified by the second one. The ISO/IEC/IEEE 29148 [12] was used for verifying the way processes tasks were interpreted.

Table 1. List of papers reviewed (adapted from [9]). P1 = Stakeholder requirements definition process. P2 = System requirements analysis process. P3 = Software requirements analysis process.

ID	Methodological proposal	Evidence	Processes addressed
s1 [13]	Adopt i* language	Case study	P1
s2 [4]	Adoption/Adaptation of several requirements techniques	Experience report	P1, P2, P3
s3 [6]	Sustainability NFR (SNFR) framework	Examples	P1, P2, P3
s4 [14]	Requirements engineering for sustainability (RE4S) framework	Case study	P1
s5 [5]	Extending requirements practices for addressing sustainability	Case study	P1
s6 [15]	Framework for modeling sustainability requirements	Case study	P1
s7 [16]	Sustainability analysis framework	Case study	P1
s8 [17]	Techniques for identifying requirements supporting positive enabling effects	Case study	P1
s9 [18]	Sustainability patterns	Student projects	P1
s10 [19]	Define green quality model and an energy profiling tools	Experiments	P1, P3
s11 [20]	Framework for sustainability profiling	Example	P1
s12 [21]	Generic sustainability model	Example	P1
s13 [22]	Techniques for identifying stakeholders	Case study	P1
s14 [23]	IMAGINE approach	Case study	P1
s15 [24]	Survey for prioritizing environmental sustainability factors	Survey	P1
s16 [25]	Analytic Network Process (ANP)	Survey	P1

Technical steps, practices, and checklists can support process activities [26]. In this work, methodological proposals were classified in two types: techniques or practices. The former refers to an explicit reference to a technique or a set of step to carry out a task. Techniques can include very simple walkthroughs, using checklists, using formal methods, among others [10]. On the other hand, a practice refers to an action relevant to requirements engineering activities (or tasks) without an explicit definition of steps. Indeed, performing a practice may require one or more techniques.

4 Results

Table 1 depicts the list of papers reviewed in this study. For each paper, the main proposal is showed as well as the type of evidence a paper describes. In addition, each paper is mapped to appropriate processes. All papers include techniques or practices that can support the tasks described in the stakeholder requirements process. Few papers describe methodological support for the other two processes (s2, s3, s10).

4.1 Stakeholder requirements definition process

Stakeholder requirements definition process is the most common process addressed by proposals for defining sustainability requirements. Almost all tasks are considered by, at least, one paper (Fig. 1). The most frequent tasks addressed by selected papers were as follows: identify stakeholders, elicit requirements, analyze requirements, and specify critical qualities. Techniques and practices for these tasks are presented in following paragraphs.

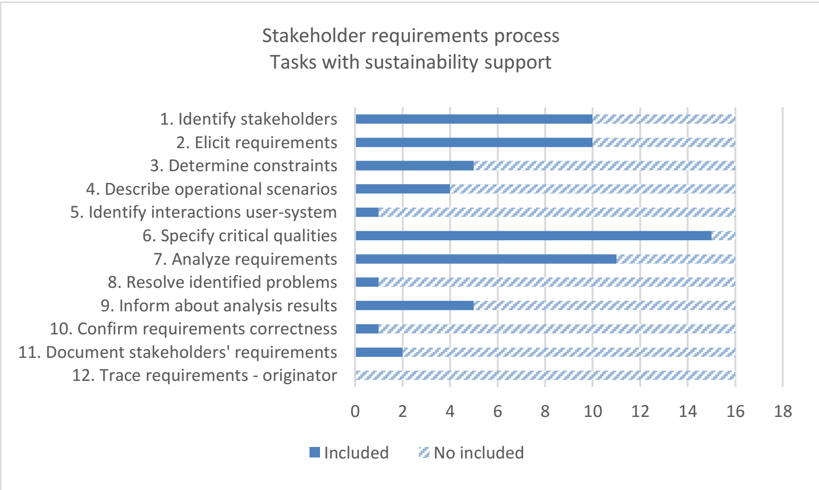


Fig 1. Techniques and practices for sustainability mapped into tasks from stakeholder definition process.

Identify stakeholders. Ten papers addressed this task, which discussed different techniques and practices to discover stakeholders related with sustainability (s1, s4, s5, s6, s7, s8, s13, s14, s15, s16). Including an expert in sustainability to understand the impact of system decision on environment is a common proposal, among others (Table 2). Others authors suggested to define a role for sustainability expert to take care of sustainability concerns. It should be assigned to a member of a project team (s2, s4). In addition, it is suggested that environmental and social sustainability dimensions require consulting with specialists of these fields (s5).

Table 2. Practices and techniques for supporting identify stakeholders (partial list).

Task	Practice	Technique
1.Identify Stakeholders	• Identify a sustainability expert within software project • Consult relevant experts in any of the sustainability dimensions • Aggregate a sustainability role in a stakeholder list	• Stakeholder impact analysis • Stakeholder model • Stakeholder matrix • Green onion model • 5-Dimension-based identification of stakeholders

Elicit requirements. A main topic on proposals is the way to elicit sustainability requirements (Table 3). The majority of methods addressed sustainability considering, at least, the environmental dimension. Other model-based proposals considered up to five sustainability dimensions and three order impact effect on sustainable development. In addition, some authors proposed to build taxonomies of sustainability goals related to sustainability dimensions as a starting point to determine sustainability requirements. Furthermore, the majority of evidence about eliciting sustainability requirements is depicted as goals and it is suggested to use participatory techniques.

Table 3. Practices and techniques for supporting elicit requirements (partial list).

Task	Practice	Technique
2 Elicit requirements from stakeholders	• Identify sustainability guidelines in the application domain • Use a sustainable goals catalog to derive sustainability requirements • Define sustainability goals by means of a model (e.g. GQM, framework, meta-model, requirements patterns) • Create a glossary of terms	• Taxonomy of sustainability goals based either on reduce, recycle, reuse approach or on sustainability dimensions • Workshops • Model based techniques: Sustainability NFR Framework, RE4S Framework, Sustainability analysis framework and others

Analyze requirements. Several approaches considered the way to depict distinct goals including those related to sustainability. Among them, modeling techniques (i* language, KAOS approach, goal modeling) allowed to represent high-level sustainability goals and the means to implement them considering different alternatives (s1) or the way a software system contributes to fulfils sustainability goals (s4, s2). SNFR framework (s3), framework for modeling sustainability requirements (s6), generic sustainability model (s12), and sustainability analysis framework (s7) categorize sustainability requirements with respect to economic, social, environment and technical dimensions. Some of these approaches allow describing different types of relationships between goals.

Other techniques determine a relative order among sustainability requirements and goals. Prioritization approaches use questionnaires to survey experts as regards important quality and sustainability factors (s15, s16). Moreover, indicators are defined to determine appropriate behavior of a system (s7, s14) and it is suggested that risk analysis considers internal and external risks as regards sustainability dimensions (s5).

Specify critical qualities. Sustainability requirements are treated as goals, softgoals, quality requirements, quality factors or only requirements. Some authors considered that sustainability is a complex and multifaceted notion (s1) that require discovering its meaning in software engineering field [27].

Several authors introduced the concept of sustainability dimensions to characterize sustainability in software. On the one hand, some authors characterized only the environmental dimension (s1, s2, s8, s10, s15, s16). On the other hand, other authors considered that sustainability dimensions are interrelated and they need to be addressed during requirements activities to carry out tradeoffs among sustainability goals (s5, s6, s4, s3, s7, s14).

4.2 System requirements analysis process

Few proposals addressed system requirements. As regards the task ‘specify system requirements’, Raturi et al. [6] (s3) described requirements considering sustainability dimensions and categorized them using a requirement facet model. By using the SNFR framework, they proposed a practice to describe systems requirements. For instance, an environmental sustainability requirement at second order effect is described as follows: “The system shall track hotel wide resource inputs and outputs that will be stored in an external resource tracking database in a standardized format.” Mahoux et al. [4] (s2), on the other hand, developed system models based on state diagrams and use cases. The task ‘evaluate system requirements’ is addressed by Mahaux et al. [4] who considered misuse case analysis to discover what system components can be harmful to the environment.

4.3 Software requirements analysis process

Software requirements analysis process has only one activity, ‘software requirements analysis’, composed of three tasks. The first task addresses the detailed specification of software requirements, including functional, nonfunctional, interface and data requirements, among others. Raturi et al. [6] proposed the sustainability NFR framework to classify requirements as function, data, performance, quality or constraint. The approach of Mahaux et al. [4] is adopting general requirements practices to consider sustainability concerns. For instance, satisfaction indicators are related with detailed requirements and misuses cases include mitigation actions.

As regards ‘evaluate software requirements’ task, Mahaux et al. [4] provided a practice to trace stakeholders’ requirements with detailed software requirements while Beghoura et al. [19] provided a tool for testing energy consumption of software to compare different software components. However, we do not find evidence related to the last task, ‘review software requirements’.

5 Discussion

As several authors mentioned, sustainability can be addressed in requirements engineering processes as other kind of nonfunctional requirement. As such, several papers presented traditional RE techniques used to discover sustainability requirements. Other proposals focused on characterizing sustainability as multi-dimensional notion and provided means to describe and analyze sustainability requirements. However, adopting or adapting traditional RE techniques face challenges related to the definition of sustainability concept and its dimensions in the context of software engineering [27]. In addition, there is a lack of techniques and tools to support quantitative measure of sustainability indicators. Supporting identification of sustainability requirements through quality models is in initial stage. Furthermore, environmental criteria are context dependent as regards software, hardware and usage scenarios (s1, s7, and s15). Therefore, it requires sound models to support appropriate identification of relevant factors influencing sustainability.

As regards the three requirements processes, the reviewed papers tend to address tasks belonging to the stakeholder requirements definition process. From 10 to 15 papers addressed the following four tasks: identify stakeholders, elicit requirements, analyze requirements and specify critical qualities. However, few papers considered tasks in both system requirements analysis process (2 papers) and software requirements analysis process (3 papers). Based on reviewed papers, approaches for addressing sustainability requirements are focused on early requirements while there is a lack of methodological support to address both how to specify sustainability software requirements and the way they can be derived from stakeholder sustainability requirements.

The main limitations in this work are the bias in selection of publications and classification of papers. The first threat was mitigated by using a set of papers identified by a SMS method [9]. Although the list of papers can serve for follow-on research activities [28], additional searching procedures were not executed, since the main goal of this work is understand the trends in developing methodological support for requirements processes. As regards the second bias, it was mitigated by using a template that included fields for the 17 tasks of the requirements processes. In addition, the second author verified the extraction data procedure. However, there is a potential misclassification of papers since some reported actions were interpreted as practices.

6 Conclusions

Addressing sustainability requirements in software projects need additional practices and techniques that both support an extended system scope and include additional stakeholders. Exploratory case studies showed that adapting traditional RE practices and techniques can be used to discover sustainability requirements. In addition, several models and frameworks define sustainability dimensions to support eliciting and analyzing sustainability requirements. As regards software processes, it was found a trend to propose methodological support for the stakeholder requirements definition process. However, there is a lack of methodological proposals for system requirements analysis process and software requirements analysis process.

As future work, it is possible to identify traditional RE techniques used to currently to perform requirements processes that can be adapted to develop sustainability requirements within the system and software requirements processes. In addition, the notion of sustainability requirements needs a sound analysis to understand the scope of each dimension and the feasibility of addressing the three-order impact on sustainable development. Since there is little empirical research in this field, it is suggested to conduct case studies in industrial settings to understand the impact of sustainability requirements techniques and practices on software requirements processes. Furthermore, validated practices to develop sustainability requirements should be adapted considering needs of small size software organizations.

References

1. Dick, M., Naumann, S.: Enhancing software engineering processes towards sustainable software product design. In: EnviroInfo, pp. 706-715. (2010)
2. Calero, C., Piattini, M.: Introduction to green in software engineering. In: Calero, C.Piattini, M., (eds.). Green in Software Engineering. pp. 3-27. (2015)
3. Penzenstadler, B., Raturi, A., Richardson, D., Tomlinson, B.: Safety, security, now sustainability: The nonfunctional requirement for the 21st century. IEEE Software. 31, 40-47 (2014)
4. Mahaux, M., Heymans, P., Saval, G.: Discovering sustainability requirements: An experience report. In: Berry, D.Franch, X., (eds.). Requirements Engineering: Foundation for Software Quality. REFSQ 2011. 6606 LNCS, pp. 19-33. Springer, Berlin, Heidelberg. (2011)
5. Becker, C., Betz, S., Chitchyan, R., Duboc, L., Easterbrook, S.M., Penzenstadler, B., Seyff, N., Venters, C.C.: Requirements: The key to sustainability. IEEE Software. 33, 56-65 (2016)
6. Raturi, A., Penzenstadler, B., Tomlinson, B., Richardson, D.: Developing a sustainability non-functional requirements framework. In: 3rd International Workshop on Green and Sustainable Software, GREENS 2014 - Proceedings, pp. 1-8. (2014)
7. ISO/IEC: Systems and software engineering -- Software life cycle processes - Redline. ISO/IEC 12207:2008(E) IEEE Std 12207-2008 - Redline1-195 (2008)
8. Chitchyan, R., Betz, S., Duboc, L., Penzenstadler, B., Ponsard, C., Venters, C.C.: Evidencing sustainability design through examples. In: CEUR Workshop Proceedings, pp. 45-54. (2015)
9. García-Mireles, G.A., Moraga, M.Á., García, F., Piattini, M.: A classification approach of sustainability aware requirements methods. In: 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1-6. (2017)
10. Chiam, Y.K., Staples, M., Ye, X., Zhu, L.: Applying a selection method to choose Quality Attribute Techniques. Information and Software Technology. 55, 1419-1436 (2013)
11. Lami, G., Fabbrini, F., Fusani, M.: Software sustainability from a process-centric perspective. In: Winkler, D., O'connor, R.V., Messnarz, R., (eds.). Systems, software and services process improvement. European Conference on Software Process Improvement. 301 CCIS, pp. 97-108. (2012)
12. ISO/IEC/IEEE International Standard - Systems and software engineering -- Life cycle processes --Requirements engineering. ISO/IEC/IEEE 29148:2011(E)1-94 (2011)

13. Cabot, J., Easterbrook, S., Horkoff, J., Lessard, L., Liaskos, S., Mazón, J.N.: Integrating sustainability in decision-making processes: A modelling strategy. In: 2009 31st International Conference on Software Engineering - Companion Volume, ICSE 2009, pp. 207-210. (2009)
14. Penzenstadler, B., Mehrabi, J., Richardson, D.J.: Supporting Physicians by RE4S: Evaluating Requirements Engineering for Sustainability in the Medical Domain. In: Proceedings - 4th International Workshop on Green and Sustainable Software, GREENS 2015, pp. 36-42. (2015)
15. Saputri, T.R.D., Lee, S.W.: Incorporating sustainability design in requirements engineering process: A preliminary study. In: Lee, S.Nakatani, T., (eds.). Requirements Engineering Toward Sustainable World. APRES 2016. 671 CCIS, pp. 53-67. Springer, Singapore. (2016)
16. Lago, P., Koçak, S.A., Crnkovic, I., Penzenstadler, B.: Framing sustainability as a property of software quality. Communications of the ACM. 58, 70-78 (2015)
17. Huber, M.Z., Hilty, L.M., Glinz, M.: Uncovering sustainability requirements: An exploratory case study in canteens. In: CEUR Workshop Proceedings, pp. 35-44. (2015)
18. Roher, K., Richardson, D.: A proposed recommender system for eliciting software sustainability requirements. In: 2013 2nd International Workshop on User Evaluations for Software Engineering Researchers, USER 2013 - Proceedings, pp. 16-19. (2013)
19. Beghoura, M.A., Boubetra, A., Boukerram, A.: Green software requirements and measurement: random decision forests-based software energy consumption profiling. Requirements Engineering. doi: 10.1007/s00766-015-0234-2, (2015)
20. Alharthi, A.D., Spichkova, M., Hamilton, M.: Sustainability profiling of long-living software systems. In: CEUR Workshop Proceedings, pp. 12-19. (2016)
21. Penzenstadler, B., Femmer, H.: A generic model for sustainability with process- and product-specific instances. In: GIBSE 2013 - Proceedings of the 2013 Workshop on Green in Software Engineering, Green by Software Engineering, pp. 3-7. (2013)
22. Penzenstadler, B., Femmer, H., Richardson, D.: Who is the advocate? Stakeholders for sustainability. In: 2013 2nd International Workshop on Green and Sustainable Software, GREENS 2013 - Proceedings, pp. 70-77. (2013)
23. Rodriguez, A., Penzenstadler, B.: An assessment technique for sustainability: Applying the IMAGINE approach to software systems. In: CEUR Workshop Proceedings, pp. 1-8. (2013)
24. Koçak, S.A., Alptekin, G.I., Bener, A.B.: Integrating environmental sustainability in software product quality. In: CEUR Workshop Proceedings, pp. 17-24. (2015)
25. Koçak, S.A., Alptekin, G.I., Bener, A.B.: Evaluation of software product quality attributes and environmental attributes using ANP decision framework. In: CEUR Workshop Proceedings, pp. 37-44. (2014)
26. Zhu, L., Jeffery, R., Staples, M., Huo, M., Tran, T.T.: Effects of architecture and technical development process on micro-process. In: Wang, Q., Pfahl, D., Raffo, D.M., (eds.). Software Process Dynamics and Agility. ICSP 2007. 4470 LNCS, pp. 49-60. Springer, Berlin, Heidelberg. (2007)
27. Naumann, S., Kern, E., Dick, M., Johann, T.: Sustainable software engineering: Process and quality models, life cycle, and social aspects. In: Hilty, L.M.Aesbischer, B., (eds.). ICT Innovations for Sustainability. 310, pp. 191-205. (2015)
28. Kitchenham, B.A., Budgen, D., Pearl Brereton, O.: Using mapping studies as the basis for further research - A participant-observer case study. Information and Software Technology. 53, 638-651 (2011)

Integrated IT Governance and Management Model: Evaluation in a Developing Country

Carlos Montenegro[1], Andrés de la Torre[1] and Natalí Nuñez[2]

[1] Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador
{cemontenegro, eadelatorre}@espe.edu.ec
[2] Universidad de las Américas, Quito, Ecuador
natinp@hotmail.com

Abstract. Following a Design Science Research (DSR) approach, this paper describes an integrated model of IT governance and management. The Model Design Phase is based on significant public and private sector IT studies contributions, on theoretical IT related models (BITA, BDIM, others), and on IT related professional frameworks (COBIT, ITIL, ISO 20000, others) that provide guidelines inspired by practice. The Model Evaluation Phase is done using three case studies with different contextual requirements and application limitations in Ecuador, a developing country. In any case, the experiences show the designed artifact operability and evidence similar implementation characteristics. The work provides a practical tool for practitioners and contributes to the existing body of knowledge, both about artifacts design and evaluation and on IT development in Private and Public Sector Organizations.

Keywords: IT governance; IT management; Public Sector; Private Sector; DSR

1 Introduction

IT Governance (ITG) defines and spreads the mechanisms required to ensure the current and future business technology alignment (BITA) objectives [1][2]; whereas IT Management (ITM) must ensure the governance mechanisms are in place implemented to fulfill the strategies [3].

Campbell *et al* identify relevant characteristics of ITG and ITM [4]. On the other hand, the BITA models [5] [6] [7] are based typically on the Strategic Alignment Model (SAM) [8]; one of SAM instances is the technological transformation alignment. A case in point can be found in COBIT 5, which states that the needs of each stakeholder must be transformed into a possible corporative strategy through its "goal cascade" process [9]. This process translates the needs into customized business goals, IT-related goals, and specific IT process.

Plenty of research has been done in the last decade about how to implement ITG in a structured and process-oriented way [10] [11]. ISO 38500 [12] has become the first international standard that depicts general directives about implementing ITG inside organizations; however, it does not include details about the required mechanisms, techniques or specific tools. De Haes *et al* [13], Van Grembergen *et al* [14], and

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_7

Almeida *et al* [15] recommend various ITG effective mechanisms, and De Haes *et al* [16] establishes the top ten mechanisms. Besides, it has also been suggested to select a set of structures, processes, and relationships of ITG; Almeida *et al* [17] suggests that these characteristics can be identified considering their appearance in the technical literature.

Concerning ITM, the Business-Driven Information Management model (BDIM) is the reference of models, practices, techniques, and tools for mapping and evaluate the interdependencies between business performance and IT. BDIM has proven its value for improving both IT service quality and business outcomes [18] [19] [20]. Currently, ITIL [21] and COBIT [9] are the most adopted frameworks on ITM.

Despite the comprehensive and detailed guidance available on IT frameworks, as above summarized, the integrated implementation of ITG and ITM have been recognized as challenging for many organizations [23] [24]. For the purpose, several studies examine the adoption of frameworks, such as COBIT or ITIL [16] [25] [26] [27] [28]. So, while many studies have looked into the adapting COBIT for a specific context [29] [30] [31], there is little research about the suitability of these customizations, much less in the public sector.

In the above subjects, the integration of Governance and Management is little studied. However, there is a background that can be used for the definition of an alternative model: existing IT standards and guidelines. The potential contributions of such integrated model can be significant to public or private sector organizations since it would facility to combine components of different models and best practices to model, complete and improve the IT processes and other aspects, as the ROI [22]. At the same time, the organizations can develop processes to deliver the capability of each technology in a way that enables work completion and the organizational objectives achievement.

2 Background and related work

For the integrated implementations of ITG and ITM, two models can be found in specialized literature: Business-Objectives driven IT Management (IT-MBO), and the Guidelines & Areas Model. The first is conceived around a set of key concepts: objectives, key performance indicators and BSC organizational perspectives [32] [33]. In this case, COBIT is a clear adaptation of this model. The second model, Guidelines & Areas [34] [35], has been reported as similar to COBIT 4.1. Since that COBIT 5 [9] covers the previous version, it can be considered as a complete framework for combining ITG and ITM [36].

Nevertheless, where a model implementation is required, doubts coexist regarding the difficulties that may come up; these uncertainties are reported for COBIT implementations [16]. Likewise, the results of the study by Debreceny & Gray [27] showed that only 16% of the surveyed organizations utilized the COBIT framework intensively, while the remainder opted to select a subset of processes that provided most of the desired IT governance benefits. Omari [37] synthesizes several previous studies referred to the public sector, and propose 13 COBIT 5 Processes as a customized referential framework, and as an option to manage the solution complexity.

On the other hand, from a services perspective, ITIL is a framework that describes best practice in IT service management (ITSM) [21]. ISO 20000 [38] is based on ITIL and adds a level of certainty and establishes a solid foundation for research on the benefits of ITSM techniques [39]. The auditability and the possibility of certification is a marked difference with ITIL so that the use of the ISO 20000 tends to IT industrialization [40]. On the other hand, according to Moura *et al* [18], COBIT and ITIL frameworks are not mutually exclusive and can be combined to offer more strong support particularly for IT Governance.

The above issues majority refer to private sector organizations (PO). Regard to the implementation of ITG and ITM in the public sector organizations (PSO), the academic contributions [41][42][43][4] have established contextual characteristics of IT; other studies propose some experiences [44][45][46]. According to Winkler [41] and others authors [4][43][47], the PSOs create more centralized structures for decision-making.

Finally, a wide range of process capability/maturity models have been developed, evolved and adapted to evaluate ITG and ITM. Some of those models are concentrated around the standard ISO/IEC 15504 [48] [49], for the process capability. The Capability levels can be applied to an organization process improvement in individual process areas [50]. COBIT 5 [9] adopts this model. The capability of each assessed process is expressed as a Level from 0 to 5, each one with ratings: Not achieved (0%-15%), Partially achieved (>15%-50%), Largely achieved (>50%-85%) or Fully achieved (>85%-100%). The model evaluates following attributes: Process Performance, Performance Management, Work Product Management, Process Definition, Process Deployment, Process Measurement, Process Control, Process Innovation, and Process Optimization [51].

3 Methodology

Design Science (DSR) is a research approach developed through the last decade [52] [53] [54]. It has been used in several domains: Information Systems [55], Business Processes Management [56] and IT Management [57], among others. According to Hevner *et al* [53], DSR constitutes a pragmatic research paradigm that encourages the creation of innovative artifacts to solve real-world problems. Thus, DSR combines a focus on IT artifacts with a high priority for its relevance in the application domain, and a set of application guidelines.

DSR follows a two-phase approach: Design and Evaluation. Besides, DSR provides a stage of feedback for the designed artifact [52] [55], which in this work is implemented by generating two models: a general and a refined model; the latter is particular to each case study. On evaluation, it is used the case study option. Besides, the qualitative method of Participant Observation is used to facilitate the feedback and gathering more detailed and accurate information [58] [59].

4 Design Phase: Integrated Model for ITG/ITM

According to the DSR Guidelines [52], the artifact is developed through a systematic search and consolidation of partial and complementary solutions, based on ITG/ITM requirements and contributions of cited base reference models. In this section the general model is presented (Table 1). The model components, or mechanisms, are comprised of processes, structures, relations, strategic alignment, best practices, and norms.

For the refined models definition (in the case studies), the general model mechanisms are applied as follows: i) COBIT Goal Cascade technique is applied to identify the ITG COBIT processes; ii) Whenever it is feasible, ITG best practices are adapted to meet the closest COBIT processes; else, the best practices are adopted; iii) It is adopted the COBIT processes related to ITM. According to the context, the processes can be added or deleted. It is convenient to adopt processes to comply a basic PDCA cycle, wich facilitate a possible improvement cycle.; iv) Optionally, it is defined a continuous improvement process, assessing the current processes capability. If gaps are identified, the improvements can be planed.

Table 1. General Integrated Model for ITG/ITM

			Public Sector Best Practices	***General Best Practices***	***IT-MBO***	***Guidelines & Areas***
ITG	Basic Framework	Structure	- IT Governance Committee - Centralized decision structure; or Hybrid structure organized as a duopolistic archetype - Liaison management	- IT Alignment Task Integration in roles and responsibilities. - Management Committee - IT organizational structure considerations. - CEO as a member of Executive Board. - CIO reporting to CEO - The Manager, as a role for business/IT relationships.	COBIT: - Governance Domain. ISO 38500 Capability Evaluation Model	**Guidelines:** - IT Objectives **Areas:** - Strategy
		Processes	- Service Q&A. - IT budgeting process - IT Project and Portfolio Management. - IT Strategic Planning.	- IT Strategic Planning. - IT Project Portfolio Management. - IT Project Follow-Up.		
		Relations	- The strength of relational networks within administration and IT stakeholders - The creation of shared knowledge among administration and IT stakeholders	- Business objectives with common understanding. - ITG awareness campaigns. - Shared management of IT accounts.		
	BITA	SAM Model	- Technological Transformation		COBIT: Goal Cascade.	**Guidelines:** - IT and Business Alignment. **Areas:** - Strategy

			Public Sector Best Practices	*General Best Practices*	*IT-MBO*	*Guidelines & Areas*
ITM	BDIM	A service-oriented model			COBIT: - Management Domains. ITIL: - Service Strategy (SS). - Service Design (SD) - Service Transition (ST) - Service Operation (SO) - Continuous Improvement (CI) ISO 20000 Capability Evaluation Model	**Guidelines:** - Business Structures/Processes/Mechanisms. - IT Processes. - IT Balanced Scorecard (BSC). - Best-practices. - Auditing. - Improvement. - Innovation. **Areas:** - Service. - Resources. - Risk. - Development. - Architecture. - Projects. - Support and Q&A. - Investment. - Outsourcing. - Compliance. - Improvement. - Innovation.
	Specific Norms	ISO 27000 ISO 22301 ISO 29110 Others.				

5 Evaluation Phase: Applying the Model in Cases Studies

The observational case study is a feasible option for this work, according to the DSR Guidelines [52]. Here, the Artifact Evaluation is done through three case studies in Ecuador, with different contextual requirements and application limitations. It should be noted that the evaluation applies to ITG/ITM mechanisms in operation, so that it is used with the objective of modeling ITG/ITM to improve these mechanisms and/or to complete those that do not exist. Details are summarized in Table 2. Other particulars can be reviewed in previous studies [60][61][62].

Table 2. Case Studies details

	CASE A	**CASE B**	**CASE C**
Objective	ITG/ITM modeling and implementation	ITG/ITM modeling and implementation	ITG/ITM modeling
Employees	3200	150	470
IT Unit	30 officials. The Unit manages IT services and a data center	20 officials. The Unit manages some outsourced services	15 officials. The Unit manages the strategic services outsourced
Design Phase outcome	ITG/ ITM model for the enterprise's network	ITG/ITM model for the institution	IT Outsourcing (ITO) model for the institution
Evaluation time	Six months	One year	Six months
Participant	Three researchers. One	Two researchers. One on	Two researchers. One on

Observation	on full time	full time	full time
Model refinement process	A combined strategic alignment scenarios between PSOs with of PDCA cycle compliance	Alternative strategic alignment scenarios.	The Model definition based on Contractual ITO services
Implementation complexity treatment criteria	Goal Cascade permits selecting COBIT and ISO 20000 processes for each enterprise. The final processes group is an accumulation of all enterprises	Goal Cascade permits selecting COBIT ITG/ITM processes. Complementary norms are used to detail the COBIT processes	Goal Cascade permits selecting COBIT ITG/ITM processes for ITO. Besides, ITIL is used to details the COBIT processes
Model Outcomes	The harmonization of 31 COBIT and 14 ISO 20000 Processes for ITG/ITM, covering six municipal companies	12 ITG/ITM Processes, and a group of Best Practices mechanisms for ITG	A reference framework composed of 16 ITG/ITM processes, 11 ITO Best Practices mechanisms, and an improving plan
Evaluation process and results	Improvement on Process Capability, over five critical processes. Average improvement: 8,3%, at Level 1 Capability	Model outcomes compliance. Implementation: all ITG mechanisms, and four processes	Compliance of Improvement Plans contents, and check using Critical Success Factors

According to Ecuadorian law [63], in the A case, the Metropolitan Public Enterprises are independent PSOs. The six selected enterprises are coordinated directly from the Town Hall [64]. Namely, they constitute an enterprise's network that, although do not share IT infrastructure, they have the same general objectives than the Metropolitan District, as well as share the financial resources. The selected enterprises have more size and infrastructure; besides, they manage a data center and do not depend on the central administration.

The less capability process in each COBIT domain was used to develop the model evaluation, through activities of an improvement plan. The plan activities included those more feasible to be executed in small time intervals; and, as far as possible, the execution risks were minimized. The activities were immersed in daily work; nevertheless, the results showed that the IT organizational context influenced for the partial implementation, and on the improvement values. The average of processes capability improvement was the 8.3% at Level 1 Capability, rating Partially; it is considered a low value.

An alignment scenario was selected in the B case, to apply the model to the Accreditation Council for Higher Education in Ecuador. Table 3 shows the summarized results using COBIT Goal Cascade, for the more likely scenario.

Table 3. Case B: IT Strategic Alignment

	Scenario		
	Corporate Goals focused on customer orientation and internal policy compliance, with emphasis on benefits and risk management optimization.		
COBIT Processes	**Priority**	**High (> 2 IT corporate goals)**	EDM03, APO13, BAI06, DSS03, DSS04, DSS05, MEA01, MEA02
		Medium (2 IT corporate goals)	EDM01, EDM02, APO01, APO02, APO08, APO09, APO10, APO012, BAI01, BAI02, BAI04, BAI10, DSS01, DSS02, DS06, MEA03
		Low (1 IT goal)	EDM05, APO03, APO05, APO07, BAI03

The high-priority processes are selected; they promote to three or more IT corporate goals. Next, considerations are made regarding corporative governance, along with the COBIT processes, so that definitive mechanisms can be established (Table 4).

Table 4. Case B: Mechanisms identified after applying the ITG/ITM model

			Mechanisms
ITG	**Basic Framework**	Structure	- ICT structure with the centralized decision, inherent of supporting public processes, according to the organization chart. - Creation of Technology Committee at Advisor Level, led by Chairman, or representative. - Integration of alignment ICT tasks in roles and responsibilities attached to the ICT Department or Technology Committee. - Presidency of ICT Department in the Technology Committee, reporting and accomplishing administrator roles of ICT business relationships.
		Processes	- COBIT Processes: EDM03, APO02, AP009, APO03
		Relationships	- A shared understanding of business objectives through the Technology Committee. - ITG Awareness Campaigns.
	BITA	SAM Model	- COBIT: Goal Cascade.
ITM	**BDIM**		- COBIT Processes: APO13, BAI06, BAI03, DSS03, DSS04, DSS05, MEA01, MEA02
	Specific Strategies and Norms		- ISO 27001 and ISO 27002 for APO13 and DSS05 - ISO 22301 and ISO 27001 for DSS03 and DSS04 - ISO 29110 and Agile Methods for BAI06

The C case includes a PO legally constituted and authorized to operate in Ecuador, in the group of medium banks of the commercial segment. Based on specific considerations, the model references a contractual mechanism of outsourcing, from a customer perspective; the ITG mechanisms and the processes are proper or adjusted to the IT Outsourcing (ITO) services. Table 5 summarizes the model components.

Table 5. Case C: Mechanisms identified after applying the ITG/ITM model

			Mechanisms
ITG	**Basic Framework**	Structure	- Strategic level: Executive Management Committee and ITO Management Office - Tactical level: Architecture Committee and Executive of ITO Projects - Operational level: Responsible for Business Units/IT
		Processes	- ITO Project Portfolio Management - ITO Project tracking - Acquisitions - Quality of service - Information Architecture Management
		Relationships	- RACI Matrix
	BITA	SAM Model	- COBIT: Goal Cascade.
ITM	**BDIM**		- COBIT Processes: EDM03, APO09, APO10, APO12, APO13, BAI01, BAI02, BAI03, BAI06, BAI09, DSS01, DSS02, DSS03, DSS04, DSS05, MEA03
	Norms/Best Practices		- Harmonizing with ITIL for each COBIT process

6 Discussion and conclusions

The integrated model was developed using a systematic approach, having the related work as a baseline. The artifact takes contributions from different theoretical models (BITA, BDIM), as well as from the frameworks (COBIT, ITIL, Guidelines & Areas). While the theoretical models contribute with the knowledge from state-of-the-art practices developed in recent years, the professional frameworks provide with knowledge cumulated through practical experiences.

The preponderant usage of COBIT in the model shows a logical result, which, unlike the academic alternatives cited, demonstrates its applicability and flexibility for strategic alignment. In this work, COBIT was considered as a part of an ecosystem of complementary and competitive frameworks (COBIT, ITIL, ISO 38500, and Guidelines & Areas). However, a model application perspective is the alternative use of ISO 38500 or Guidelines & Areas, which have not been taken into account in the case studies.

The integrated model allows ITG/ITM modeling, and it is evaluated through the case studies, showing its versatility in different contexts. About it, some important issues are discussed below.

IT governance has tended to focus on within a single organizational environment [65], and the understanding of IT governance in an inter-organizational context remains limited [66]. This work contributes to evidence about the inter-organizational influence in the complexity treatment. In the A case study, to minimize the number of COBIT processes is not an alternative to contributing to the IT improvement implementation success. In the B and C cases, the goal cascade process constitutes an adequate filter regard to processes to be into account in the refined model.

The improvement level in the case study A, reveals the contextual implementation difficulties into the public sector, issue studied by various authors [4][41][43]. Particularly, the model requirements reveal the multifaceted goals of stakeholders (municipal enterprises), and the low IT evaluation outcomes evidence the risk aversion that difficult the changes; these issues are negatively empowered on developing country context. Likewise, in the case study B, the implementation level was partial; the fundamental reason was the risk aversion. This reality suggests complementing this work, exploring the critical success factors for the implementation stage.

On the other hand, the case study C shows the model applicability and versatility for to use it in ITG and ITM subdomains, as the ITO services.

In any case, the experiences show model operability and applicability. Thus, the work provides a practical tool for practitioners and contributes to the existing body of knowledge both about artifacts design and on IT development in a developing country Private and Public Sector Organizations.

References

1. P. Webb, C. Pollard and G. Ridley, "Attempting to Define IT Governance: Wisdom or Folly?," in Proceedings of the 39th Hawaii International Conference on System Sciences, pp 1-10 (2006).
2. M. Gerrard, "Defining IT Governance: Roles and relationships," Gartner. ID Number: G00139986 (2006).

3. L. Wilbanks, "IT Management and Governance in Equal Parts," IT Professional, vol. 10, no. 1, pp. 60-61 (2008).
4. J. Campbell, C. McDonald, and T. Sethibe, "Public and Private sector IT Governance: identifying contextual differences," Australasian Journal of Information Systems, vol. 16, no. 2, pp. 5-18 (2009).
5. F. Amarilli, "A Framework for Business-IT Alignment in Turbulent Environments," Athens Journal of Technology Engineering, vol. 1, no. 2, pp. 103-118 (2014).
6. A. Cataldo, R. McQueen, and J. Hardings, "Comparing Strategic IT Alignment versus Process IT Alignment in SMEs," Journal of Research and Practice in Information Technology, vol 44, no.1, pp 43-57 (2012).
7. Y. Chan and B. Reich, "State of Art. IT alignment: what have we learned?," Journal of Information Technology, vol. 22, pp. 297-315 (2007).
8. J. Henderson and N. Venkatraman, "Strategic Alignment: Leveraging information technology for transforming organizations," IBM Systems Journal, vol. 32, no. 1, pp. 4-16 (1993).
9. ISACA, COBIT 5.0. A business framework for governance and management of IT (2012).
10. S. De Haes and W. Van Grembergen, Information Technology Governance: Models, Practices, and Cases, IGI Publishing (2008).
11. R. Almeida, R. Pereira, M. Mira da Silva. IT Governance Mechanisms: A Literature Review. In: Falcão e Cunha J., Snene M., Nóvoa H. (eds) Exploring Services Science. IESS 2013. Lecture Notes in Business Information Processing, vol 143. Springer, Berlin, Heidelberg (2013)
12. ISO, ISO/IEC 38500: 2008. Corporate governance of information (2008).
13. S. De Haes and W. Van Grembergen, "IT Governance and Its Mechanisms," Information Systems Control Journal, vol. 1 (2004).
14. W. Van Grembergen, S. De Haes and W. Guldentops, "Structures, Processes and Relational Mechanisms for IT Governance," in Strategies for Information Technology Governance, Idea Group Inc., pp. 1-36 (2004).
15. R. Almeida, R. Pereira, M. Mira da Silva. IT Governance Mechanisms Patterns. In: Franch X., Soffer P. (eds) Advanced Information Systems Engineering Workshops. CAiSE 2013. Lecture Notes in Business Information Processing, vol 148. Springer, Berlin, Heidelberg (2013)
16. S. De Haes and W. Van Grembergen, "An Exploratory Study into the Design of an IT Governance Minimum Baseline through Delphi Research," Communications of the Association for Information Systems, vol. 22, no. 24 (2008).
17. R. Almeida, "Implementing IT Governance. Thesis to obtain the Master of Science Degree," Técnico Lisboa, Lisboa (2013).
18. A. Moura, J. Sauvé and C. Bartolini, "Research Challenges of Business-Driven IT Management," in 2nd IEEE/IFIP International Workshop on Business-Driven IT Management, pp 19-28 (2007).
19. C. Bartolini and C. Stefanelli, "Business-driven IT Management," in IFIP/IEEE International Symposium on Integrated Network Management (IM), pp 964-969, Dublin, Ireland. 23-27 May (2011).
20. J. Sauvé, A. Moura, M. Sampaio, J. Jornada and E. Radziuk, "An Introductory Overview and Survey of Business-Driven IT Management," in The First IEEE/IFIP International Workshop on Business-Driven IT Management, BDIM '06, pp 1-10 (2006).
21. ITIL, "ITIL. Best Management Practices product," TSO (2011).
22. J. Siviy, P. Kirwan, L. Marino and J. Morley, "The Value of Harmonization Multiple Improvement Technologies: A Process Improvement Professional's View," SEI, Carniege Mellon (2008).
23. Y. Bartens, S. De Haes, Y. Lamoen, F. Schulte, and S. Voss, "On the Way to a Minimum Baseline in IT Governance: Using Expert Views for Selective Implementation of COBIT 5.," in 48th Hawaii International Conference on System Science, pp 4554-4563 (2015).
24. S. De Haes and W. Van Grembergen, "Enterprise Governance of Information Technology," Springer International Publishing, Heidelberg (2015).
25. Q. Liu and G. Radley, "IT Control in the Australian Public Sector: an International Comparison," in European Conference on Information Systems, Germany (2005).

26. L. Omari, P. Barnes, and G. Pitman, "An Exploratory Study into Audit Challenges in IT Governance: A Delphi Approach," in Symposium on IT Governance, Management & Audit (SIGMA2012), Kuala Lumpur, Malaysia (2012).
27. R. Debreceny and G. Gray, "IT Governance and Process Maturity: A Field Study.," in Proceedings of the 42nd Hawaii International Conference on System Sciences (2009).
28. G. Mangalaraj and A. Singh, "IT Governance Frameworks and COBIT - A Literature Review," in Twentieth Americas Conference on Information Systems, pp 1-10, Savannah (2014).
29. L. Gerke and G. Ridley, "Towards an abbreviated COBIT framework for use in an Australian State Public Sector," in 17th Australasian Conference on Information Systems, Adelaide, Australia (2006).
30. L. Gerke and G. Ridley, "Tailoring COBIT for Public Sector IT Audit: An Australian Case Study," in Information Technology Governance and Service Management: Frameworks and Adaptations, A. Cater-Steel, Ed., New York, IGI Global, pp 101-124 (2009).
31. R. Lubbad, "Towards An Abbreviated Model of IT governance for Palestinian government sector According to COBIT 5 framework," (2014).
32. C. Bartolini, M. Sallé, and D. Trastour, "IT service management driven by business objectives: an application to incident management," in Network Operations and Management Symposium, NOMS 2006. 10th IEEE/IFIP, pp 45-55 (2006).
33. C. Bartolini, "Business-driven IT Management. DOTTORATO DI RICERCA IN SCIENZE DELL'INGEGNERIA INGEGNERIA DELL'INFORMAZIONE," Università degli Studi di Ferrara (2009).
34. R. Pereira and M. Mira da Silva, "Towards an Integrated IT Governance and IT Management Framework," in IEEE 16th International Enterprise Distributed Object Computing Conference, pp 191-200 (2012).
35. R. Pereira and M. Mira da Silva, "A Literature Review: IT Governance Guidelines and Areas," in ICEGOV '12, Albany (2012).
36. S. De Haes, W. Van Grembergen, and R. Debreceny, "COBIT 5 and Enterprise Governance of Information Technology: Building Blocks and Research Opportunities," Journal of Information Systems, vol. 7, no. 1, pp 307–324 (2013).
37. L. Omari, "IT Governance evaluation: adapting and adopting the COBIT framework for Public Sector Organizations," (2016).
38. ISO, "ISO/IEC 20000-1:2011. Information technology -- Service management -- Part 1: Service management system requirements," (2011).
39. S. Cots, M. Casadesús and F. Marimon, "Benefits of ISO 20000 IT service management certification," Information Systems and E-Business Manage, vol. 14, no. 1 (2016).
40. G. Disterer, "ISO 20000 for IT," Business Information System Engineering, vol. 6, pp. 463-467 (2009).
41. T. Winkler, "IT Governance Mechanisms and Administration/IT Alignment in the Public Sector: A Conceptual Model and Case Validation," in 11th International Conference on Wirtschaftsinformatik, pp 831-845 (2013).
42. N. Helbig, J. Hrdinová, and D. Canestraro, "Enterprise IT governance at the state level: An emerging picture," in The Proceedings of the 10th International Digital Government Research Conference, pp 172-179 (2009).
43. T. Sethibe, J. Campbell, and C. McDonald, "IT Governance in Public and Private Sector Organisations: Examining the Differences and Defining Future Research Directions," in ACIS 2007 Proceedings, pp 833-843 (2007).
44. P. Bermejo, A. Tonelli, and A. Zambalde, "Developing IT Governance in Brazilian Public Organizations," International Business Research, vol. 7, no. 3, pp. 101-114 (2014).
45. I. Kurti, E. Barolli, and K. Sevrani, "Effective IT Governance in the Albanian Public Sector – A critical success factors approach," The Electronic Journal of Information Systems in Developing Countries, vol. 63, pp. 1-22 (2014).
46. N. Al Qassimia and L. Rusua, "IT Governance in a Public Organization in a Developing Country: A Case Study of a Governmental Organization," in Conference on ENTERprise Information Systems, pp 450-456 (2015).

47. J. Denford, G. Dawson and K. Desouza, "An Argument for Centralization of IT Governance in the Public Sector," in 48th Hawaii International Conference on System Sciences, pp 4493-4501 (2015).
48. ISO, "ISO/IEC 15504-4. Information technology — Process assessment — Part 4: Guidance on use for process improvement and process capability determination," (2009).
49. H. J. von Wangenheim, C. Salviano and A. von Wangenheim, "Systematic Literature Review of Software Process Capability/Maturity Models," in Proceedings of International Conference on Software Process. Improvement And Capability dEtermination (SPICE), Pisa, Italia (2010).
50. SEI, CMMI® for Services, Version 1.3, Hanscom AFB, MA: Carnegie Mellon University (2010).
51. ISACA, Process Assessment Model (PAM): Using COBIT 5 (2013).
52. A. Hevner, S. Ram, S. March and J. Park, "Design Science in Information Systems Research," MIS Quarterly, vol. 28, no. 1, pp. 75-105 (2004).
53. A. Hevner and S. Chatterjee, Design Research in Information Systems, New York: Springer Publishing (2010).
54. S. Gregor and A. Hevner, "Positioning and presenting Design Science Research for maximum impact," MIS Quarterly, pp. 337-355 (2013).
55. K. Peffers, M. Rothenberger and B. Kuechler, Design Science Research in Information Systems. Advances in Theory and Practice. 7th International Conference, DESRIST 2012, K. Peffers, M. Rothenberger and B. Kuechler, Eds., Springer-Verlag Berlin Heidelberg (2012).
56. M. Chiarini, D. VanderMeer, M. Rothenberger, A. Gupta and V. Yoon, Advancing the Impact of Design Science: Moving from Theory to Practice. DESRIST 2014, Miami: Springer (2014).
57. M. Helfer and B. Donnellan, Practical aspects of Design Science. European Design Science Symposium, EDSS2011, M. Helfer and B. Donnellan, Eds., Springer-Verlag Berlin Heidelberg (2012).
58. J. Mason, Qualitative Researching, London: SAGE Publications (2002).
59. SAGE, The SAGE Encyclopedia of Qualitative Research Methods. Given, L. (Editor), Los Angeles: SAGE Publications (2008).
60. A. De la Torre and M. Néjer, «Propuesta del modelo de control de gestión de servicios de TI para las Empresas Municipales del Distrito Metropolitano de Quito. Caso de estudio,» EPN. Quito (2017).
61. C. Montenegro and D. Flores, "An Integrated Model for ICT Governance and Management," in ICCCS, pp 1-9, Pointe aux Piments, Mauritius (2015).
62. C. Montenegro, N. Nuñez, and A. Larco, "IT Outsourcing Improvement: Contractual Model of Governance and Management from Customer Perspective," in WorldCIST 2017. Advances in Intelligent Systems and Computing, pp 616-627 Cham (2017).
63. Asamblea Nacional - Ecuador, "Ley Orgánica de Empresas Públicas," Quito (2009).
64. Consejo Metropolitano de Quito, "Ordenanza 039," (2010).
65. J. Chong and F. Tang, "IT Governance in Collaborative Networks: A Socio-Technical Perspective," Pacific Asia Journal of the Association for Information Systems, vol. 4, no. 2, pp. 31- 48 (2012).
66. S. Trang, N. Opitz, and L. Kolbe, "IT Governance in a Network Context: Literature Review and Agenda for Research," in Proceedings of the Nineteenth Americas Conference on Information Systems, pp 1-10, Chicago, Illinois (2013).

Analysis of environmental factors in the adoption of ISO/IEC 29110. Multiple case study

Stuardo Lucho[1], Karin Melendez[2] and Abraham Dávila[2]

[1]Escuela de Posgrado, Pontificia Universidad Católica del Perú, Lima 32, Lima, Perú
[2]Departamento de Ingeniería, Pontificia Universidad Católica del Perú, Lima 32, Lima, Perú
{stuardo.lucho, kmelendez, abraham.davila}@pucp.edu.pe

Abstract. The software industry comprising small companies represents an interesting opportunity for economic development. But these still have problems in productivity and the quality of their delivered products. In this context, the ISO/IEC 29110 standard has been developed, which represents an opportunity for small companies whose adoption depends on several factors, in particular, those related to the environment. In this paper, we study the influence of environmental factors on the adoption of ISO/IEC 29110 standard. For this research, a multiple case study was carried out, which includes four organizations as units of analysis. In a controlled environment, each organization implemented the processes belonging to the basic profile of the ISO/IEC 29110 standard. After the implementation, an analysis has been made of the environmental factors that influenced the adoption of the standard. Of the 16 environmental factors analyzed, it was found that 6 factors influenced all the organizations and those related to the support and trust of the partners had the greatest positive influence on the standard adoption, while the defense factor of the partners had the most negative influence. All other factors had minimal or no influence.

Keywords: ISO/IEC 29110, adoption factors, software process improvement.

1 Introduction

In the software industry, companies define and improve their processes for [1][2]: (1) managing the software lifecycle, guiding professionals to perform their tasks; and (ii) providing a general understanding to all those involved in software development about the tasks for which they are responsible. When companies decide to improve their software processes, there are several standards and models [3]. Some of them focused on large companies or workgroups with more resources such as CMMI or ISO/IEC 15504 [4]; and others oriented to small organizations as MoProSoft or ISO/IEC 29110, among others [5].

However, the adoption of a process model depends on many factors, such as company size [6][7], organizational culture, resources to be invested, expected results [8][9], the correct use of a reference model, the participation of the top management of the organization; and especially the high ratio of people who are part of the organization's work team [10]. All these factors have been studied and organized by various authors in categories: organizational factors, top management and the environment,

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_8

among others; or as frameworks such as those presented in various studies [11], [12], [13], [14].

In this context, the main task of the ProCal-ProSer project [15] is to study the different behaviors of people in software development companies that seek to adopt the ISO/IEC 29110 standard, to determine why or not to adopt this model. Identified factors will allow defining adoption strategies in the software industry. Also, to reduce bias, companies with less than 25 professionals in the software development area have been considered.

This article presents an exploratory research on the factors of adoption related to the environment in a group of companies that implement the ISO/IEC 29110 standard. The rest of the paper is arranged as follows: Section 2, shows background; Section 3 research protocol; Section 4, units of analysis; Section 5, presents the interpretation of results and Section 6, the final discussion and future work.

2 Background

2.1 Software process models

There are several ways to achieve a better quality of the software product [16],[17], including quality in the software development process [18] based on software processes models.

Among the most representative models are: CMMI [19] and ISO/IEC 12207 standard [20]; however from the perspective of small companies, both CMMI and ISO/IEC 12207 have adoption problems and generate rejection when trying to be implemented in small organizations [4], [21], [22]. Given the situation described, other models were generated for small companies and some of them used in Peru: MoProSoft [23], Competisoft [24], MPS.BR [25], and ISO/IEC 29110 standard [26].

From the above, ISO/IEC 29110 is a standard developed with the aim of improving the quality of software products through improving the processes of small companies [27]. The standard defines a Very Small Entity (VSE) as an organization, company, department or project that has up to 25 professionals engaged in software projects [26]. The standard is organized by profiles such as: basic, intermediate and advanced; and in each profile a set of processes is established [26]. The basic profile is detailed in the document ISO/IEC 29110-5-1-2 [28] and is composed of two processes: (i) Project Management (PM) which aims to systematically carry out the tasks of the software development project, aiming to meet the objectives of the project in relation to quality, time and budget, and (ii) Software Implementation (SI) which aims to establish the systematic relationship of the analysis, construction, design, integration and testing activities of the software product according to the specified requirements. Each process and its main activities are shown in Fig.1 and Fig. 2.

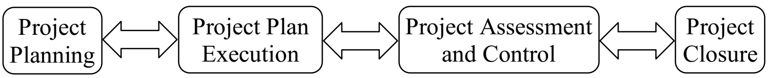


Fig. 1. Main activities of the Project Management Process [28]

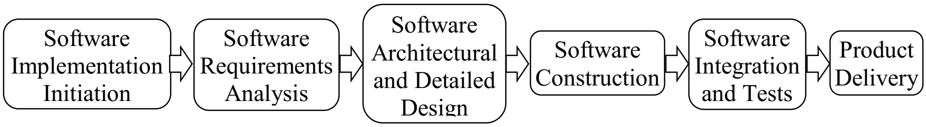


Fig. 2. Main activities of the Software Implementation Process [28]

2.2 Factors models in Software Process Improvement

The adoption of practices from a model following process improvement can be influenced positively or negatively by a set of factors. This software processes improvement (SPI) in a company can be considered as an innovation in information technologies [29].

In this context, there are research works such for VSEs as [30] and [31], where methods are proposed to identify key factors to be considered if the company wishes to implement a successful software process improvement. Some factors identified were: software maintenance, product quality and change management.

Likewise, there are other studies that have collected a set of factors that can influence positively or negatively when performing a SPI. These studies include: (i) the analysis performed in [32] which shows key points to be taken into account in the different phases of software development and in the people who implement the SPI; (ii) the 18 success factors to achieve SPI, in small and medium-sized Pakistani companies described in [33], in the context of web development applications; (iii) the study of [34] which analyzes the influence of other factors such as: organizational culture, the exchange and distribution of knowledge among employees, among other; (iv) the drawbacks associated with adopting an agile methodology described in [35]; where the author identifies a set of factors such as: the size of the project, assimilation gaps or cultural problems, among others; and (v) finally the conceptual model shown in [11], where 107 factors were collected from the current literature, which may influence the adoption of an information technology innovation. Factors were grouped into 5 categories: innovation (20 factors), organizational (41 factors), environment (16 factors), CEO (8 factors) and user acceptance (22 factors).

2.3 Related work

The ISO/IEC 29110 standard is focused on enabling companies to improve their software processes without having the complexity of other standards or models [9]; thus, there are studies such as [10] and [36], which show the perception of the VSEs about adopting the ISO/IEC 29110 standard in the future. These studies discuss potential benefits and barriers to possible adoption. Also, in these papers [37], [38], [39] and [40] some recommendations and guidelines to be considered when implementing the ISO/IEC 29110 standard in VSEs are analyzed, such as: "Deployment Packages", close assistance to the team for the implementation of the standard and the use of support tools for the management of SI and PM processes. The papers previously mentioned in this section are considered of high relevance for the present study; However, the purpose of this article is to analyze the factors after the implementation of the ISO/IEC 29110 standard, not before, which is why it differs from these papers.

On the other hand, there is another group of articles that analyze which factors influence the VSEs, when implementing the SPI using the ISO/IEC 29110 standard. One of them is [41], where it is collected (from the Current literature) and classify 37 factors (on success factors and barriers) that may influence the adoption of the ISO/IEC 29110 standard. Our research uses another conceptual model with different groups of factors, as a first difference with the previous article, then analyze and infer which are the factors that influence positively and negatively. However, in the discussion section, it will be analyzed the relation between both groups of factors.

Finally, papers [42] and [43] show the point of view of some VSEs regarding the adoption of the ISO/IEC 29110 standard, information obtained through the use of grounded theory methodology and interviews. As a result, topics such as: level of interest, acceptance level and barriers present in a VSE to adopt a software process improvement with ISO/IEC 29110 standard are shown. Our research has similarity to the previous two; however, the purpose of this study is to know which factors have a greater or lesser influence in the software processes improvement (based on a conceptual model [11]), unlike the previous studies which obtain the factors as a result of the methodology used.

3 Research methodology

In the present study, an exploratory analysis of an improvement process is carried out, under similar contexts, but in a set of different organizations, so we have used a case study as a method of qualitative research method.

The protocol that has been followed for the case study is the one defined in [30] and [31], likewise, we have [11] as document reference, which shows a practical case of protocol implementation, in an analysis of factors of adoption in a company. The process followed consists of 4 phases and several stages (see Table 1).

In the following paragraphs, the stages of the case study related to the planning phase (P) are developed:

(P1-1) Purpose of the study: to qualitatively analyze the way in which environmental factors influence the adoption of ISO/IEC 29110 standard, the work was doing in work teams of four VSE, which are described in Section 4. To achieve this study a software process improvement cycle was performed in all units of analysis using the basic profile of ISO/IEC 29110 standard, which involves software development processes and project management processes. Within the ProCal-ProSer project [44], it was established to work with factor models used in the context of process improvement and models of technological innovation factors, with the purpose of evaluating their application in several VSEs. For this particular research, it was established to work with the conceptual model defined in [11], which groups adoptions factors into 5 categories as shown in section 2.2. This article uses and analyzes the 16 environmental factors defined in the model and the study of the rest of the factors is described in the future work section.

Table 1. Phases and steps of the case study process [15]

Phase	Step
P1: Design the case study	(P1-1) The purpose of the study (P1-2) What is studied (P1-3) Theory required (P1-4) Research questions (P1-5) Methods of collection (P1-6) Selection of data
P2: Prepare data collection	(P2-1) Definition of how the collection is performed
E1: Collect the data	(E1-1) Collection of data
E2: Analyze and interpret data	(E2-1) Analysis of data (E2-2) Interpretation of data
R1: Report the results	(R1-1) Carrying out the report

(P1-2) What is studied: the units of analysis are four VSEs. The work teams are smaller than 25 people, so there was no restriction to implement process improvement based on ISO/IEC 29110 standard. The detailed description of the companies and the selected projects is presented in Section 4. Consequently, the study object is based on a group of analysis instances each under its own context, which allows defining a holistic multiple case study [31].

(P1-3) Theory required: see Section 2.

(P1-4) Research questions: the research question for the present study is: How do environmental factors influence the adoption of the basic profile practices defined in the ISO/IEC 29110-5-1-2?

(P1-5) Methods of collection: the information collected was of the qualitative type and was obtained in two stages- In the first stage, relevant information was obtained on the adopted practices of the ISO / IEC 29110-5-1-2 standard, through documents and reports. The data collection process, for this first stage, can be considered as a third degree, according to [30]. In the second stage, the information was obtained through a semi-structure interview, to achieve a descriptive and exploratory study of the environmental factors that influenced the SPI. In this case, the technique can be considered as a first degree due to there was direct contact between the study subjects and the researcher. The interview questions were formulated based on the description of each factor and were reviewed and approved by an external reviewer with more than 10 years of experience in SPI issues and quality of software processes. Additional details of the interview are shown in Section 5.

(P1-6) Selection of data: it was carried out in two stages: The first, by obtaining the report of the level of adoption based on ISO/IEC 15504 assessment and performed by formal assessor, which could determine which practices of the Standard were adopted and which were not, and to what degree (see Section 4). In the second stage, the interviews were refined and consolidated, to obtain a diagnosis of the factors related to the environment that influenced the adoption (see Section 5).

(P2-1) Definition of how the collection is performed: it was carried out in two stages. The first one, where the adoption level of the standard was evaluated and the second one, through semi-structured interviews, whose central focus was to know which factors [7] positively or negatively influenced the adoption of the basic profile.

The remaining steps were analyzed as follows: (E1-1) in section 4, (E2-1) and (E2-2) in section 5, (R1-1) in the whole article.

4 Analysis units

We have four units of analysis for the multiple case study. The VSEs, that for confidentiality issues will be called alpha, beta, gamma and delta, are described in Table 2.

Table 2. Units of analysis

Item	Alpha	Beta	Gamma	Delta
Description	Software area within an educational organization. It develops software projects to be used in educational environments.	Private sector VSE that develops hardware, software and networking solutions for the gaming industry.	VSE that offers solutions in diverse sectors, mainly commercial, retail, health, mining and engineering and education.	VSE that offers solutions for iPhone, iPad, Android and Windows Phone.
Team	1 project manager 2 junior developers 1 graphic designer	1 project manager 3 senior developers 2 analysts 1 junior developer 1 test manager	1 project manager 2 analysts 3 senior developers	1 general manager 1 sub-manager 4 senior developers 2 administrative staff.
Enterprise Expertise	5 years	13 years	8 years	10 years
SPI experience	No previous experience	No previous experience	No previous experience	No previous experience
SPI support	post graduate student	post graduate student	Undergraduate student (remunerated)	Undergraduate student (remunerated)
SPI duration	4 months	6 months	6 months	6 months

The process followed includes an initial diagnostic evaluation, setting the goal of improvement, implementation and final evaluation of the improvement cycle. The diagnostic evaluation was carried out at the beginning of the project and the basic profile, PM and SI processes were evaluated in order to know the current state of the companies and propose improvement proposals. The process followed for this assessment was performed using ISO/IEC 15504-2, as specified in ISO/IEC 29110-3 [32]. The diagnostic evaluation was performed by a qualified evaluator together with the project manager of each company and presenting the documentation required by each practice or group of practices.

Column IE (initial evaluation) in Table 3, shows the percentage of adherence of companies in each of the processes implemented, based on the four levels defined by ISO/IEC 15504-2: N (not met), P (partially Complies), L (long is fulfilled) and F (Full - has a great achievement of the process). Likewise, percentages of adhesion to the guide have been placed; those are representing the degree of compliance based on the practices adopted by the company. These percentages were placed by the authors to facilitate comparison and analysis between the initial and final evaluation. The base practice qualification from process assessment process is showed in Appendix A. The improvement goals were established based on the diagnostic evaluation and were different for each company: Alfa aimed to achieve at least an L level in both processes of the basic profile; While Beta aimed to achieve a P level in PM and a L level in SI and finally, Gamma and Delta aimed to achieve F level in SI and PM. The implementation was carried out in all the VSE using the basic profile (PM and SI) of ISO/IEC 29110-5-1-2. The final evaluation was carried out by an external evaluator according to what was established in the Project.

5 Analysis and results

The process followed for the final evaluation was the same followed for the diagnostic evaluation. The results of the final evaluation are presented in column FE (final evaluation) of Table 3. It can be seen in Table 3, for example, that Alfa achieved a significant improvement in its PM process, going from a N rating with 10.6% compliance to an L with 62.5% compliance.

Table 3. Comparison Initial and Final evaluation

	Project Management			Software Implementation		
Organiz. Id	Adherence (%) IE → FE	Level variation IE → FE	Capacity variation IE → FE	Adherence (%) IE → FE	Level variation IE → FE	Capacity variation IE → FE
Alfa	10.6 → 62.5	N → L	0 → 0	37.9 → 42.8	P → L	0 → 0
Beta	14.6 → 48.1	N → P	0 → 0	21.9 → 49.4	P → P	0 → 0
Gamma	89.4 → 96.2	F → F	1 → 1	52.9 → 71.8	L → L	0 → 0
Delta	27.9 → 84.6	P → L	0 → 1	29.9 → 77.4	P → L	0 → 0

After the evaluation, the factors influencing in the adoption were analyzed through a semi-structured interview (see Section 3, P1-5), the activity was performed with project managers from each company. The purpose of the research is to analyze the factors related to the environment, which are shown in Table 4, together with the description of each factor. For each factor, the researcher proposed a question that would make it possible to know during the interview whether the factor influenced the SPI, and if so, at what level. These questions are shown in Appendix B.

Likewise, each factor was evaluated according to the following criteria: (i) influence on process improvement: positive or negative; (ii) level of influence, on a Likert scale, from 1 to 5, where 1 is very low influence and 5 is high influence; (iii) justification of the influence of the factor; (iv) what problems were presented on this factor in the improvement? (v) Criticality of the problem, on a Likert scale, from 1 to 5, where 1 is very low criticality and 5 is high criticality; and (vi) how did you deal with this problem?

Table 5 presents the consolidated of the interviews, considering two columns for each company, the first is the type of influence (Inf.) and the second, the level of that influence. If the factor evaluated had no influence, it is placed N.A, if the influence was positive is placed (+) and if it was negative is placed (-) in column Inf.; and the level is placed in the adjacent column. In this research we used narrative synthesis, according to those established in [45], to analyze the information of the environmental factors described in [11], which through interviews were collected and mapped in Table 5.

Table 4. Environmental factors [11]

ID	Factor	Description
E01	Competitive pressure	Pressure exerted by competitors in the same sector.
E02	External pressure	Influences that arise from various sources in the competitive environment surrounding the organization.
E03	Government support	The government supports with financial incentives, tax cuts, pilot projects to benefit innovation.
E04	Vendor support	Assistance from the seller of a product or service that the company own.
E05	Partners support	Degree of Support from the company's partners to corporate decisions.
E06	Partners readiness	Availability of partners to support process improvement.
E07	Environmental Uncertainty	The company has little information about its external environment; therefore, its behavior is erratic.
E08	Vertical linkage	Coordination and communication activities from the highest to the lowest positions in the company.
E09	Partners defense	The partners of the company support the improvement of processes.
E10	Government pressure	Pressure exerted by the government towards the adoption of standards and obtaining certifications.
E11	Number of competitors	Number of competitors identified by the company.
E12	External expertise	The company requests the external consulting from a third-party supplier.
E13	Consultant effectiveness	The requested consultancy offered a solution and / or helped to clarify the present problems.
E14	Trust with partners	Trust given by partners in the process improvement adoption process.
E15	Globalization	Increased number of software companies in the world, customers of the company in other countries, new markets.
E16	Social influence	The perception of the public (or other company) about the position of the company.

Among the justifications provided by the project managers, we can observe the comment for the factor E02 (External pressure) from Alfa: "Customers requested documents to verify the progress and what was being proposed as a solution, the standard supported compliance with this requirement and required the work team to comply with the required documentation. "In two of the four companies, this factor influenced a high level, as did the factor E05 (Partner support), factor E08 (Vertical link) and factor E14 (Trust in partners).

Regarding the factors that had a greater influence in the adoption of process improvement, in Alpha, E14 obtained a value of 5 and the comment was as follows: "The reason is that the technical leader is also the software specialist and the area manager relied fully on his capabilities to implement process improvement and get the results sought". In relation to the most influential factor in Beta, the support of the partners (E06) was rated 5 and the interviewee's comment was: "the manager always agreed with the decisions made by the project leader, he trusted in what the leader said, he trusted people a lot."

Table 5. Results of Factor Evaluation

	Alfa case		Beta case		Gamma case		Delta case	
Factor ID.	Inf.	Level	Inf.	Level	Inf.	Level	Inf.	Level
E01	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
E02	+	4	+	3	N.A.	N.A.	N.A.	N.A.
E03	+	3	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
E04	+	1	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
E05	+	3	+	5	+	2	+	5
E06	+	3	N.A.	N.A.	+	3	+	5
E07	N.A.	N.A.	+	1	N.A.	N.A.	N.A.	N.A.
E08	+	4	+	4	+	5	+	5
E09	-	2	-	3	+	4	+	3
E10	N.A.	N.A.	+	2	+	2	N.A.	N.A.
E11	N.A.	N.A.	+	2	N.A.	N.A.	N.A.	N.A.
E12	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
E13	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
E14	+	5	+	4	+	3	+	4
E15	+	1	+	2	+	3	+	4
E16	N.A.	N.A.	N.A.	N.A.	+	5	N.A.	N.A.

On the other hand, it can be observed that factor E09 (Defense of partners) had a negative influence on two companies. An excerpt from the interview comment to the Beta company regarding this factor: "The manager of the company knew about the improvement of processes, but not being a technical person (not an engineer), he preferred not to go into the detail of the implementation of the Improved and left it in the hands of the project leader.

Regarding the possible threats of the study, it should be considered that the SPI was guided by a member of the ProCal-ProSer project team in gamma and delta, which did not happen in alfa and beta, and above all may not always be applied to other VSEs. Likewise, regarding the analysis of the factors, the interviews were made to the team leaders and perhaps team members had a different opinion and other levels of influence from the same factors.

6 Final discussion and future work

The adoption of a process improvement involves a series of steps and good practices that must be followed to achieve the goals set by the company. When it comes to a small organization, based on previous studies and the results obtained, this improvement will be implemented better if there is a total commitment of the highest positions of the company to the lowest positions.

The results obtained allow concluding that, of the 16 factors analyzed, 5 of them influenced all the organizations, regardless of the context in which they were. This group of 5 factors, E05 partners support, E08 vertical linkage and E14 trust with partners, which are related to top management support in SPI, are factors identified as positive influence, which is in agreement with the research [41] (Section 2.2), where two factors mapped, "Involvement and participation", and "Organizational policy", both related to the support of the top management and the work team, are considered as success factors. Also, these three factors are related to the research [42] as it is mentioned that two barriers of adoptions are lack of perceived need and resources need, which in this case, all de VSEs could afford all the resources needed for the SPI, overcoming the barriers.

The other two factors that influenced all companies, E09 partners defense and E15 globalization, have not been identified as success factors or barriers to adoption in any previous paper, therefore, these two factors should be considered in the future, in similar research works, when studying factors that affect SPI.

Finally, of all the factors analyzed, the only one that influenced negatively in two companies was E09 partners support, which allows to conclude that this factor must be supervised with great care when performing an SPI using the ISO/IEC 29110 standard.

As a future work, it is proposed to perform the procedure described in this paper, but in focus on the factors that have not been analyzed in this study, which belonging to [11], such as: innovation, organizational, CEO and user acceptance. Also, it is proposed to analyze if there is any correlation between the level of improvement after SPI and the factors that have influenced the adoption of that SPI.

Acknowledgments. This work has been carried out within the ProCal-ProSer project funded by Innóvate Perú under Contract 210 - FINCYT-IA-2013 and by the Department of Engineering and the Software Engineering Research and Development Group (GIDIS) of the Pontifical Catholic University of Peru.

References

1. Humphrey, W.S.: A Discipline for Software Engineering. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1995).
2. Unterkalmsteiner, M., Gorschek, T., Islam, a. K.M.M., Permadi, R.B., Feldt, R.: Evaluation and Measurement of Software Process Improvement—A Systematic Literature Review. IEEE Trans. Softw. Eng. 38, 398–424 (2012).
3. Vasconcellos, F.J.S., Landre, G.B., Cunha, J.A.O.G., Oliveira, J.L., Ferreira, R.A., Vincenzi, A.M.R.: Approaches to strategic alignment of software process improvement: A systematic literature review. J. Syst. Softw. 123, 45–63 (2017).
4. Huang, D.B., Zhang, W.: CMMI in medium & small enterprises: Problems and solutions, (2010).
5. Dávila, A., Cano, C.: Revisión sistemática de comparación de modelos de procesos software, http://tesis.pucp.edu.pe/repositorio/handle/123456789/6521.
6. Horvat, R.V., Rozman, I., Györkös, J.: Managing the complexity of SPI in small companies. Softw. Process Improv. Pract. 5, 45–54 (2000).
7. Basri, S. Bin, O'Connor, R. V.: Organizational commitment towards software process improvement: An Irish software VSEs case study. Proc. 2010 Int. Symp. Inf. Technol. - Syst. Dev. Appl. Knowl. Soc. ITSim'10. 3, 1456–1461 (2010).

8. Ebert, C.: Technical controlling and software process improvement. J. Syst. Softw. 46, 25–39 (1999).
9. Larrucea, X., O'Connor, R. V., Colomo-Palacios, R., Laporte, C.Y.: Software Process Improvement in Very Small Organizations. IEEE Softw. 33, 85–89 (2010).
10. Basri, S., O'Connor, R. V.: Understanding the perception of very small software companies towards the adoption of process standards. Commun. Comput. Inf. Sci. 99 CCIS, 153–164 (2010).
11. Hameed, M.A., Counsell, S., Swift, S.: A conceptual model for the process of IT innovation adoption in organizations. J. Eng. Technol. Manag. - JET-M. 29, 358–390 (2012).
12. Bayona, S., Calvo-manzano, J.A., Feliu, T.S.: Review of Critical Success Factors Related to People in Software Process Improvement. 20th EuroSPI Conf. 179–189 (2013).
13. Bayona, L.S., Villalon, J.A.C.-M., Agustin, G.C., Gilabert, T.S.F.: Taxonomia de Factores Criticos para el Despliegue de Procesos Software. REICIS, Rev. Española Innovación, Calid. e Ing. del Softw. 6, 6–22 (2010).
14. Sulayman, M., Urquhart, C., Mendes, E., Seidel, S.: Software process improvement success factors for small and medium Web companies: A qualitative study. Inf. Softw. Technol. 54, 479–500 (2012).
15. Dávila, A., Pessoa, M.: Factors driving the adoption of ISO/IEC 29110: A case study of a small software enterprise. Proc. - 2015 41st Lat. Am. Comput. Conf. CLEI 2015. (2015).
16. Scholar, P.G., Science, C., Coimbatore, T.: Automated Testcase Generation for Software Quality Assurance. 2016 10th Int. Conf. Intell. Syst. Control. (2016).
17. Svensson, R.B., Gorschek, T., Regnell, B., Torkar, R., Shahrokni, A., Feldt, R.: Quality Requirements in Industrial Practice - An Extended Interview Study at Eleven Companies. IEEE Trans. Softw. Eng. 38, 923–935 (2012).
18. Elhag, A.A.M., Elshaikh, M.A., Mohamed, R., Babar, M.I.: Problems and future trends of software process improvement in some Sudanese software organizations. Proc. - 2013 Int. Conf. Comput. Electr. Electron. Eng. 'Research Makes a Differ. ICCEEE 2013. 263–268.
19. Equipo del Producto CMMI: CMMI ® para Desarrollo, Versión 1.3. (2010).
20. ISO/IEC JTC 1/SC 7: ISO/IEC 12207:2008, https://www.iso.org/obp/ui/#iso:std:iso-iec:12207:ed-2:v1:en.
21. Brodman, J.G., Johnson, D.L.: A Software Process Improvement Approach Tailored for Small Organizations and SmaU Projects. In: Proceedings of the (19th) International Conference on Software Engineering. pp. 661–662 (1997).
22. Ruiz, J.C., Osorio, Z.B., Mejia, J., Muñoz, M., Chávez, A.M., Olivares, B.A.: Definition of a hybrid measurement process for the models ISO/IEC 15504-ISO/IEC 12207:2008 and CMMI Dev 1.3 in SMEs. Proc. - 2011 IEEE Electron. Robot. Automot. Mech. Conf. CERMA 2011. 421–426 (2011).
23. Oktaba, H., Alquicira, C., Angélica, E., Ramos, S., Martínez, A., Quintanilla, G., Mara, O., López, R., López, F., Hinojo María, L., Rivera, E., María, L., Orozco, J., Yolanda, M., Ordóñez, F., Ángel, M., Lemus Versión, F., Índice, A.: Modelo de Procesos para la Industria de Software: MoProSoft, (2005).
24. Piattini, M., Oktaba, H., Orozco, J., Alquicira, C.: Competisoft. Mejora de procesos software para pequeñas y medianas empresas y proyectos. RA-MA EDITORIAL (2008).
25. Kalinowski, M., Weber, K., Santos, G., Franco, N., Duarte, V., Travassos, G.: Software Process Improvement Results in Brazil Based on the MPS-SW Model. Softw. Qual. Prof. 17, (2015).
26. ISO/IEC-JTC-1/SC-7: ISO/IEC TR 29110-1 Lifecycle profiles for Very Small Entities (VSEs) — Part 1: Overview. (2011).
27. Laporte, C.Y., O'Connor, R. V., Paucar, L.H.G.: The implementation of ISO/IEC 29110 software engineering standards and guides in very small entities. Commun. Comput. Inf. Sci. 599, 162–179 (2016).
28. ISO/IEC-JTC-1/SC-7: ISO/IEC TR 29110-1 Lifecycle profiles for Very Small Entities (

VSEs) — Part 5-1-2:Management and engineering guide: Generic profile group: Basic profile. (2011).
29. Anacona, D., Bastidas, M.I., Pino, F.J., Pardo, C.: Innova-Procedure : A procedure to guide the innovation of software development processes in VSEs Innova-Procedure : Un procedimiento para guiar la innovación de procesos de desarrollo software en VSEs. 108–115 (2015).
30. Jezreel, M., Mirna, M., Pablo, N., Edgar, O., Alejandro, G., Sandra, M.: Identifying findings for software process improvement in SMEs: An experience. Proc. - 2012 9th Electron. Robot. Automot. Mech. Conf. CERMA 2012. 141–146 (2012).
31. Jezreel, M., Mirna, M., Edrisi, M., Diego, E., Al Rumeni, a, Juan, M.: Leading the effort of software process improvement in SMEs. Inf. Syst. Technol. (CISTI), 2013 8th Iber. Conf. 1–7 (2013).
32. Cuevas, G., Calvo-Manzano, J.A., García, I.: Some Key Topics to be Considered in Software Process Improvement. 119–142 (2014).
33. Sulayman, M., Mendes, E., Urquhart, C., Riaz, M., Tempero, E.: Towards a theoretical framework of SPI success factors for small and medium web companies. Inf. Softw. Technol. 56, 807–820 (2014).
34. Lee, J.-C., Shiue, Y.-C., Chen, C.-Y.: Examining the impacts of organizational culture and top management support of knowledge sharing on the success of software process improvement. Comput. Human Behav. 54, 462–474 (2016).
35. Paulk, M.C., Duncan, S., Garzás, J.: On Empirical Research into Scrum Adoption. Presented at the (2011).
36. Sánchez-Gordón, M.L., O'Connor, R. V.: Understanding the gap between software process practices and actual practice in very small companies. Softw. Qual. J. 24, 549–570 (2016).
37. O'Connor, R. V., Laporte, C.Y.: Using ISO/IEC 29110 to harness process improvement in very small entities. Commun. Comput. Inf. Sci. 172, 225–235 (2011).
38. O'Connor, R. V., Sanders, M.: Lessons from a Pilot Implementation of ISO/IEC 29110 in a Group of Very Small Irish Companies. Commun. Comput. Inf. Sci. 349 CCIS, 243–246 (2013).
39. Paucar, L.G., Laporte, C.Y., Arteaga, J., Bruggmann, M.: Implementation and Certification of ISO/IEC 29110 in an IT Startup in Peru. 17, 14 (2015).
40. O'Connor, R. V.: Early Stage Adoption of ISO/IEC 29110 Software Project Management Practices: A Case Study. 226–237 (2014).
41. Wongsai, N., Siddoo, V., Wetprasit, R.: Factors of influence in software process improvement: An ISO/IEC 29110 for very-small entities. Proc. - 2015 7th Int. Conf. Inf. Technol. Electr. Eng. Envisioning Trend Comput. Inf. Eng. ICITEE 2015. 12–17 (2016).
42. Sanchez-Gordon, M.L., O'Connor, R. V., Colomo-Palacios, R.: Evaluating VSES viewpoint and sentiment towards the ISO/IEC 29110 standard: A two Country grounded theory study. Commun. Comput. Inf. Sci. 526, 114–127 (2015).
43. O'Connor, R. V.: Evaluating management sentiment towards ISO/IEC 29110 in very small software development companies. Commun. Comput. Inf. Sci. 290 CCIS, 277–281 (2012).
44. Dávila, A.: Presentación ProCal-ProSer, https://sites.google.com/a/pucp.pe/procal-proser/.
45. Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., Duffy, S.: Guidance on the Conduct of Narrative Synthesis in Systematic Reviews. A Prod. from ESRC Methods Program. 211–219 (2006).

Appendix:

Available in:
https://drive.google.com/drive/folders/0B_K_llz5juqPMll5TC1sNkliWGc

A Theoretical Analysis of Digital Marketing Adoption by Startups

Sérgio Teixeira[1], José Martins[1, 2], Frederico Branco[1, 2], Ramiro Gonçalves[1, 2], Manuel Au-Yong-Oliveira[3] and Fernando Moreira[4]

[1] University of Trás-os-Montes e Alto Douro
5000-801 Vila Real, Portugal
sergioteixeira13@hotmail.com
[2] University of Trás-os-Montes e Alto Douro
5000-801 Vila Real, Portugal
INESC TEC (coordinated by INESC Porto), Faculty of Engineering, University of Porto
4200-465 Porto, Portugal
{fbranco, jmartins, ramiro}@utad.pt
[3] GOVCOPP, Department of Economics, Management, Industrial Engineering and Tourism
University of Aveiro
3810-193 Aveiro, Portugal
mao@ua.pt
[4] Univ Portucalense, Portucalense Institute for Legal Research – IJP
Research on Economics, Management and Information Technologies – REMIT
Porto & Universidade Aveiro, IEETA
Aveiro, Portugal
fmoreira@uportu.pt

Abstract. With the rapid growth in the use of digital platforms for the dissemination and expansion of a company's business reach, it is vitally important that startup firms are firmly aware of the options when deciding whether to adopt a particular technology because they often have low resource availability, which reduces their margin for error. In order to help these companies to adopt Digital Marketing in a more secure way by knowing the most relevant factors that they can find as concerns the adoption of technologies, this study will analyze the factors that influence the adoption of technology, identifying them initially through the systematic literature review of similar scientific works.

Keywords: Digital Marketing; Startups; IT Adoption; Literature Review.

1. Introduction

Nowadays it is mandatory for companies to have an online presence, either through an institutional website or an online store, with the ultimate goal of expanding the market associated with the sale of a product or service offered by that company. In order to make the most of this online business method, a set of techniques have been

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_9

developed to bring end users to these digital platforms, in order to trigger their purchase behaviour. These techniques, typically associated with the "digital marketing" concept, have main functions such as the analysis, planning, implementation and control of projects aimed at satisfying consumers' demand, wishes and needs, with quality, hence bringing profit to the company. Thus, digital marketing is considered to play an important role in the survival, development, and success of small and new ventures.

As digital marketing is highly correlated with new information technologies, it is expected that traditional marketing techniques and characteristics will not only serve as a basis but also will be greatly developed, in order to improve the overall business performance. Nevertheless, this principle applies mainly to new and small businesses that rely heavily on new customers and new investors in their products and services.

Considering the widespread set of perspectives on the adoption of digital marketing at the organizational level, the present work aims at identifying the most relevant determinants for the adoption of digital marketing by Portuguese startups. This output will serve, from our perspective, as the basis for future research on the topic.

2. Theoretical Framework

2.1. Digital Marketing Conceptualization

The Internet, as well as other digital technologies, played a crucial role in the progress and evolution of marketing. It allowed the existence of a wide range of products, services, means of purchase, prices, and new modes of communication that make information more quickly available to consumers. The Internet has showed companies a novel way to reach new markets, as well as having provided the opportunity to offer new products and services, using online communication techniques, placing them in the same market of larger companies when used appropriately and corresponding to the requirements of the sector concerned.

In summary, marketing functions involve the analysis, planning, implementation and control of projects designed to achieve and satisfy demand, whether with products or services, considering the desires and needs of the consumers/users, with quality and in order to make a profit for the company. Marketing is considered to be of utmost relevance in the survival, development, and success of small and new ventures. Marketing is now using technology to innovate and to enable firms to be inserted in expanding markets in numerous forms. The companies that use these new technologies will have different perspectives and objectives with their use, and it should be noted that the learning stage of digital marketing may result in a slow process [1].

According to Levitt [2] marketing "*is the process of attracting the customer*". In a similar note, Kotler [3] argues that marketing "*is the social and management process through which individuals and groups obtain what they need and desire through the creation and exchange of products and values*". A similar perspective is presented by

Skacel [4], according to whom, "*it is what we do to get more consumers to use more services or buy more products, to meet more needs and more often*".

2.2. Importance and Benefits of Digital Marketing

In sum, the functions of marketing involve the analysis, planning, implementation and control of projects designed to achieve and satisfy demand, whether with products or services, considering the desires and needs of consumers / users, with quality and in a way to bring profit to the company. Marketing is considered to be extremely relevant in the survival, development, and success of small and new ventures.

According to Chaffey, Ellis-Chadwick, Mayer and Johnston [5], associated to digital marketing are several tangible and intangible benefits. The authors consider tangible benefits to increase sales of new sales opportunities resulting in increased revenue from new customers, new markets, and existing customers (repeat sales or cross selling); cost reduction due to reduced time for customer service, online sales and reduced costs of printing and distribution of marketing communications. With respect to intangible benefits, these are presented as the communication of corporate image; enhancement/improvement of the brand; marketing communications now faster and more responsive (including public relations); improved customer service; improved learning for the future; the meeting of customers' expectations for having a website; identifying of new partners and better support for existing partners; better management of marketing information and information about the customer; more customer feedback on products.

Marketing is very important for the creation of companies, especially as they face environmental challenges, such as the rapid evolution of technology, globalization and increasingly more sophisticated competitors. As such, they need to be able to identify and pursue market opportunities, adapting to a dynamic environment, despite the limited financial and human resources that they possess [6] and the set-back of a limited market and small customer base [7]. The main tool to take advantage of these opportunities, currently, is clearly the Internet – where companies promote their products/services to a global target audience, reaching large masses at a greatly reduced cost. For this reason it is mandatory to give maximum attention to digital marketing to obtain the maximum competitive advantage that this medium provides to companies. Marketing professionals are required to constantly update their knowledge and to constantly search for information, because this sector is always undergoing constant change.

2.3. Digital Marketing in Startups

In the case of an SME (Small and Medium-sized Enterprise), the distinction between marketing and sales becomes very tenuous, since marketing in SMEs takes place during the sales process [8]. What causes that is the fact that many managers of SMEs have a wrong perception of what the sale is for Marketing [9], however it is undenaible that marketing exists to increase sales. [10] performed some tests taking as main actors the entrepreneurs themselves and what marketing means to them. As you would expect, the conclusion which was reached was that entrepreneurs have an incomplete meaning and perception of what marketing is. The majority considered marketing to be synonymous with sales or advertising, thereby emphasizing short-term goals, rather than long term profitability. Managers should see marketing as a tool for strategic planning or strategic orientation, however the vast majority of them seemed to opt for a product and sales orientation [10].

Flexibility is a key word for marketing in SMEs, since marketing is bound to evolve throughout the life cycle of the company, to better position the product in the market, and in order to satisfy the needs of customers. The flexibility and reduced response time SMEs manage in practice are a competitive advantage compared with large enterprises [9, 11-13]. SMEs practice creative marketing, of an alternative and instinctual nature even given weak financial resources. Typically, the success of innovation in SMEs is determined by the novelty, the power of new adoptions that translate into a growing number of opportunities for growth.

3. Digital Marketing Adoption

3.1. Web-Based Information Technology Adoption

There are several studies on the use of models to characterize the adoption and use of new technologies at the enterprise level. These models enable researchers to have an idea of the major determinants that must be taken into account as potential influencers in the process of incorporating the referred technologies. As argued by Karahanna, Straub and Chervany [14], these models offer valid arguments that allow managers, users, as well as the academic community, to gain a better perception of the application and potential of new technologies, with the purpose of leading to more productive modes of operation.

As digital marketing is a recent concept, especially when perceiving its application to SMEs (that have reduced resources and cannot make risky investments), there is an increasing demand for studies focusing on the adoption of the referred technology by small and resource-limited companies, in order to reach a full understanding on what might be the obstacles and determinants of the adoption process.

In order to perform a valid analysis of the existing scientific literature, a set of international scientific repositories, such as the ACM Digital Library, the IEEE Xplore Digital Library, the Web of Science, ScienceDirect, SCOPUS and Google Scholar, have all been used as the source for literature that focuses upon the adoption of web-based technologies, such as digital marketing, at the firm level.

An initial analysis of the existing literature allowed us to perceive that the majority of the existing research was drawn from more "traditional" IT adoption models, such as Rogers [15] theory on the Diffusion of Innovation (DOI), Davis, Bagozzi and Warshaw [16] technology-organization-environment framework (TOE) and, also, from the merging of these models with other existing models, such as the Institutional Theory and Iacovou, Benbasat and Dexter [17] IT adoption model.

When assessing the existing literature, we were able to identify 41 articles that directly addressed the adoption of web-based information technologies at the firm level. A direct analysis allowed to perceive that since 2014 the topic has been directly approached a number of times (Figure 1).

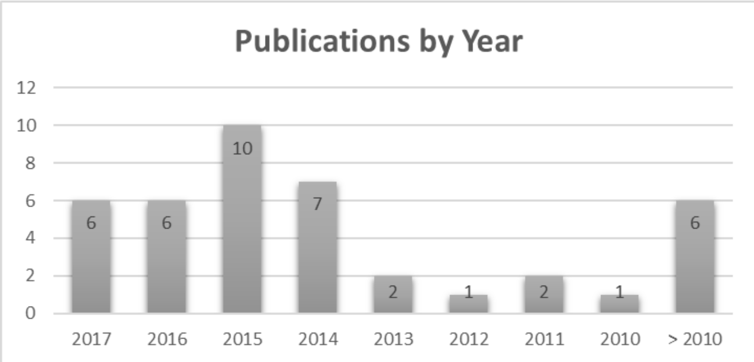


Fig. 1. Evolution of the number of publications that directly addressed the adoption of web-based information technologies at the firm level.

3.2. Digital Marketing Adoption Determinants

The adoption of digital marketing by startups is of vital importance since this technology might represent a communication tool that can be used at a small cost and with lower risk. However, any decision is not without error, hence the need to understand the variables that may influence, directly or indirectly, the referred technology adoption.

We analyzed 72 articles, on which we performed an initial filtering based on their keywords, title, methodological approach and contribution. Of this initial set, only a sub-set of 41 articles were considered relevant to our research.

Through a detailed analysis of the referred articles, we were able to identify 45 variables associated with the adoption of digital marketing by organizations. In Table 1 it is possible to view the referred variables, in descending order of the number of occurrences in the literature [11, 18-57].

A brief description of the identified variables, with the biggest number of usages in the literature, is presented below:

- **Ease of Use/Complexity**: The degree to which an innovation is perceived as easy to use and understand. New technologies that are simpler to understand are adopted faster than others which require more learning time.
- **Perceived Benefits**: Degree to which a company can take advantage of using a certain technology. Measured by the gains or improvements derived from the use of the technology, such as the recruiting of new clients, or the improved feedback provided by the customer base.

- **Cost**: This factor is also often taken into account for the implementation of new technologies. Relevant factors are how much cheaper a technology is as well as how much more easily and quickly it is implemented.
- **Perceived Industry Pressure**: This factor is connected to the enterprise's obligation to evolve to be able to maintain / increase its market power. This takes into account the entry of new operators, the threat of substitute products and services, the bargaining power of customers, as well as the power of suppliers of resources and the intensity of rivalry.
- **Expectable Performance**: The degree to which an organization believes that the use of a given technology will help the organization to achieve performance gains. The performance at the level of the organization can be translated in terms of business efficiency, performance in sales, customer satisfaction and the development of a relationship with the client.
- **External Pressure**: External pressures encompasses several factors present in the environment of the organization. These include competitors' actions, market conditions and legal regulations.
- **Compatibility**: The degree to which a given technology is compatible with the existing business processes, practices and systems, as some technologies may require significant change.
- **Government Pressure**: The level of pressure imposed by government entities through regulations and mandatory procedures.
- **Readiness**: The quantity and quality of existing (human, technical and financial) resources held by an organization that assure the capability to be ready to adopt a given technology.
- **Size**: Refers to the dimension of the organization both in terms of the number of existing employees and its business volume.
- **Business Dependence on IT**: Degree of readiness that a company has to assign the implementation of new technology to IT professionals and the measurement of the maintenance of the technology base.
- **Availability/Manager Attributes**: Degree of ease that the person responsible for the project will have to accept the implementation of new technology.
- **Customer Power/Pressure**: Degree of demand on the part of an important customer for the company to implement a certain technology. This type of pressure has a major role in encouraging companies to adopt a technology. An innovative customer could force the organization to use and adopt a new innovation.
- **Social Influences**: Degree to which an organization believes that through someone influential it is important to use a certain technology.
- **Trustworthiness**: Degree of trust that a company can expect for the implementation and use of the new technology in their business.

Table 2. Adoption determinants identified in the analysed literature.

Variable / Paper	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[35]	[34]	[38]	[36]	[37]	[57]	[39]	[40]	[41]	[42]	[43]	[44]	[45]	[11]	[46]	[47]	[48]	[49]	[50]	[51]	[52]	[53]	[54]	[55]	[56]	#
Ease of Use/Complexity			x	x	x		x	x	x	x	x	x		x	x				x			x	x	x			x	x	x	x	x					x	x				x	23
Perceived Benefits			x	x	x	x		x	x	x	x	x		x	x		x	x		x				x	x		x			x					x						x	20
Cost		x					x					x	x				x			x				x	x				x	x					x						x	12
Perceived Industry Pressure									x		x							x					x			x							x	x	x				x	x		10
Expectable Performance/Readiness	x						x											x	x				x					x	x	x		x	x		x		x					12
External Pressure											x							x				x			x						x		x	x			x			x	x	10
Compatibility			x					x	x											x		x			x		x									x	x				x	10
Government Pressure			x		x		x						x						x		x			x										x					x	x		10
Size																				x		x	x		x	x						x		x	x				x			9
Business Dependence on IT									x							x					x				x			x		x			x								x	8
Availability /Manager Attributes									x							x				x		x				x										x		x	x			8
Customer Power/Pressure					x				x								x	x			x				x							x										7
Social Influences	x	x	x				x							x														x	x					x								8
Trustworthiness		x	x	x		x								x					x																	x						7
Relative Advantage	x	x							x													x	x													x						6
Technological Infrastructure and Support					x						x		x						x														x							x		6
Top Management Support									x				x								x			x																x	x	6
Trialability									x		x		x						x								x									x						6
Expectable Effort/Organization Readiness	x																							x					x	x					x							5
Innovativeness			x			x																			x	x					x											5
Strategic Complexity																x				x	x		x								x											5
It Experience/It Ability				x									x								x				x																	4
Perceived Value for Money					x	x				x								x																								4
Normative Pressures		x												x																		x									x	4
Communication Process																x	x																						x		x	4
Technology Readiness			x												x										x							x										4
Competitive Advantage										x	x							x																								3
Educational Level													x									x									x											3
Information Intensity									x											x		x																				3
Organizational Learning Ability/Self-concept																x													x			x										3
Security						x			x					x																												3
Behavioural Intention to Use								x	x																																	2
IT Equipment																														x	x											2
Market Readiness																		x														x										2
Observability									x																		x															2
Organizational Complexity																		x													x											2
Centralization																																						x				1
External Support																									x																	1
Organizational Slack																																							x			1
Hedonic Motivation																													x													1
Information Strategy/Politics																				x																						1
Service Differentiation										x																																1
Strategic Intent											x																															1
Top Management Emphasis																																x										1
Waiting Time																													x													1

- **Relative Advantage**: The degree to which an innovation can bring benefits to an organization. Firms must have a vision that defines the benefits that will be gained by the performance of a certain technology and affecting the performance, exponential growth, financial growth and competitive advantages of the company.
- **Technological Infrastructure and Support**: Degree to which an organization believes that an organizational infrastructure and technology exists to give support to the system.
- **Top Management Support**: A large literature on organizational innovation suggests that management support is considered one of the most important factors that influence the adoption of innovation. The successful adoption of innovation requires the support of top management to integrate innovation in business as well as in other processes.
- **Trialability**: Degree to which an innovation can be experienced for a limited time.
- **Expectable Effort/Organizational Readiness**: Degree of ease of learning and of implementation of the technology.
- **Innovativeness**: Ability of a company to be open to new ideas and use of new solutions for its projects.
- **Strategic Complexity**: Degree of difficulty that a company has to define a good strategic plan to take part in a certain technology. The literature on marketing stresses the importance of strategic guidance to help companies build dynamic capabilities in environments of permanent change.

4. Results Analysis

The study allowed for the collection of a set of determinants considered critical to the adoption of digital marketing at the firm level. We were also able to acknowledge that the majority of the considered literature is drawn from traditional information technology adoption models.

During the present study, a set of 45 variables was identified as being associated to the adoption of digital marketing, of which we have selected 21, considering that these were the ones with the most representativity in the analysed literature.

Based on the results achieved it is possible to perceive that digital marketing is a vital technology for the development, sustainability and increase in performance of organizations. However, its adoption by new and small firms, such as startups, might be a complex process considering the number of determinants that should be addressed in order for the adoption process to be successful.

From the literature analysis undertaken it was also possible to perceive the existence of limited knowledge on the adoption of digital marketing at the firm level, the difficulties (highlighted by firms) in recruiting and training skilled professionals to enhance the results due to the adoption of digital marketing, and the existing decrease of the price associated with outsourcing digital marketing services that, when strategically used in aligned marketing strategies, tend to produce very positive results.

5. Final Considerations

5.1. Theoretical and Practical Implications

The present research seeks to fill a knowledge gap concerning the characterization of the digital marketing adoption process at the firm level, with a particular focus on the adoption of the referred technology by startups. During our research we were able to identify several studies focused on the adoption of similar technologies, such as e-commerce and cloud computing, that can be used as similar studies to the one we were performing herein and whose results can also be analysed and taken into consideration.

With the above in mind, the present research aims to contribute, in a positive manner, to both the development of existing knowledge on the adoption of digital marketing by startups, and to a more effective adoption of the referred technology, digital marketing, by the same set of organizations, through the identification of the determinants with the most relevancy to the adoption process.

5.2. Conclusions and Future Work

The performed literature analysis allowed us to acknowledge that the adoption of digital marketing at the firm level is a very up-to-date topic and its application to new and small enterprises, such as startups, needs to be analysed in a more detailed manner considering the lack of literature focusing on this topic.

As concerns future work, we will be performing a focus group interaction, comprising a group of ten experts, aimed at assessing their considerations on the variables identified as the ones with the most relevancy to the digital marketing adoption process at the firm level. To do this, the list of variables, identified in the present study, will be put up for discussion within the group of experts, hence aiming on filtering the sub-set of variables considered as the most relevant to the previously mentioned technology adoption process.

References

1. Hall, B.H., Khan, B.: Adoption of new technology. National bureau of economic research (2003)
2. Levitt, T.: Marketing myopia. Harvard business review 38, 24-47 (1960)
3. Kotler, P.: Reinventing marketing to manage the environmental imperative. Journal of Marketing 75, 132-135 (2011)
4. Skacel, R.K.: The marketing plan: how to prepare it, what should be in it. NTC Business Books (1990)
5. Chaffey, D., Ellis-Chadwick, F., Mayer, R., Johnston, K.: Internet marketing: strategy, implementation and practice. Pearson Education (2009)

6. Aldrich, H., Auster, E.R.: Even dwarfs started small: Liabilities ofage and size and their strategic implications. Research in organizational behavior 8, 165-186 (1986)
7. Carson, D.J.: The evolution of marketing in small firms. European Journal of Marketing 19, 7-16 (1985)
8. Oakey, R.: Innovation and the management of marketing in high technology small firms. Journal of Marketing Management 7, 343-356 (1991)
9. Mc Cartan-Quinn, D., Carson, D.: Issues which impact upon marketing in the small firm. Small business economics 21, 201-213 (2003)
10. Alberto, M., Gianluigi, G., Peluso, A.: What is marketing for SME entrepreneurs? The need to market the marketing approach. Journal of Marketing Trends-Small & Medium Enterprises, I 67-74 (2010)
11. Gilmore, A., Gallagher, D., Henry, S.: E-marketing and SMEs: operational lessons for the future. European Business Review 19, 234-247 (2007)
12. Carson, D., Gilmore, A., Perry, C., Gronhaug, K.: Qualitative marketing research. Sage (2001)
13. Heathfield, P.: SME business leaders need powerful on-board computers. Industrial Management & Data Systems 97, 233-235 (1997)
14. Karahanna, E., Straub, D.W., Chervany, N.L.: Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs. MIS quarterly 183-213 (1999)
15. Rogers, E.M.: Diffusion of Innovations: modifications of a model for telecommunications. Die Diffusion von Innovationen in der Telekommunikation, pp. 25-38. Springer (1995)
16. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User acceptance of computer technology: a comparison of two theoretical models. Management science 35, 982-1003 (1989)
17. Iacovou, C.L., Benbasat, I., Dexter, A.S.: Electronic data interchange and small organizations: Adoption and impact of technology. MIS quarterly 465-485 (1995)
18. Goswami, A., Dutta, S.: E-Commerce Adoption by Women Entrepreneurs in India: An Application of the UTAUT Model. Business and Economic Research 6, 440-454 (2017)
19. Wu, K., Vassileva, J., Zhao, Y.: Understanding users' intention to switch personal cloud storage services: Evidence from the Chinese market. Computers in Human Behavior 68, 300-314 (2017)
20. Kumar, R., Sachan, A.: Empirical Study to Find Factors Influencing e-Filing Adoption in India. In: Proceedings of the Special Collection on eGovernment Innovations in India, pp. 52-57. ACM, (Year)
21. Sun, E., McLachlan, R., Naaman, M., Tech, C.: TAMIES: A Study and Model of Adoption in P2P Resource Sharing and Indirect Exchange Systems. In: CSCW, pp. 2385-2396. (Year)
22. Hasan, M.R., Lowe, B., Petrovici, D.: Antecedents of Adoption of Pro-poor Innovations in the Bottom of Pyramid: An Empirical Comparison of Key Innovation Adoption Models—An Abstract. Marketing at the Confluence between Entertainment and Analytics, pp. 1081-1082. Springer (2017)
23. Kim, Y., Park, Y., Choi, J.: A study on the adoption of IoT smart home service: using Value-based Adoption Model. Total Quality Management & Business Excellence 1-17 (2017)
24. Engotoit, B., Moya, M.B., Mayoka, K.G., Bonface, A.: A Mobile-Based Communication Adoption Model for agricultural market information dissemination in Uganda. Global Journal of Computers & Technology Vol 5, (2016)
25. Kanchanatanee, K., Suwanno, N., Jarernvongrayab, A.: Factors Affecting the Intention to use E-marketing of Small and Medium Sized Businesses in the Three Southern Border Provinces of Thailand. International Journal of Business and Social Science 5, (2014)
26. Iqbal, T., El-Gohary, E.: An attempt to understand e-marketing: an information technology prospective. International Journal of Business and Social Science 5, (2014)

27. Abubakar, S., Ahmad, A., Umar, M.a.: E-Marketing Adoption and Competitive Advantage: Study of Commecial Banks in Nigeria. I S E R 27th INTERNATIONAL CONFERENCE, pp. 1-6. ISER, Riyadh, Saudi Arabia (2016)
28. Iddris, F., Ibrahim, M.: Examining the relationships between e-marketing adoption and marketing performance of small and medium enterprises in Ghana. Journal of Marketing and Consumer Research 10, 160-169 (2015)
29. Mochoge, O.C.: SMEs Adoption of Web-based Marketing: Empirical Evidence from Kenya. IJCSI International Journal of Computer Sciences2014 11, 2 (2014)
30. Kazungu, I., Panga, F.P., Mchopa, A.: IMPEDIMENTS TO ADOPTION OF E-MARKETING BY TANZANIAN SMALL AND MEDIUM SIZED ENTERPRISES: AN EXPLANATORY MODEL.
31. SHOTER, A.M., BATAINEH, A.Q., SALHAB, H.A.: Building a Model for Determining the Factors Affecting Mobile Marketing Acceptance and Adoption. IRMBR - International Review of Management and Business Research 5, 22 (2016)
32. BAŞGÖZE, P.: Integration Of Technology Readiness (Tr) Into The Technology Acceptance Model (Tam) For M-Shopping. International Journal of Scientific Research and Innovative Technology 2, 26-35 (2015)
33. Musa, H., Li, S.C.H., Abas, Z.A., Mohamad, N.: Adoption Factor of Mobile Marketing: The Case of Small Medium Enterprises (SMEs) in Malaysia. International Review of Management and Marketing 6, (2016)
34. Kithinji, L.W.: Internet marketing and performance of small and medium enterprises in Nairobi county. University of Nairobi (2014)
35. Mokhtar, N.F.: Internet Marketing Adoption by Small Business Enterprises in Malaysia. International Journal of Business and Social Science 6, (2015)
36. Herzallah, F., Mukhtar, M.: Organization Information Ecology and E-Commerce Adoption: Effect on Organizational SMEs Performance. Journal of Computer Science 11, 540 (2015)
37. Al-Somali, S.A., Gholami, R., Clegg, B.: Determinants of B2B e-commerce adoption in Saudi Arabian firms. International Journal of Digital Society (IJDS) 2, 406-415 (2011)
38. Albesher, A.A., De Coster, R.: The impact of IT resources on SMEs innovation performance.
39. White, G.R., Afolayan, A., Plant, E.: Challenges to the adoption of e-commerce technology for supply chain management in a developing economy: a focus on Nigerian SMEs. E-commerce Platform Acceptance, pp. 23-39. Springer (2014)
40. Al-Alawi, A., Al-Ali, F.: Factors affecting e-commerce adoption in SMEs in the GCC: An empirical study of Kuwait. Research Journal of Information Technology 7, 1-21 (2015)
41. Rahayu, R., Day, J.: Determinant factors of e-commerce adoption by SMEs in developing country: evidence from Indonesia. Procedia-Social and Behavioral Sciences 195, 142-150 (2015)
42. Wamba, S.F., Carter, L.: Social media tools adoption and use by SMEs: An empirical study. Social Media and Networking: Concepts, Methodologies, Tools, and Applications, pp. 791-806. IGI Global (2016)
43. Ndekwa, A.G.: Drivers of electronic commerce (e-commerce) among small and medium tourist enterprises (SMTEs) in Tanzania. International Journal of Science and Research (IJSR) 4, 2512-2517 (2014)
44. Fortes, N., Pereira, J.H., Costa, J.F.d.: The adoption of cloud computing services by Portuguese companies: The impact of marketing efforts. RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação 33-48 (2016)
45. Dwivedi, Y.K., Shareef, M.A., Simintiras, A.C., Lal, B., Weerakkody, V.: A generalised adoption model for services: A cross-country comparison of mobile health (m-health). Government Information Quarterly 33, 174-187 (2016)

46. Vilaseca, J., Torrent, J., Meseguer, A., Rodríguez-Ardura, I.: An integrated model of the adoption and extent of e-commerce in firms. International Advances in Economic Research 13, 222-241 (2007)
47. Wu, F., Mahajan, V., Balasubramanian, S.: An analysis of e-business adoption and its impact on business performance. Journal of the Academy of Marketing science 31, 425-447 (2003)
48. Brodie, R.J., Winklhofer, H., Coviello, N.E., Johnston, W.J.: Is e-marketing coming of age? An examination of the penetration of e-marketing and firm performance. Journal of interactive marketing 21, 2-21 (2007)
49. Tsiotsou, R.H., Vlachopoulou, M.: Understanding the effects of market orientation and e-marketing on service performance. Marketing Intelligence & Planning 29, 141-155 (2011)
50. Braun, P.: Networking tourism SMEs: e-commerce and e-marketing issues in regional Australia. Information Technology & Tourism 5, 13-23 (2002)
51. Tanakinjal, G.H., Deans, K.R., Gray, B.J.: Third screen communication and the adoption of mobile marketing: A Malaysia perspective. International Journal of Marketing Studies 2, 36 (2010)
52. Dlodlo, N., Dhurup, M.: Drivers of e-marketing adoption among small and medium enterprises (SMEs) and variations with age of business owners. Mediterranean Journal of Social Sciences 4, 53 (2013)
53. Alford, P., Page, S.J.: Marketing technology for adoption by small business. The Service Industries Journal 35, 655-669 (2015)
54. Dahnil, M.I., Marzuki, K.M., Langgat, J., Fabeil, N.F.: Factors influencing SMEs adoption of social media marketing. Procedia-social and behavioral sciences 148, 119-126 (2014)
55. Fazli, S., Shirdastian, H., Laroche, M.: Effective factors of successful cloud marketing adoption by SMEs: the case of Iran. International Journal of Business Environment 7, 415-434 (2015)
56. Maduku, D.K., Mpinganjira, M., Duh, H.: Understanding mobile marketing adoption intention by South African SMEs: A multi-perspective framework. International Journal of Information Management 36, 711-723 (2016)
57. Almoawi, A., Mahmood, R.: Applying the OTE model in determining the e-commerce adoption on SMEs in Saudi Arabia. Asian Journal of Business and Management Sciences 1, 12-24 (2011)

Financial impact on the adoption of software validation tasks in the analysis phase: A business case

David Allasi[1], Abraham Dávila[2]

[1]Escuela de Posgrado, Pontificia Universidad Católica del Perú, Lima 32, Lima, Perú
[2]Departamento de Ingeniería, Pontificia Universidad Católica del Perú, Lima 32, Lima, Perú
{dallasi, abraham.davila}@pucp.edu.pe

Abstract. Some information systems in a group of kind of organizations (for example government administrative organizations) have a critical role so they must achieve requirements in a rigorous way. However, it is not a common practice to use validation tasks, so these requirements usually do not correspond to the need. This study aims to analyze the financial impact on the adoption of practices related to validation in the phase of requirements management and analysis of software development. This study was carried out based on pre-test and post-test approach; in our case, we perform: first, an evaluation of a relevant and representative project; second, the organization introduces software validation tasks; and finally, an evaluation of a equivalence project with the improved process. We found that both project performed change during software life cycle. However, in the first one, the change was performed in the final stage, while the second project was made during the analysis stage. The requirements management and analysis process were improved and the results were observed during first stages and it is represents a 16% of saved.

Keywords: validation, requirement management, software process improvement.

1 Introduction

The most used criteria to measure the project successful are: the fulfillment of the objectives of times, cost, functionality and quality [1] [2] [3]. In particular, in software development project, the software functional requirements management is critical and it is compound of a set of activities that help to the work team to identify, control [4] and follow the requirements and it changes at any time. The well defined software functional requirements help to work team to get right scope of the project that must be implemented in the software [5] and this will allow to avoid over-costs by wrong implementation based on requirements incomplete, not understand, incorrect or out of scope [6].

The Chaos Report [7] points out that around of 29% of the projects are considered successful and in this report also show that around 6 of 10 causes why software development project fail are attributable to related to software requirements. In the Peruvian software industry, we identify the same problems descripted in Chaos report

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_10

and, based in our knowledge of industry; we believe that our indicators are worse than Chaos report.

The situation described before is usually common in different kind of organizations and different kind of software products. In particular, the software for information systems is critical in groups of kind of organizations as government administrative organizations, nonprofit, educational institution, etc. For these situations is very important that software for information system meet requirements and the requirements must keep correspondence with the organization needs, however it is not a common practice to use validation tasks, so these requirements usually do not correspond to the need.

The organization under study was called Gob.Alpha by confidential agree. In this organization the main functional mistakes are the incorrect requirements because there are user misunderstanding, lack of precision about domain rules and lack of comprehension the relations and dependences with others software of information systems. In this context we propose two fundamental questions: (i) ¿how could the requirements management process be improved to ensure that the defined scope is full and correct? (ii) ¿what is the financial impact of this improvement on the organization? For this situation we improved software process based in the introduction of validation tasks and assess their impact.

In this paper we present a study about financial impact of improving a process based on software validation practices. The paper is organized as follows: in Section 2, present a background; in Section3, show unit of analysis; in Section 4, describe the research method; in Section 5, show the results of the systematic mapping; in Section 6, present the results of the case; in Section 7, we discuss the findings and conclusions.

2 Background

In this section we present two important aspects of our study. One is related to software process improvement used and the other is related to works obtained by a systematic mapping.

Gob.Alpha is a government institution and for this reason we must use the standard ISO/IEC 12207 [8], in particular, we used software validation process (SUP.5) whose practices and results can be seen in ISO/IEC 15504-5:2012 [9]. The purpose to consider SUP.5 was include validation tasks, products or tools to help that the functional scope in critical information systems projects became most complete and correct possible [4].

The systematic mapping (SM) is a defined method to build classifications and conducting thematic analyzes in order to obtain a visual map of existing knowledge within a broad theme [10]. In this study, our research question was RQ.1. What software validation activities or techniques are used in the requirements management process? The SM was performed in two iterations. First one was performed without results about adoption experiences of practices validation in information systems in government entities. Second one was performed on the application of validation techniques in the stage of requirements. In the Appendix A show in details a SM performed to answer the research question, using PICO method (Population, Intervention, Comparison, Outputs) [11] and a taxonomy defined by Koskinen [12] to perform the classification of validation works. In Table 1, it is displayed the results of SM performed classified according to [12]. As shown in the results, Pre-review is the most used validation technique in the management of requirements, so we include this validation technique as an improvement of the process in the case study carried out.

Table 1. Results (Absolute y Relative) classified

Class / Sub-class	#	Percentage
Technical view	(18)	100%
Pre-review	6	33.3
Requirements Inspection	1	5.6
Inspection based on test cases	2	11.1
Requirements Reading	5	27.8
Prototyping	1	5.6
Based on models	1	5.6
Based on user point of view	1	5.6
Based on tests	1	5.6

3 Unit of Analysis

The definition of unit of analysis corresponds, according to Piattini [13] quoting Benbasat, the single type and in this particular case to a small company as a unit of analysis. The following sections describe the organization in the pension sector, the business of this company and the description of the projects selected.

- Business area of selected organization. The Gob.Alpha is an institution that belongs to the pension sector in Peru, which began operations in June 1993 [14]. Its administration is regulated by the standards rules by the Superintendence of Banking and Insurance (SBS), which must be strictly fulfil with.
- A project development team was defined and it was formed by: (i) two functional analysts, (ii) two functional users, (iii) one technical analyst of Gob. Alpha, and (iv) two functional analyst's provider. There was also a research team made up of the authors of this study.
- Organization description. The purpose of pension sector organizations is to manage the pension fund of the members of their portfolios in order to provide them with a pension in accordance with the law [15]. SBS constantly issues new standards that allow regulate current standards, and create, modify, or remove existing standards. All this evolution of the legal framework, that makes the business logic that must handle the information systems on pension sector, is complex and often creates problems in the definition of scope in new software developments. It was therefore decided to use an institution in this sector for the present case study.
- Description of the considered projects. The evaluation of the current process and the improved process, from the requirements management and analysis stage, in relation to the over-costs incurred by requests for changes, was made in high-

impact software development projects, which were: (i) implementation of the new regulations on the creation of a required fund type 0 (zero risk) [16] and (ii) implementation of the new 95.5% refund rule for its members Age 65 [16], called projects Alpha1 and Alpha2 respectively. Both projects involve a very critical and sensitive operating module such as the module of pensioner benefits and pension procedures. In Glob. Alpha, a project is considered high impact when the number of man hours of development exceeds 200 hours. Besides, for this type of project, it always handles a reserve of 8% of the total project cost for change requests.

4 Research Method

To carry out this study we use partially two research models. We define some activities based on the pre-test and post-test approach for the technical part and we use case study structure and philosophy for perform and reporting our research. In Table 2, it is presented the phases of our research based on the recommendation of Runeson [17] and Piattini [13] used in the ProCal-ProSer Project.

Table 2. Phase and activity in case study scheme adopted by ProCal-Proser. [18]

Phase	Responsibility
P1	(P1-1) Objective of the study
	(P1-2) What is being studied
	(P1-3) Theory acquired
	(P1-4) Research Questions
	(P1-5) Collections Methods
	(P1-6) Selection of data
P2	(P2-1) Definition of how harvesting is performed
E1	(E1-1) Data collection
E2	(E2-1) Data analysis
	(E2-2) Interpretation of data
R1	(R1-1) Reporting

In the first part of this study we remark that in order to reduce bias from projects characteristics, it was established to work with two similar projects (in pre-test and post-test approach). In the follow paragraphs we use two perspectives to show our activities: process level and product level; and the results are displayed in Section 5.

Since process perspective, for the initial technical stage (pre-test) evaluation of a completed project selected was carried out, the number of re-work actions was identified and one assessment (1st assessment) of requirements management process was performed using ISO/IEC 15504 [9]. After reporting results of first assessment, some validation tasks are introduced in the requirements management process, see the original and final process diagram in Appendix B. Also, personnel were trained and a new project was coordinated to prove the effect of the changes made. For the final technical stage (pro-test) the second assessment was performed and identified projects characteristics need for comparing both projects.

Since product perspective, in the initial phase, we used a metric developed for this research to measure the percentage of functional mistakes detected, classified as incomplete or incorrect functional definition. This metric was performed in the product testing and certification stage. These errors were considered because they generate change requests within the initial scope of the project, thus affecting the budget because the mistake was detected in a final stage of the project.

The financial impact was evaluated by the over-cost that is obtained in both projects. This over-cost was generated by the solution of defects found in the tests stage and the changes due to the ambiguity of the definition made.

5 Results and analysis

In this Section, we present the following: (i) the result of the diagnostic evaluation of the current requirements management and analysis process (ii) the results obtained from the evaluations to the projects indicated in the previous point, and (iii) the findings found.

5.1 Diagnostic evaluation of the current process

We perform the diagnostic evaluation of the current Process of Requirements Management and Analysis under the protocol defined in the ISO / IEC 15504-5: 2012 for the SUP 5. Software Validation (see Table 3).

From the diagnostic evaluation carried out and the results obtained in the systematic mapping performed the following validation activities were adopted within the requirements management and analysis process.

- Validation of the functional scope using a checklist of critical points defined for each module based on the expert judgment of the functional analysts of both parties (customer and supplier).
- Checklist feedback at the end of each project with the inclusion of some critical points that could be detected based on errors by incorrect or incomplete definition of the functional scope found in the QA testing process. This activity is added to the post-production support process.

5.2 Metrics for evaluation

The metric defined has as purpose to evaluate the defined functional scope in requirements analysis and management. This metric is calculated by percentage of functional mistakes detected in the stage of product test and certification, regarding the total of functional errors found (see Appendix C). This metric will allow us to observe the impact in time and cost that could have the system due to a poor definition of the scope. It is observable since these ends up in change requests.

5.3 Comparison of characteristics of the projects

In Table 4 we compared the initial characteristics of the projects studied before the improvement (Alpha1) and the project studied after the improvement (Alpha2).

Table 3. Diagnostic Evaluation in SUP 5. Software Validation

Base Practice	Assessment of the practice based in the current process
SUP5.BP1. Develop a validation strategy	In the review of the RFP (request for proposal) delivered by the client there is no global validation strategy since only a functional validation is performed only at the level of the requested requirements and not in a global context of all the legal regulations that applied, That these requirements could give rise to gaps in other functionalities.
SUP5.BP2. Develop validation criteria	The validations that are made in the reviews of the RFP document and observations are made to expert judgment of the analyst in charge and there are no criteria or guidelines that allow a standardized validation that contemplates common points for that type of business whose norm is complex.
SUP5.BP3. Carry out validation activities	The validation activities are done but without a formal technique, only under expert judgment.
SUP5.BP4. Identify and record problems	The problems that can be detected in the reviews of the RFP documents are not registered in any file or database, they only feed the judgment of an expert, which could generate that in case of an eventual departure of the expert, and these problems can be given in Future projects.
SUP5.BP5. Provide data resulting from the completion of the validation activities	This point, if it is finally fulfilled in the solution document that is delivered to the client, contrasts that all the functional requirements requested in the RFP are included there.
SUP5.BP6. Provide validation results available to the client and other interested parts.	Due to the problems found are not registered in a repository, the client does not have access to them and only the functional analyst of the client that participates in the meetings increases his expert judgment, generating the same risk that occurs in point SUP5.BP4

Table 4. Alpha1 & Alpha2 projects: stages, efforts (man-hours) and costs

	Alpha 1		**Alpha 2**	
Stage	**M/H**	**Total (S/.)**	**M/H**	**Total (S/.)**
Project management	80	15,840	40	7,920
Requirements Manag. &Analysis	80	10,560	60	7,920
Development	640	73,920	580	66,990
Testing QA	360	17,820	260	12,870
Post Production Support	60	5,940	50	4,950
Total		990		100,650

5.4 Evaluation of Alpha1 project

The results of the product testing and certification stage for this requirement are in Table 5.

Table 5. Summary table of functional errors detected in the alpha 1 project

Module	Number of functional mistakes of the system	Number of functional mistakes by functional scope	Number of functional mistakes per construction
Collection	5	0	5
Pension procedures	35	12	23
Pension payments	12	0	12
Change of Fund	8	4	4
Membership	3	0	3
Transfers	2	0	2
Total	65	16	49

From Table 5 the % of functional errors can be determined by incorrect or incomplete definition of the functional scope. PorcErrorAlcance = (NEFA) / NEF where NEFA = 16 and NEF = 65, then PorcErrorAlcance is 24.62%, errors that do not apply are not considered.

The percentage of mistakes by incorrect or incomplete definition of the functional scope is high, and as can be seen, was basically presented in 2 modules: (i) Pension Procedures and (ii) Change of Fund.

5.5 Evaluation of Alpha2 project

As part of the improvement included in the process, a checklist of critical points was developed and used in the process of validation (pre-revision) of the scope. The checklist is shown in Table 6.

Table 6. Functional validation checklist

ID	Point to validate	Module	¿Is it within range?	¿Does it need to consult the SBS?
1	Stock Cases (applications in progress)	Pension procedures	Yes	No
2	Operational Codes	Return to affiliate	No	Yes
3	Withholding of 4% for the payment of pension.	Pension payments	Yes	Yes
4	Special flows of pension procedures (Special schemes)	Pension procedures	No	Yes
5	Interface with accounting and cash satellites	Return to affiliate	Yes	No
6	File upload Contingency	Return to affiliate	Yes	No

As a result of the revision of this checklist, certain actions were taken in the functional scope precisions document (see Table 7).

Table 7. Validation results with the checklist

ID	Point to validate	Module	Action made in the document of scope precisions.
1	Stock Cases (applications in progress)	Pension procedures	No accuracy was realized because it is already in range.
2	Operational Codes	Return to affiliate	It was defined not to include the new operational codes for these movements because the SBS would define them after the norm goes into effect. The code 9C87 will be temporarily handled
3	Withholding of 4% for the payment of pension.	Pension payments	It is specified that the remaining 4% of the 95.5% returned will not be considered for the pension payment and must be "frozen" in the same account.
4	Special flows of pension procedures (Special schemes)	Pension procedures	Consultation with the SBS is made and specifies that for the moment only the formalities REJA (Retirement by Unemployment)
5	Interface with accounting and cash satellites	Return affiliate	No accuracy was realized because it is already in range.
6	File Upload Contingency	Return affiliate	No accuracy was realized because it is already in range.

The results of the product testing and certification stage for this requirement are in Table 8.

Table 8. Summary table of functional errors detected in the alpha 2 project

Module	Number of functional mistakes of the system	Number of functional mistakes by functional scope	Number of functional mistakes per construction
Pension procedures	8	0	8
Pension payments	4	0	4
Return to affiliate	10	2	8
Total	22	2	20

With these precisions, the solution document was elaborated and the functionalities requested in the RFP were implemented and limited in the document of functional scope precisions. From Table 8 and using de metric defined before PorcErrorAlcance = 9.09 % (NEFA = 6 and NEF = 22). The percentage of errors by incorrect or incomplete definition of the functional scope is low, and as you can see, it was basically presented in a module: Return to the member.

5.5 Analysis of obtained results.

In a comparative way we can say that in the Alpha 1 project was requested two changes and its impacts are shown in Tables 9; and in the Alpha 2 project was requested only one change and its impact is shown in Table 10. The total over-cost in Alpha 1 project for these changes were 20,955 which represents 16.9% of the initial budget. The total over-cost in Alpha 2 project for these changes were 10,164 which represents 10.1% of the initial budget. In the study carried out, it was not considered that the users involved in the project could be more predisposed due to the improvement. This could be a threat in the validity of the obtained result, generating that it is smaller than the one obtained in this study.

Table 9. Data and costs of change request 1 & 2 (CR1 & CR2) of the alpha1 Project

Stage	CR 1		CR 2		Total CR
	M/H	Total (S/.)	M/H	Total (S/.)	(S/.)
Project management	10	1,980	10	1,980	3,960
Requirements Manag. &Analysis	20	2,640	10	1,320	3,960
Development	60	6,930	25	2,887.5	9,817.5
Testing QA	15	742.5	10	495	1,237.5
Post Production Support	10	990	10	990	1,980
Total	115	13,282.5	65	7,672.5	20,955

Table 10. Data and costs of change request (CR) of the alpha2 Project

Stage	Total M/H	Total amount S/.
Project management	10	1,980
Requirements Management and Analysis	15	1,980
Development	40	4,620
Testing QA	12	594
Post Production Support	10	990
Total	87	10,164

6 Discussion, Conclusions and Future Work

In concordance with to the theory of the software process improvement, the introduction of validation task in software requirements was successful and the organization saved money. Some related ideas are:

The defining functional scope is very critical issue in any kind of project and, in according with this experience, introduces validation tasks or techniques are beneficial or positive for the software team, users and other stakeholders.

Before this experience the common idea was that validation tasks or techniques because were complex and applicable in critical system (not information systems) and now developers accept introduce another validation tasks or techniques.

In this experience, the adoption of validation tasks became a very simple project of software process improvement and the software team accepted these changes. All stakeholders made grateful comments about this effort.

A problem found in the literature review was that in the last 10 years there are a lot of studies or research of validation in critical systems (for example medical or military context) but nothing in information system for administrative organization.

Calculate financial impact of this effort was very important for all members that believed in validation tasks. Also, using numbers was easier showing the benefits to the other members and the main managers.

The client also supported this kind of improvement because his users waste less time in the process of adoption the components or systems.

The user perception, by showing functional validation tools, is more comfortable because they feel that the software developer area, which is a support area, helps them to avoid mistakes instead of blaming them for it.

In custom systems, there is little that the customer can do in the event that the supplier fails to meet the time and cost commitments and may incur higher costs in post-production activities for unplanned delays or rework [19]. By means of this study we can provide that applying validation techniques in the stage of requirements management in customized systems helps to reduce this risk, since the change requests and the cost of the same will be minimized.

The investment cost of adopting these validation tasks in the requirements management and analysis stage was much lower than the over-cost that could be generated by not using it. This situation allowed this improvement to be included in the process without causing projects get out of their budgeted costs.

After this effort, many stakeholders asked question about next steps. This is a better condition to start other project to generalization this validation tasks. Also, researchers consider relevant extend this study in other projects to collect more data; introduce another validation tasks to increase the product reliability in the first stage of the software development project using a research protocol.

Acknowledgments. This work is framed within ProCal-ProSer Contract N° 210-FINCYT-IA-2013 (Innovate Perú): "Productivity and Quality Relevance Factors in small software development or service organizations adopted ISO standards", the Software Engineering Development and Research Group and the Department of Engineering of Pontificia Universidad Católica del Perú.

References

1. B. S. D. M. A. Anda, «Variability and reproducibility in sowftare engineering: a study of four companies that developed the same system,» IEEE Trans. Softw. Eng. 35, pp. 407-429, 2009.
2. R. Atkinson, «project mangement: cost, time y calidad,» J. Proj. Manage., pp. 337-342, 199.
3. K. K. A. El Emam, «A replicated survey of IT software project failures.,» IEE Software, 25, pp. 84-90, 2008.

4. N. Z. Sourour Maalema, «Challenge of validation in requirements,» ScienceDirect, pp. 1-7, 2016.
5. P. A. F. Julio Cesar Sampaio do Prado Leite, «Requirements validation through viewpoint resolution.,» IEE Trans. Softw. Eng., p. 12, 1991.
6. I. S. Gerald Kotonya, «Requirements Engineering Processes and Techniques.,» John Wiley & Sons., 1998.
7 S. &. W. S. Hastie, «Standish Group 2015 Chaos Report - Q&A with Jennifer Lynch,» 15 October 2015. [En línea]. Available: http://www.infoq.com/articles/standish-chaos-2015. [Último acceso: 15 June 2017].
8. I. 12207:2008, Systems and Software Engineering Software Life Cycle Processes, 2008.
9. I. 15504-5:2012, Information technology - software process assessment - part 5: an exemplar process assessment model, 2012.
10. R. F. a. S. M. K. Petersen, «Systematic mapping studies in software engineering,» Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, pp. 1-10, 2008.
11. S. PW, «Popping the (PICO) question in research and evidence-based practice,» Appl Nurs Res, nº 197-8, 2002.
12. S. K. a. J. Koskinen, «Survey of Software Inspection Research,» Open Software Engineering Journal, pp. 15-34, 2009.
13. M. G. B. J. A. C. Mario G. Piattini, Métodos de investigación en ingeniería del Software, Malaga: Ra-Ma, 2014.
14. J. Rojas, «La rentabilidad del sistema privado de pensiones en el Perú: 1993-1997,» p. 3, 1998.
15. A. e. a. Medina Giacomozzi, «Efecto sobre la rentabilidad que tiene para el afiliado la comisión cobrada por las administradoras de fondos de pensiones,» Elsevier, pp. 24-33, 2013.
16. Andina, «RPP,» 29 marzo 2016. [En línea]. Available: http://rpp.pe/economia/economia/afp-este-1-de-abril-entra-en-vigencia-el-fondo-cero-noticia-949428. [Último acceso: 5 julio 2016].
17. P. R. Martin Host, «Guidelines for conducting and reporting case study research in software engineering,» Empirical software engineering, vol. 14, nº 2, p. 131, 2009.
18. A. Dávila Ramón y M. Pessoa, «Factors driving the adoption of ISO/IEC 29110: A case study of a small software enterprise» de XLI Latin American Computing Conference, Arequipa, 2015.
19. Y. Angus G, Software crisis, what software crisis?, Stirling: IEEE, 2009.

Appendix

https://drive.google.com/open?id=0B_K_llz5juqPNkhESFVfYVo3SWs

Multiple Software Product Lines: applications and challenges

Guadalupe Isaura Trujillo-Tzanahua[1] Ulises Juárez-Martínez[1], Alberto Alfonso Aguilar-Lasserre[1] and María Karen Cortés-Verdín[2]

[1]Division of Research and Postgraduate Studies. Instituto Tecnológico de Orizaba, Veracruz, México.
[2]Facultad de Estadística e Informática. Universidad Veracruzana, Xalapa, Veracruz, México.
[1]gtrujillot@ito-depi.edu.mx, ujuarez@ito-depi.edu.mx, aaguilar@itorizaba.edu.mx
[2]kcortes@uv.mx

Abstract. The goal of a software product line is to create a suitable platform for fast and easy production of software for same market segment. However, a software product line is limited because it needs to meet new stakeholder requirements either through upgrades or the introduction of new technologies. A Multi Product Line aims at deriving new software products from reuse of a set of features provided by several heterogeneous software product lines without modifying or altering the independent operation of the same. This paper presents a study about the application of Multi Product Lines in the software development process. It shows some domains that illustrate applications of multi product lines principle in the process and the product. Also, the main current challenges in applying multi product line in software engineering are described. This paper aims to show the importance and usefulness of applying multi product lines approaches in Software Engineering.

Keywords: Software Engineering, Software Product Lines, Multi Product Lines.

1 Introduction

Product lines are successfully used in both software and non-software domains such as automotive, metallurgy and manufacturing to support systematic reuse. A classic example is automobile manufacturing consisting of creating variations of a single car model with a set of parts and a factory specifically designed to configure and assemble such parts. Software Product Lines (SPL) are analogous to industrial manufacturing, in which similar products are configured and assembled from reusable prefabricated parts for fast and easy software development focused on a specific market.

The Software Engineering Institute (SEI) defines a Software Product Line as a "set of software-intensive systems that share a common, managed set of features satisfying the specific needs of a particular market segment or mission and that are developed

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_11

from a common set of core assets in a prescribed way" [1]. Software Product Lines is a means to improve the processes of software development by reducing the cost and substantially improving the productivity and the quality of the products developed. For the development of a Software Product Line, a key element is the analysis, specification and management of common and variable elements within the set of products produced by the SPL.

Frequently, Software Product Lines need to meet the changing requirements of the market either by functionality, focus or technology and consequently, it is necessary to add and configure new products in the SPL. However, experts in the development of SPL recognize that they are limited and that in most cases, it is impossible to extend or adapt the platform. An alternative solution is a Multiple Software Product Line which derives new software products from the reuse of a set of features provided by different heterogeneous software product lines without modifying or altering the independent operation of the same. The purpose of this study is to examine Multi Product Lines applications in software engineering.

This document is organized as follows. Section 2 presents the research methodology. In section 3, basic concepts on Multi Product Lines are explained. Application of Multi Product Lines on different domains is shown in section 4. Challenges faced by multi product lines in the field of software engineering are presented in section 5. Section 6 presents the discussion. Finally, conclusions are presented in section 7.

2 Research methodology

The methodology is composed of three stages. The first stage was the research of works related to Multiple Software Product Lines in several databases of scientific journals. The second concerned the classification of these works in the different fields of knowledge. Finally, the third stage of the methodology involved the report of detailed literature review that identifies challenges addressed and technologies, tools and programming languages used in the implementation of the proof of concept applications in some of the reported works. For this review, we first searched in the major databases of electronic journals for a comprehensive bibliography of relevant research of MPL. The digital libraries considered were: (1) ACM Digital Library, (2) CiteSeerX, IEEE Xplore Digital Library, IGI Global, ScienceDirect (Elsevier), Semantic Scholar and SpringerLink.

3 Application of Multi Product Lines on different domains

Nowadays, there are many examples that illustrate the original idea of using MPLs, among which stand out mainly in the metallurgy, steel industry and mechatronics systems.

3.1 Metallurgy

In Metallurgy, alloys are a good example to achieve special desired properties by combining different metals. For copper to be versatile, its characteristics are modified through mixing with other metals depending on the desired end use. From this mixture, it is possible to obtain more than 400 alloys, such as bronze, brass, alumni bronze. Another alloy identified is derived from nickel and steel and is known as Invar or FeNi36. Invar is used in the manufacture of precision parts such as watchmaking, apparatus of physics, the valves of engines, among others and in instruments for measuring length such as those used in topography due to its small coefficient of expansion. This notion of combining, integrating or composing different approaches motivates the approach called Invar (Integrated View on Variability) [2, 3] for the development of software using multiple product lines. Invar facilitate the exchange of heterogeneous models of variability during the configuration of the product regardless of the techniques, notations and tools used in the organization.

3.2 Steel Industry

Another example of implementation of MPLs occurs in the Steel industry, specifically in the mini-mills that make up the product portfolio of SIEMENS VAI [4–7]. Unlike the integrated steel mill, the mini-mill is a facility that produces steel products using scrap steel as an iron source. A mini-mill is integrated by several product lines such as the electric furnace, the caster, the rolling mill, and the maintenance and setup system (MSS), a software tool used by customers for customizing the mini-mill software solution during operation. Although the mini-mill is subject to the same requirements that the integrated steel mill differs in that the plant is flexible with ability to be upgraded technically, diversity in the styles of management, labor relations and markets for the product. A mini-mill and the different subsystems can be customized in terms of the amount of iron, furnace type, number of filaments, type of lamination train or lamination capability.

3.3 Mechatronics systems

Mechatronics combines various disciplines as mechanical engineering, electronics, automatic control and software for the design of products and processes. Mechatronics systems have several applications: robotics, aeronautics, automotive industry, medical industry, home automation, among others. In the field of mechatronics, products are described by multiple models belonging to different engineering domains such as mechanical, electrical, and electromechanical. In medical domain, multiple product line development is identified for Philips Medical imaging systems through hierarchical product lines [8]. Typically, several product lines are available because products are developed in different parts of the world, and within different product groups such as magnetic resonance, X-ray, and ultrasound tomography. To handle the complexity in product lines for mechatronic systems, in [9] is proposed a MPL approach to distinguish between software and hardware by using different feature models for each.

Table 1 shows the analysis of four different domains of MPL application reported in the literature. This table highlights the main similarities and differences in the MPL applied to the industry for the manufacture of physical products and software products. It is important to note that in an MPL applied to the manufacturing industry, production capacity is considered because it defines the competitive limits of the company since the quantity of products or services that can be obtained by a period of time depends on demand and company infrastructure. While the capacity of an MPL in the context of software is unlimited because it can generate n software products. Both industrial or software MPLs agree that it is possible to involve different suppliers, obtain different versions of products (variability) and reuse processes and tools. Another aspect to highlight in MPL in the context of software is that similar to Software Product Lines not only takes care of generating software product but also generates requirements, code, architecture, tests, documentation.

Table 1. Analysis of application domains

Element	Domains			
Domain	Metallurgy	Steel Industry	Mechatronics systems	Software Engineering
Assets	Metals	Scrap steel	Hardware Software	Software artifact and such as requirements, feature model, architecture, libraries, test.
Reuse	Process, machines and tools	Process, machines and tools	Process, machines and tools	Process and software tools
Providers	Several	Several	Several	Several
Variability	Yes	Yes	Yes	Yes
Production Capacity	Limited by infrastructure	Limited by infrastructure	Limited by infrastructure	Unlimited but depends on product configuration
Line Balancing	Yes	Yes	Yes	Yes
Application	Alloys	Mini-mill	Philips Medical Imaging Systems, Aselsan REHIS	SECO, SOS, ERP, Software Supply Chain

MPL is a multi-domain approach that arises to develop large and complex systems through several independent product lines which are developed by several organizations with different approaches and technologies for different geographic areas and in any context. This approach is investigated for the development of Systems of Systems (SoS) [5, 10] and Software Ecosystem (SECO) [11, 12] that result from the integration of several operationally independent systems. The feasibility of the approach and its implementation could be useful for implementation of ERP systems [2] or software supply chains[13].

4 Multiple Software Product Lines

Software artifact reuse from different software product lines, referred to as multi product lines [14, 15] is addressed by several authors [16–18]. The reuse and composition of multiple software product lines is also known as Nested Software Product Lines [19, 20], Hierarchical Product Lines or Composite Product Lines.

A Multiple Software Product Line (MSPL) also called Multi Product Line (MPL) is a software product line that results from combining components or products developed from several independent and heterogeneous software products lines [15]. It means that software product lines are provided by different organizations and use diverse approaches and technologies.

According to Holl [14], an MPL is ”a set of several self-contained but still interdependent product lines that together represent a largescale or ultra-large scale system”.

Multi Product Line configurator is an assembly entity that is responsible for controlling and reusing the artifacts of software product lines according to the needs of stakeholders (see Fig1). Given the need to combine, integrate or compose different software product lines, the inclusion of different approaches to model variability, annotations and tools is detected.

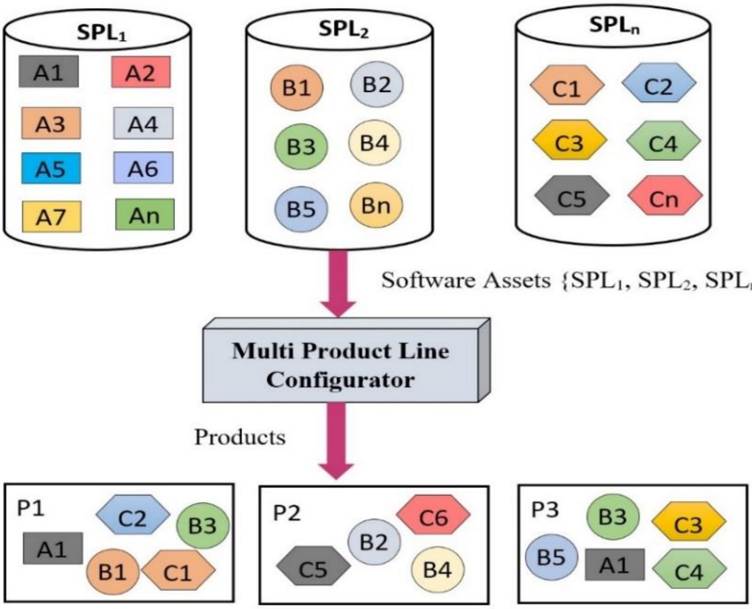


Fig. 1. Multi Product Lines

The decision of using a traditional approach or mass production through SPL or MPL for software development depends on requirements such as: reusing components, implementing non-functional requirements, launching or modifying a product to focus on a different market segment, acquisition of a competitive advantage, incorporation of new technologies (hardware, software), resource sharing, among others [21].

5 Challenges

Some of the general challenges faced by multi product lines in the field of software engineering are presented below.

5.1 Software Product Lines Reuse

Reuse in Multi Product Lines refers to the process of deploying or upgrading software systems using existing software assets across multiple software product lines. This results in a composition of SPLs that facilitates to reuse SPLs within other SPLs. Consequently, techniques, methods or approaches are required to facilitate and maximize the reuse of artifacts from software product lines to generate valid product families in a Multi Product Line [22–24]. Likewise, the fusion and reuse of feature models provided by different companies or suppliers are required. The degree of reuse depends on the scope of the available SPLs, so for the development of an MPL not only reuse the implementation but it is possible to reuse processes, tools, requirements, tests, technologies.

5.2 Interoperability between software product lines

Interoperability refers to the ability of two or more systems or components to exchange information and use the information exchanged. In an MPL, the interoperability and the integration of software product lines and their products must be facilitated [24, 25]. An adequate interoperability is useful to promote cooperation between independent systems in order to integrate them as a System of Systems [5, 10] or Software Ecosystem [11, 12] and provide more complex functions.

5.3 MPL Reference Architecture

Reference Architecture is a type of software architecture that captures knowledge and experience about how to structure architectures of software systems in a domain. Its purpose is therefore to be a guidance for the development, standardization, and evolution of systems of a single domain or neighbor domains for example Autosar (AUTomotive Open System Architecture Open Systems) for the Automotive sector. One or more reference architectures could be used as a basis of MPL. In an MPL, is necessary to define a reference architecture or consistent MPL architecture [26] that represents the common and variable artifacts of the software product lines available for incremental product development.

5.4 Automating product derivation in MPL Engineering

Product derivation in an MPL refers to complete process of building a product through the software assets of many software product lines. Automating the derivation of the product means automating this building process. The subprocesses of product derivation are product configuration and product generation.

Product configuration in an MPL is a process of collaboration between different people and teams with different knowledge regarding the domain, notations and programming languages used in the development of software product lines. Problems arise at the moment of making the decisions to assign the interested parties in the right order and at the right time based on the knowledge of the domain of the people.
Generating products in an MPL relies on a configurator to make decisions, taking as input the configuration to build the final product (or parts of it, such as executables, documentation, tests, and so on).

5.5 User Guide on Product Derivation

During the derivation of products in an MPL, it is necessary to ensure that the multiple users involved in the configuration and generation of products are aware of the chosen decisions [4]. For this reason, users need to be aware of the variability and dependencies between software product lines.

5.6 MPL tools

Supporting tools offer the developers a complete environment for development and maintenance of software product line, aiming at facilitating its adoption. Although there is a huge variety of tools for software product lines development, it is not possible to ensure that all needs of SPL engineers are being fulfilled. It is necessary to better investigate the scope, the availability and the utility of these tools for MPLs development. For this reason, SPL engineers need to adapt and extend of software product line tools for feature modeling, SPL composition, variability management, configuration and derivation of products in an MPL. In addition, tools are required to facilitate automated analysis of dependent features models.
Another aspect that stakeholders need to consider for MPL implementation is the use of different approaches in individual software product lines such as Feature-oriented Programming (FOP), Aspect-Oriented Programming (AOP) and Delta-oriented Programming (DOP).

6 Discussion

Multi Product Lines are successfully used in domains such as metallurgy, steel industry, mechatronics systems and software (Section 3). The results of this study confirmed that Multi Product Lines approach is a feasible option for software mass customization. This is because the principles used in other areas mainly in the metallurgical industry, steel industry, mechatronic systems are completely applicable to software development in order to optimize the individual systems development by taking advantage of their common features and managing their differences.

In the software industry, MPLs emerge as a flexible and viable development paradigm that enables companies to enhance their products from reusing and mass customization to a rapid market introduction, reducing costs and maximizing quality of products.

Currently, several approaches and proposals that support the development of software using MPLs are reported. However, as mentioned in Section 5, MPLs in software context present certain limitations and challenges that need to be addressed. Also, several areas of opportunity were identified in this field of research since the use of MPL depends mainly on the capabilities of the development team, market and technologies.

7 Conclusion

This study addresses the main aspects necessary to understand the importance of multi product lines and its application in the field of software engineering. Multi product line is an area with a great potential of applications in the field of software engineering, particularly in the study of software development processes. In addition, the different challenges and problems faced by multi product lines are identified.

The current study on Multi Product Lines allowed to detect that despite the experience that has in the industry, the Software Engineering requires to meet certain challenges that arise when increasing the complexity of the systems, by addition of new products, make updates by technology or change of requirements, all this in order to meet the needs or expectations of the market. Also, several areas of opportunity are detected. This work serves as a basis for future research on Multi Product Lines.

For the future work, we aim at contributing to mature the area of Multi Product Lines and propose means to better explore the software product development using this approach.

References

1. Clements, P., Northrop, L.: Software Product Lines: Practices and Patterns. (2001).
2. Dhungana, D., Seichter, D., Botterweck, G., Rabiser, R., Grünbacher, P., Benavides,

D., Galindo, J.A.: Configuration of Multi Product Lines by Bridging Heterogeneous Variability Modeling Approaches. In: Proceedings of the 2011 15th International Software Product Line Conference. pp. 120–129. IEEE Computer Society, Washington, DC, USA (2011).

3. Dhungana, D., Seichter, D., Botterweck, G., Rabiser, R., Grünbacher, P., Benavides, D., Galindo, J.A.: Integrating Heterogeneous Variability Modeling Approaches with Invar. Proc. Seventh Int. Work. Var. Model. Software-intensive Syst. 8:1--8:5 (2013).
4. Rabiser, R., Grünbacher, P., Holl, G.: Improving Awareness during Product Derivation in Multi-User Multi Product Line Environments. In: Proceedings 1st Int'l Workshop on Automated Configuration and Tailoring of Applications (ACoTA 2010) in conjunction with 25th IEEE/ACM Int'l Conference on Automated Software Engineering (ASE 2010), Antwerp, Belgium, September. pp. 1–5 (2010).
5. Holl, G., Elsner, C., Grünbacher, P., Vierhauser, M.: An Infrastructure for the Life Cycle Management of Multi Product Lines. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing. pp. 1742–1749. ACM, New York, NY, USA (2013).
6. Holl, G., Thaller, D., Grünbacher, P., Elsner, C.: Managing Emerging Configuration Dependencies in Multi Product Lines. In: Proceedings of the Sixth International Workshop on Variability Modeling of Software-Intensive Systems. pp. 3–10. ACM, New York, NY, USA (2012).
7. Holl, G., Grünbacher, P., Elsner, C., Klambauer, T., Vierhauser, M.: Constraint Checking in Distributed Product Configuration of Multi Product Lines. In: 20th Asia-Pacific Software Engineering Conference (APSEC). pp. 347–354. IEEE Computer Society, Washington, DC, USA (2013).
8. van der Linden, F., Wijnstra, J.G.: Platform Engineering for the Medical Domain. Presented at the (2002).
9. Brink, C., Peters, M., Sachweh, S.: Configuration of Mechatronic Multi Product Lines. In: Proceedings of the 3rd International Workshop on Variability & Composition. pp. 7–12. ACM, New York, NY, USA (2012).
10. Klambauer, T., Holl, G., Grünbacher, P.: Monitoring System-of-Systems Requirements in Multi Product Lines. In: Doerr, J. and Opdahl, A. (eds.) Requirements Engineering: Foundation for Software Quality. pp. 379–385. Springer Berlin Heidelberg (2013).
11. Urli, S., Blay-Fornarino, M., Collet, P., Mosser, S., Riveill, M.: Managing a Software Ecosystem Using a Multiple Software Product Line: a Case Study on Digital Signage Systems. In: Euromicro Conference series on Software Engineering and Advanced Applications(SEAA'14). pp. 344–351 (2014).
12. Schmid, K., Eichelberger, H.: EASy-Producer: From Product Lines to Variability-rich Software Ecosystems. In: Proceedings of the 19th International Conference on Software Product Line. pp. 390–391. ACM, New York, NY, USA (2015).
13. Hartmann, H., Trew, T.: Using Feature Diagrams with Context Variability to Model Multiple Product Lines for Software Supply Chains. In: Proceedings of the 2008 12th International Software Product Line Conference. pp. 12–21. IEEE Computer Society, Washington, DC, USA (2008).
14. Holl, G., Grünbacher, P., Rabiser, R.: A Systematic Review and an Expert Survey on

Capabilities Supporting Multi Product Lines. Inf. Softw. Technol. 54, 828–852 (2012).
15. Rosenmüller, M., Siegmund, N.: Automating the Configuration of Multi Software Product Lines. In: Proceedings of Fourth International Workshop on Variability Modelling of Software-Intensive Systems. pp. 123–130 (2010).
16. Aoyama, M.: Continuous and Discontinuous Software Evolution: Aspects of Software Evolution Across Multiple Product Lines. In: Proceedings of the 4th International Workshop on Principles of Software Evolution. pp. 87–90. ACM (2001).
17. Aoyama, M., Watanabe, K., Nishio, Y., Yasuyuki, M.: Embracing requirements variety for e-Governments based on multiple product-lines frameworks. In: Requirements Engineering Conference, 2003. Proceedings. 11th IEEE International. IEEE (2003).
18. Bühne, S., Lauenroth, K., Pohl, K.: Why is it not Sufficient to Model Requirements Variability with Feature Models? Work. Automot. Requir. Eng. AURE04. 4, 5–12 (2004).
19. Krueger, C.W.: New methods in software product line development. In: Software Product Line Conference, 2006 10th International. pp. 95–99. IEEE (2006).
20. Marinho, F.G., Andrade, R.M.C., Werner, C., Viana, W., Maia, M.E.F., Rocha, L.S., Teixeira, E., Filho, J.B.F., Dantas, V.L.L., Lima, F., Aguiar, S.: MobiLine: A Nested Software Product Line for the domain of mobile and context-aware applications. Sci. Comput. Program. 78, 2381–2398 (2013).
21. Savolainen, J., Mannion, M., Kuusela, J.: Developing Platforms for Multiple Software Product Lines. In: Proceedings of the 16th International Software Product Line Conference - Volume 1. pp. 220–228. ACM, New York, NY, USA (2012).
22. Rosenmüller, M., Siegmund, N., Kästner, C., Ur Rahman, S.S.: Modeling Dependent Software Product Lines. Engineering. 13–18 (2008).
23. Altintas, N.I., Cetin, S.: Managing Large Scale Reuse Across Multiple Software Product Lines. In: High Confidence Software Reuse in Large Systems. pp. 166–177 (2008).
24. Schröter, R., Siegmund, N., Thüm, T.: Towards modular analysis of multi product lines. Proc. 17th Int. Softw. Prod. Line Conf. co-located Work. - SPLC ’13 Work. 96 (2013).
25. Nakagawa, E.Y., Oquendo, F.: Perspectives and Challenges of Reference Architectures in Multi Software Product Line. In: Proceedings of the 17th International Software Product Line Conference Co-located Workshops. ACM, New York, NY, USA (2013).
26. Tekinerdogan, B., Erdoğan, Ö.Ö., Aktuğ, O.: Chapter 10 – Archample—Architectural Analysis Approach for Multiple Product Line Engineering. In: Relating System Quality and Software Architecture. pp. 263–285 (2014).

Automated software generation process with SPL

Jesús-Moisés Hernández-López[1,1], Ulises Juaréz-Martínez[2,2] and Ixmatlahua-Díaz Sergio-David[1,3]

[1] Instituto Tecnológico Superior de Zongolica, Veracruz, México,
[2]Instituto Tecnológico de Orizaba, Veracruz, México
[1]jesus_hernandez_pd197@itszongolica.edu.mx, [2]ujuarez@ito-depi.edu.mx,
[3]sergio.ixmatlahua.pd169@itszongolica.edu.mx

Abstract. Software Product Line (SPL) is a set of applications with a common architecture and shared components, with each application specialized to reflect different requirements. SPLs manage a large number of artifacts, for each of them it is necessary to define *commonality* and *variability*. Consequently, the applications generation from SPL becomes complicated and it often must be done manually. This paper shows the development of a SPL with an automated software generation process, in which *commonality* and *variability* have been defined for each artifact (i.e., product management, requirements, design, realization and test). Some technologies used were: XML files, Scala, AspectJ, Apache Maven and Junit. To automate the applications generation process, we have developed a configurator that enables *features selection* for each application and generate it in an automated way from SPL. The software generated includes: executable and documentation. Further, we propose a model-driven architecture for support the evolution in the SPL.

Keywords: variability, MDA, inmotic, traits, aspects, mixins

1 Introduction

Nowadays, there is a demand for enhancing the quality of software, reducing costs and accelerating their development processes [1]. SPLs are focused on these aspects (i.e., quality, cost and time).

A SPL is a set of applications with a common architecture and shared components, with each application specialized to reflect different requirements [1]. SPL development requires generating a large number of artifacts, such as: product management, requirements, design and test [2]. For each artifact generated, *variability* and *commonality* must be defined and managed. Variability defines the flexibility of a SPL to generate products with different features and commonality defines artifacts

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_12

that will be reused in each application from SPL. Due to the number of artifacts, the generation process becomes in a complicated task for developers, which many times must be done manually. The drawback of doing it manually is that it increases the time of software generation and there is a risk of making mistakes in the configuration.

In this paper, we show the development of a SPL based on *features selection* with an automated software generation process. The development of this SPL was carried out through the classic methodology for Software Product Line Engineering, in which 2 processes are distinguished: Domain engineering process.– *This process is responsible for establishing the reusable platform and thus for defining the commonality and the variability of the product line*. Application engineering process.- *This process is responsible for deriving productline applications from the platform established in domain engineering* [2]. *Commonality* and *variability* were defined with different technologies, such as: XML files, Scala, AspectJ, Apache Maven and Junit. To automate software generation process, we developed a configurator, which is in charge of reusing common artifacts and exploiting variability defined in the SPL, in order to generate complete applications (i.e., documentation and/or executable application). The process is executed in an automated manner, based only on *feature selection.* Further, we propose a model-driven architecture for support the evolution in the SPL. To verify our approach, we choose inmotic domain as case study.

Document is organized as follows: Section 2 presents related works. Section 3 explains the automated software generation process with a simple example. Section 4 presents a case study. Section 5 shows a proposal to support evolution in SPLs. Section 6 summarize conclusions and future work.

2 Related works

In [3], variability was introduced through integration of Model-Driven Software Development (MDSD) and Aspect-Oriented Software Development (AOSD). Combination of MDSD and SPL facilitates traceability from problem domain to solution domain. Aspect-oriented language is useful to generate code where an architecture doesn´t provide links.

In [4], an Aspect Oriented Analysis (AOA) on product requirements was presented to design Product Line Architectures (PLA). AOA scheme consists of: (1) requirements are separated in each aspect of original requirements, (2) requirements of each aspect are analyzed and the architecture is examined for each of them, (3) results and design options are analyzed.

In [5], an MDE (Model-Driven Engineering) and Aspect-Oriented Software Development approach was presented to facilitate implementation, management and traceability of variability. Features are separated into models, which are composed by AO techniques at model level. By integrating MDSD into SPLE (Software Product

Line Engineering), DSL (Domain Specific Language) manages the variability with respect to its structure or behavior.

In [6], a new approach was presented to implement SPL by fine-grained reuse mechanisms. Featherweight Record-Trait (FRTJ) was introduced and product functionality units are modeled with traits and records. Reuse degree of traits and records is higher than the potential to reuse hierarchies based on standard static classes.

In [7], an approach to facilitate variability management was proposed to model architectures of SPLs. Domain requirements and architecture are captured into models. For application engineering, DSL was used to specify requirements of particular applications. AO techniques were used during the domain engineering to modulate concerns in models, transformers, and generators.

In [8], delta-oriented programming (DOP) was proposed. DOP is a programming language designed to implement SPLs. A delta module can add classes for products implementation or remove classes from them. A SPL implemented with DOP is divided into a core module and delta modules. The core module comprises a set of classes to implement a complete product with valid feature configuration. It enables a flexible modular implementation for product variability, starting from different core products.

In [9], an automated assembly for domain components of a PLA was presented at low level abstraction. To address this problem, Model-Driven Engineering (MDE) was used. Benefits obtained with MDE are: (1) improving development of PLA with integration of modeling tools and specific domain components, (2) model-based structures help keep stability of domain evolution in MDE-based systems, (3) improving robustness and ability of models transformation, further debug support to correct errors in transformation specifications.

Related works show different approaches to manage variability and commonality at different levels of abstraction. There are some approaches that enable managing variability at code level, such as: *aspects*, *traits* and DOP [1] [3] [4] [5] [6]. We have also researched approaches to manage variability at high level abstraction: for example: MDA and AOA approaches [2] [7]. In conclusion, approaches based on *aspects* enable code manipulation at compile time and *traits* have a higher degree of reuse than classic class inheritance. MDA approaches support the evolution on PLA.

3 Automated software generation process

The automated application generation process is described step by step in Figure 1 through a simple example. Step (1), the administrator of the SPL makes *features selection* from *features model*, which is shown with a graphic interface in the configurator. The *features model* was defined through Common Variability Language (CVL) [10]. This example contains a *features model* with 3 features (A, B and C).

The Variability Specification Tree Resolution (VSpectree Resolution) is generated in step (2), applying CVL rules [10] defined in the configurator. The example of Figure 1 shows a VSpectree Resolution with two features (A and C), which is a valid selection for the features model.

VSpectree Resolution is verified in step (3) with the *configuration file* based on XML. This file contains features defined in the SPL, features definition for each application, and location of each artifact. The example of Figure 1 shows a *VSpectree Resolution* is equal to “Product 1”, defined in the *configuration file*.

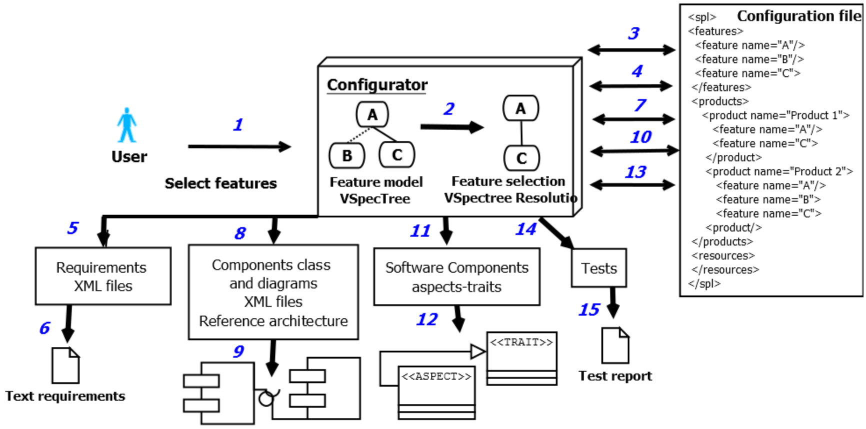


Figure 1. Automated application generation processes

Requirements artifacts (i.e., requirements on XML files) are reused in steps (4), (5) and (6). Common requirements for all applications are reused directly, and specific requirements are added depending on the application that will be generated. *Requirements artifacts* are result of *domain engineering process [2]*, which it defines communality and variability. Figure 1 shows the configurator working like a traceability link mechanism, linking features between “Product 1” from *configuration file* until *requirements artifacts.* The configurator generates application requirements in the step (6), reusing requirements with the value “Base” for the attribute condition and adding requirements where the value is equal to features name in “Product 1”. . This process is observed in Figure 2.

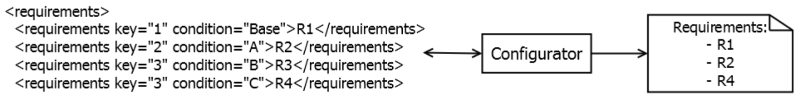


Figure 2. Requirements generation for “Product 1”

Like steps (4) and (5), steps (7) and (8) work with design artifacts developed on XML files, such as: components and class diagrams. Communality and variability were defined using a tool called PlantUML [11]. The artifact generated by *domain* *engineering* is the *reference architecture.* This artifact is reused for all applications of the SPL in order to generate complete architectures for applications that will be generated. *To* generate a complete architecture in step (9), *application engineering process is* carried out *as* observed in Figure 3. The process is the following: Configurator chooses components, interfaces and relations with value “Base” in order to generate *reference architecture*. Subsequently, the configurator chooses and assembles components, interfaces and relations, verifying that the value in *condition* is equal to features name from “Product 1” from Figure 1.

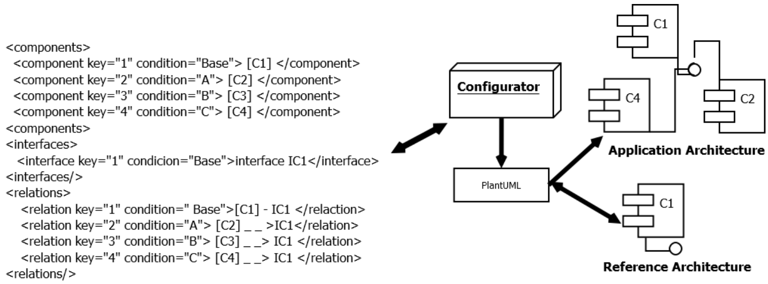


Figure 3 Example of architecture generation

Steps (10), (11) and (12) of Figure 1 are carried out in order to generate executable applications. In step (10), the configurator is responsible for verifying whether each *trait* is available, comparing *traits* and *features* defined for “Product 1” *in file configuration.* If there is no problem then the process continues. The application generation works in a systematized way from step 12. This process is shown more broadly in Figure 4. Before starting compilation, the configurator identifies the product that will be generated at time compile to weave *aspects* and *traits* through Maven [12], using command “*mvn install*”. “*AddFeatures*” aspect is responsible for assembling features “A” and “C” in *the reference architecture* during compilation. Components assembly is performed by *application engineering process.* Finally, whether there are no faults in the testing phase, then executable application (.jar) is generated.

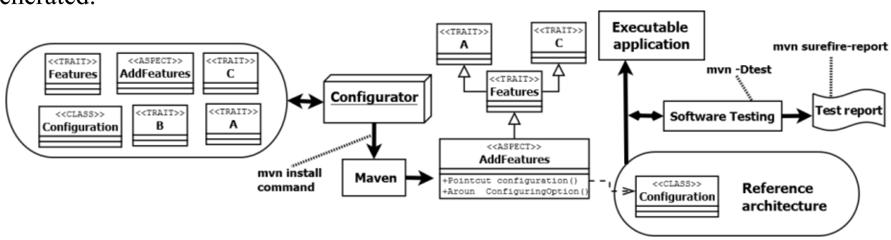


Figure 4 Applications generation process

Steps (13), (14) and (15) are carried out in order to apply software tests for the application configured in step (12). We have developed test suites with JUnit framework to distinguish domain and application tests. Before executable application

(.jar) is generated, tests are executed by the configurator with "*mvn –Dtest"* command, as observed in Figure 4. Whether there are no faults in the testing phase, then executable application (.jar) is generated with its test report, using "mvn surefire-report" command. Whether there are faults, the test report is also generated but the application generation process (.jar) is aborted until it is corrected. The following is an example of *Test Suite* for "Product 1" from Figure 1, in which will only execute tests for Product 1, specifically test for A and C features.

```
@RunWith (classOf [Categories])
@IncludeCategory (classOf[Product1])
@SuiteClasses(Array(classOf[TestA],classOf[TestC))
class SuiteProducto7Test {}
```

4 Case study

The case study chosen for this SPL was the inmotic domain. Inmotic is the incorporation of numerous subsystems in installations of tertiary or industrial use, in order to optimize resources, reduce costs and unnecessary energy consumption. The reason for dealing with this domain was to get a domain in which the variability is necessary. Examples of variability about inmotic domain, such as: sensors (e.g., flame, light or temperature), actuators (e.g., relays, valves or motors) and user interfaces (e.g., TV interface, a web-based interface or mobile interface).

Figure 5 shows features model defined with CVL [10] for the SPL. Lights feature and device feature define a large part of *commonality* in the SPL. They are defined with continuous line (i.e., mandatory features), therefore, applications generated will always have LIGHTS feature and DEVICE feature (i.e., a microcontroller). Variability is defined at different levels VSpectree. The most notorious variability is defined in device feature . There are applications that work with ARDUINO and others with RASPBERRY, but they don't work in the same application.

.

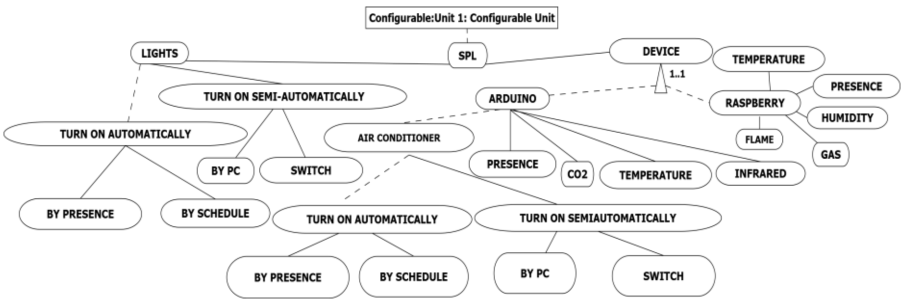


Figure 5 Features model defined with CVL (VSpectree)

The SPL is able to generate 8 applications for inmotic domain, which include executable applications, requirements, design and test report. Different artifacts are generated for each application, however, some components are shared between them. Two valid aplications are shown in Table 1, in which there are notable differences. For example: application 1 doesn´t has any automation features, such as: Turn on Lights automatically by presence sensor or turn on lights automatically by schedule. Aplication 2 have features for light automation and other sensors, such as: flame, humidity or gas.

Table 1. Valid applications obtained from VSpetree of Figure 5

Application 1	Application 2
Lights	Lights
Turn on lights semi-automatically	Turn on lights semi-automatically
Lights control by pc	Lights control by pc
Lights control by switch	Lights control by switch
Arduino	Turn on lights automatically
Air conditioner	Turn on lights automatically by schedule
Turn on air conditioner semiautomatically	Turn on lights automatically by precense sensor
Air conditioner control by pc	Raspberry
Air conditioner control by switch	Presence sensor
Presence sensor	Humidity sensor
CO2 sensor	Gas sensor
Temperatura sensor	Flame sensor
Infrared sensor	

The *reference architecture* of this SPL is shown in Figure 6. *Reference architecture* is an incomplete architecture that will be reused and completed to generate each application of this SPL, as is explained in Section 3. Architectural style of the *Reference architecture* is the classic client-server.Therefore, after every automated application generation process, SPL in conjunction with the configurator will deliver two executables (i.e., client.jar and server.jar).

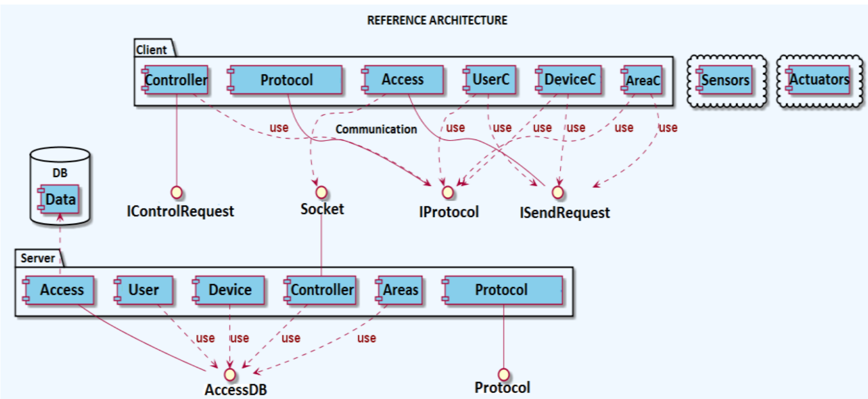


Figure 6. Reference architecture of the SPL

Design artifacts and realization artifacts for the applications described in Table 1 are shown to observer the reuse of the same components and differences between them caused by variability. Figure 7 shows a diagram class in which application 2 is configurated, as observed in Table 1. Trait called "*Features*" inherits functionality from traits called "*Light*", "*allData*", "*Presence*","*Temperature*", "*Flame*", "*Gas*"*and* "*Humidity*" creating mixin composition . Subsequently, the *aspect* called "*Access*" inherits the *trait* "*Features*" configured to add it in the *reference architecture* to generate the *application architecture of the application 2.*This process is generated at binary level by the configurator, which is explained in Section 3.

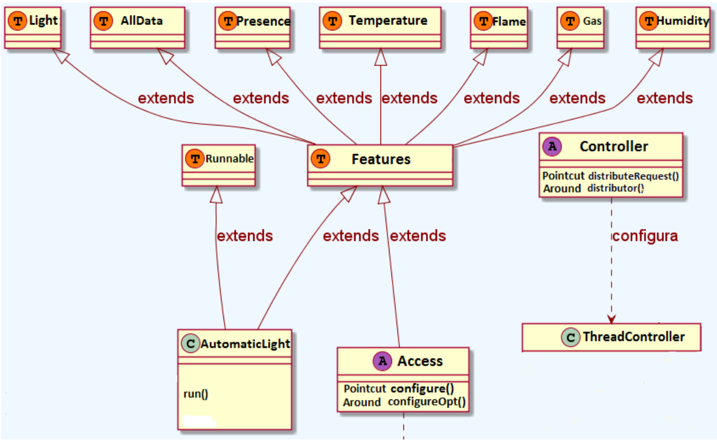


Figure 7. Combination of *aspects* and *traits* to configure Application 2

Coming up next is a fragment of the configuration for some features in Application 1. In this code is shown the reuses of "Presence", "Light" and "Temperature" traits. The "*AirConditioner*", *and* "*Infrared*" assembly is equally performed for the 8 applications from SPL. The "*AccessAspect*" aspect inherits "*Features*" trait and with a cutting weaves the full functionality where the *Pointcut* has indicated.

```
trait Features extends Light
with AirConditioner
with Presence
with Temperatura
With Infrared
with CO2
with Arduino{}

@Aspect
class AccessAspect extends Features {}
@Pointcut("call()) && args()
def configure() = {}
@Around()def configureOpt() = {}
```

5 MDA to Support Evolution in SPLs

Our proposal to address problems in terms of evolution was represented through two models of MDA approach: Platform-Independent Model (PIM) and Platform-Specific Model (PSM). PIM will represent all elements of the SPL at high level abstraction, without indicating technologies. PIM can be observed in Figure 9. Coming up next PIM elements are described:

Domain. - Corresponds to market segment for one SPL.
Features model. - There are different languages to model features. PIM does not indicate any specific language.
Features selection. - This concept will get a model features instance.
Platform. – It contains artifacts obtained from domain process and application process
Configurator. - It's responsible to generate applications using automatically artifacts (i.e., domain artifacts and application artifacts). Also, it allows verify all domain artefacts, before generating applications.
Domain or application artifacts. - They are artifacts generated by the configurator.

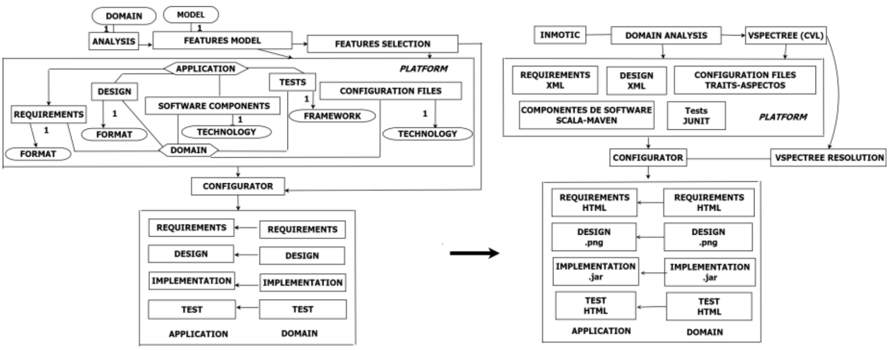


Figure 8 PIM to PSM

From PIM, is possible obtaining one or more PSMs, showing a projection of the PIM. PSM generation is based on transformation rules established in the PIM. The PSM generated in Figure 8 shows a model with elements of a SPL for inmotic domain. In this way is possible generating others PSMs for different domains.

6 Conclusions and future work

The *application engineering process* [2] was automated by the configurator, which is responsible to generate artifacts for 8 applications from SPL. Commonality and variability of the SPL was managed at different levels of abstraction with technologies quite flexible. *Aspects* and *traits* allow weaving binary code at specific points of the reference architecture to configure applications. MDA was used to propose a PIM to support evolution in SPLs, thinking not to depend on technology, but rather to adapt to unexpected changes. PSMs generated from PIM will be constantly changing at technological level but without modifying the PIM, this way, the evolution will be in both models. Although the goal of this paper neither comparative nor statistical, it is quite remarkable that the automated software generation process from SPL reduces time to generate software instead of manually configuring it. An automated software generation from this SPL takes about 10 minuts, depending the number of features chosen for one application. The software quality increases also due to the constant work with the same software components, further, test reports generated indicate whether the software has bugs and where they were located. Future work will be intended to perform a comparative and statistical analysis of advantages and disadvantages between manual processes and automated process to generate applications, and to evaluate attributes focused on time and quality. It will also increase the number of applications to the product portfolio [2] of the SPL.

Reference

1. Sommerville, I.: Software Engineering 9th ed, (2010)
2. Van Der Linden, F., Pohl, K.: Software Product Line Engineering: Foundations, Principles, and Techniques. In: Springer Science Business Media, (2005)
3. Voelter, M., Groher, I.: Product line implementation using aspect-oriented and model-driven software development. In: Software Product Line Conference, 2007. SPLC 2007. 11th International, pp. 233-242 IEEE, (2007)
4. Kishi, T., Noda, N.: Aspect-oriented analysis for product line architecture. In: Soft-ware Product Lines pp. 135-145, Springer, (2000)
5. Groher, I., Voelter, M.: Aspect-oriented model-driven software product line engi-neering. In: Transactions on aspect-oriented software development VI pp. 111-152, Springer, (2009)
6. Bettini, L., Damiani, F., Schaefer, I.: Implementing software product lines using traits. In: Proceedings of the 2010 ACM Symposium on Applied Computing pp. 2096-2102, ACM, (2010)
7. Groher, I., Voelter, M., Schwanninger, C.: Integrating Models and Aspects into Product Line Engineering. In: SPLC pp.355, (2008)
8. Schaefer, I., Bettini, L., Bono, V., Damiani, F., & Tanzarella, N. (2010, September). Delta-oriented programming of software product lines. In *International Conference on Software Product Lines* (pp. 77-91). Springer Berlin Heidelberg.
9. Deng, G., Schmidt, D. C., Gokhale, A., Gray, J., Lin, Y., Lenz,G.: Supporting Evolution in Model-Driven Software Product-line Architectures. In: ACM SIGSOFT Software Engineering Notes, ACM, (2007)
10. CVL Submission Team: Common variability language (CVL), OMG revised sub-mission ,(2012)
11. PlantUML, http://plantuml.com/
12. Maven, I.: Apache maven project, (2011)

Systematic Review: Cybersecurity Risk Taxonomy

A.M. Rea-Guaman[1], T. San Feliu[1], J.A. Calvo-Manzano[1] and I.D. Sanchez-Garcia[2],

[1] Universidad Politécnica de Madrid, ETS Ingenieros Informáticos
Madrid, España
marcelo.rea.guaman@alumnos.upm.es
{tomas.sanfeliu, joseantonio.calvomanzano}@upm.es

[2] Instituto Politécnico Nacional, Escuela Superior de Ingeniería Mecánica y Eléctrica
Ciudad de México, México
issanchez@ipn.mx

Abstract. In cybersecurity, the identification of risks is a fundamental part because this activity is not unique to cybersecurity and it is hard to know what the risks in this area are. This study aims to identify if there are some risk taxonomies in cybersecurity. For this, a systematic review of the studies published from 1990 to 2017 was carried out. We found 132 papers and some of them mention some risk taxonomies within the scope of IT (information technologies) cybersecurity, although only five primary elements were selected, identifying the main risk taxonomies. A classification of cybersecurity risk taxonomy types has been adapted, with the inclusion of new categories, categorized according to their perspective and domain. We have analysed the taxonomies form a proposed five level perspective. Finally, it has been observed that risk taxonomies may be shifting the focus from the asset level to service and business level.

Keywords: Cybersecurity risk taxonomy; Cybersecurity risk; Risk taxonomy, Cyber risk taxonomy.

1 Introduction

In recent years, interest in cybersecurity has been increasing. As our world becomes increasingly interconnected, real-time availability of systems becomes more and more necessary. Therefore, the impact that cyber threats can cause to organizations can be large, and the care and protection of information assets within organizations is of great importance. And the assets are important and crucial in critical business processes. Also, the information that is shared in the different technologies has an increasing value for consumers and users. The information contained in the systems is more valuable than the technologies that contain the systems, more and more private and governmental organizations have combined efforts to standardize the identification of risks that affect cybersecurity.

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_13

How organizations address the cybersecurity risks in their organizations is essential in order to implement effective, efficient and sustainable cybersecurity. Therefore, it is necessary to identify the risks that affect organizations in the area of cybersecurity. There are risks in almost all areas of the organizations. Risks that fall into cybersecurity and are difficult to identify, cybersecurity risk taxonomies assist in determining the risks that exist within the scope of cybersecurity.

Based on the above, a systematic review was carried out to identify the main taxonomies and classification of cybersecurity risks. Thus, this paper is organized as follows: Section 2 presents the concept of risks and it includes the cybersecurity risk, Section 3 presents the details of the systematic review process; Section 4 analyses and interprets the results of the systematic review; Section 5 reports on the results of the systematic review based on the different taxonomies of existing or applicable cybersecurity risks, includes the adaptation of a new classification of taxonomies with the inclusion of new categories, according to their perspective and, finally, Section 6 presents the conclusions.

2 Context

A definition of risk according to ISACA is: "The combination of the probability of an event and its consequence" [1]. This is a definition that applies to any field, whether it is an environmental risk, a work risk or a risk in the field of information technology. Different classifications have been made to identify the cybersecurity risks, for the different areas because have particular characteristics.

To better understand what a cybersecurity risk is, the following concepts are presented according to ISACA. To understand the term cybersecurity, we must first define the term cybersecurity risk. "Cybersecurity risk is not one specific risk. It is a group of risks, which differ in technology, attack vectors, means, etc. We address these risks as a group largely due to two similar characteristics: A) they all have a potentially great impact B) they were all once considered improbable" [2]. "Cybersecurity is the sum of efforts invested in addressing cybersecurity risk, much of which was, until recently, considered so improbable that it hardly required our attention" [2].

By the previous definitions, we know that the handling of a cybersecurity risk is different from the other types of risks and, in turn, the risks in cybersecurity are being very variable. It is therefore important to have a taxonomy that helps to identify and classify the risks inherent in cybersecurity.

In this paper, it is intended to answer through a systematic literature review the following question: Are there any risk taxonomy related to cybersecurity published?

In order to answer the question, the proposed systematic review technique by Kitchenham is used [3], [4]. A systematic review is a formal and verifiable process that the researcher performs to document the state of knowledge in a specific subject. The systematic review [4] makes possible to: (1) review the relevant work that has been done in the area of study, (2) examine the results, evaluate and contrast them, and (3) identify gaps in current research in order to do an appropriate proposal for a new research activity.

3 Systematic review

The systematic review includes the following activities: (A) identifying the needs, (B) proposing a review protocol, (C) conducting the review (identifying primary studies, evaluating studies, and synthesizing information), (D) analysing and interpreting the results, and (E) reporting the results of the systematic review.

Next, the process of the systematic review related to cybersecurity risk taxonomies published is detailed. Section 4 and 5 of the paper present the activities D and E of the systematic review.

3.1 Identification of the Needs for the Systematic Review

A systematic review was required to identify the different taxonomies of cybersecurity risks that have been published.

3.2 Review Protocol

At this point the following tasks were defined: Formulation of the questions to be asked, the criteria for selecting the database sources, the database sources to be used for the search, the elaboration of the search strings according to the defined criteria and the search in the sources, to locate and select the studies.

Formulation of the question. Question(s): What risk taxonomies have been proposed for cybersecurity?

The issues and questions related to the needs and objectives of the review were raised.

- Problem: there is a need to implement cybersecurity risk taxonomies in organizations, but it is not known which have been proposed. This also makes possible to determine a trend of implementation.
- Questions*:* What risk taxonomies have been proposed for cybersecurity?
- Population*:* publications related to risk taxonomies in cybersecurity, security management in information systems, and applications in organizations.
- Intervention*:* different taxonomies of cybersecurity risks that have been published/used.
- Effect: to know if there are any taxonomy that cover all areas of cybersecurity or not.
- Result measurement*:* proposed cybersecurity risk taxonomies, their descriptions and approaches.
- Application*:* to know the different taxonomies of cybersecurity risk, and their approaches.

Criteria for the selection of sources. The criteria for the selection of sources were: database that include journals and papers focused on cybersecurity risk taxonomy and have advanced search mechanisms, making use of the terms and synonyms used in search queries; availability of full text papers; papers available on the web for free; specialized magazines are available in the library of the Universidad Politécnica de Madrid.

Identification of the sources. Specialized databases such as IEEE Computer, Science Direct, ACM Digital Library and Web of Knowledge are among the selected sources.
Search string. The terms used in the systematic review were constructed using the following criteria: (1) "Cybersecurity Risk Taxonomy", (2) "Cybersecurity Risk", (3) "Risk Taxonomy" and (4) "cyber risk taxonomy". These keywords, combined with the logical operators AND and OR, as well as the NOT operator to refine the search, were used in the search engines of the specialized databases.

The words used in the search string include: "Taxonomy", "cybersecurity", "risk", "cyber risk".

Search in the sources. We searched the sources using the criteria defined previously for their selection. All sources of the identified databases were included. Search strings were applied to electronic databases and other sources (journals and conferences). To evaluate the list of sources obtained, were involved two experts in cybersecurity risk taxonomies.

3.3 Review

At this point, the search of the papers in the databases selected with the predefined search strings is displayed. The review is done in three phases. The inclusion and exclusion criteria established below were applied to the search results.

Criteria for selecting studies and procedures for inclusion and exclusion within the primary studies. The Table 1 lists the inclusion and exclusion criteria that were applied to the results of the initial search. The selection of studies was focused on those related to cybersecurity risk taxonomies published.

Table 1. Inclusion and Exclusion Criteria.

Inclusión (I)	Exclusión (E)
I1. Empirical studies of cybersecurity risk taxonomies that have related content in showing some risk classification that applies to cybersecurity.	E1. Papers that are based only on a particular opinion that does not address cybersecurity.
I2. Papers that discuss taxonomies or classifications of cybersecurity risks.	E3. Studies that are not relevant to the research question or are not related to the particular study.
I3. Papers that use keywords.	E4. Studies that are unclear or ambiguous.
I4. Papers whose title, summary or content are related to the topic.	E5. Duplicate publications.
I5 Free access documents.	E6 Pre-1990 studies due to constant updating in the area.

In a first phase of the review, making use of the search engines of the identified databases and putting the search string elaborated in the task “the search string” of the activity "review protocol", a total of 132 studies were found.

In the second phase of the review, each study was reviewed taking into account the previous inclusion and exclusion criteria, obtaining a total of 14 relevant studies. To select the relevant studies, the following steps were taken: read the title, read the summary, read the conclusions and fill in template created, if they provide enough information, the study is selected and saved. Otherwise, it is deleted.

In the third phase of the review, five primary studies were obtained that answered the questions formulated initially. Table 2 shows the sources, the total number of papers, and the number of primary paper selected by source.

Table 2. Distribution of studies by source.

Source	Total	Relevant	Primary
IEEE Explore	11	3	2
Science Direct	6	0	0
ACM Digital Library	8	0	0
Web of Knowledge	107	11	3
Total	**132**	**14**	**5**

Evaluation of study quality. To assess the quality of the studies, the following questions were asked:

- Is the primary study relevant to the research being done?
- Do primary studies provide enough information for the results to benefit from the systematic review?

From the previous questions, it was verified that the five primary studies selected are relevant, provide sufficient information and add value to the systematic review.

Data extraction and synthesis. To extract the important information of each paper, a template was designed. The template contains the following fields: (1) paper identification, (2) reference (title, author, year), (3) type of paper (case study, survey, experiment, research), (4) purpose of the study, (5) context of the study, (6) type of study (improvement, deployment or both), (7) depth of analysis (high, medium, low), (8) cybersecurity risk taxonomies published, (9) area and approach to taxonomy, and (10) observations

For each paper selected, after reading the full text, the information was recorded on the form, which allowed for the subsequent analysis of the results.

4 Analysis and interpretation of the results of the systematic review

At the conclusion of the systematic review, the context of each document was analyzed, for those documents that proposed a taxonomy in cybersecurity, the proposed risk classification was analyzed, and its focus and scope.

According to their context they were classified in:

- Framework related risk.
- Concepts on how to perform a taxonomy.
- Taxonomy focused on a specific area.

There is one framework enterprise risk management (ERM), one concept on how to perform a taxonomy of information security risks and three taxonomies focused in: social engineering, software providers and information security risk.

The results of the systematic review are presented below, the primary items found are described in the Table 3 for further analysis.

Table 3. Primary studies

Paper	Context	Author
Three key enablers to successful enterprise risk management.	Risk management	Kanel (2010)
Taxonomy of information security risk assessment (ISRA)	Information security risk	Alireza (2016)
Understanding Taxonomy of Cyber Risks for Cybersecurity Insurance of Financial Industry in Cloud Computing	Perform a taxonomy of information security risks	Elnagdy (2016)
Analysis of Unintentional Insider Threats Deriving from Social Engineering Exploits Developing a Risk Management	Taxonomy focused on social engineering	Greitzer (2014)
Process and Risk Taxonomy for medium-sized it solution providers	Taxonomy focused on software providers	Herzfeldt (2012)

In order to obtain complementary studies, the technique of snowballing has been applied. This has allowed to obtain two additional studies from [8] and one additional study starting from [5]. Additionally, a search in Google Scholar provided two complementary taxonomies.

These complementary studies are shown in the Table 4.

Table 4. Primary studies

Paper	Context	Reference
A Taxonomy of Operational Cyber Security Risks.	Enterprise	Cebula (2010)
The structure of a cyber risk a scenario based approach in cyber risk assessment	Enterprise	Delmee (2016)
Securing smart grid: cyber attacks, countermeasures, and challenges	Smart grid	Li (2012)
Classification and trend analysis of threats origins to the security of power systems	Energy systems	Bompard (2013)
Development of methodical social engineering taxonomy	Social engineering	Larebee (2006)

5 Results report of systematic review

Once each of the primary studies was selected and analysed, the published cybersecurity risk taxonomies were identified, and the area and focus were determined, as well as the frameworks and risk management models.

- The document of Kanel, J [9] addresses a risk taxonomy that includes cybersecurity risks and financial risks in organizations based on the COSO (Committee of Sponsoring Organizations of the Treadway Commission) framework.
- In the document of Elnagdy, S [8], it was found a taxonomy of basic cyber-attacks in smart grid, and a classification of malicious threats.
- The document of Greitzer F [5] presents a taxonomy of risks focused on social engineering.
- We identified that the author Herzfeldt A [6] presented a taxonomy of shrouded risks and IT system vendors is provided, which categorize the risks from the need identification to the implementation and maintenance of the system.
- The paper of Cebula, J [10] presented a Taxonomy of Operational Cyber Security Risks that mentions an adequate taxonomy of risks in cybersecurity. This is divided into four classes: actions of people, systems and technology failures, failed internal processes, external events.
- The paper of Delmee, F [11] created a cyber risk taxonomy based on a scenario based risk approach. The scenarios are further elaborated with contextual information about the victim and threat agent to identify all the relevant concepts within the scenario.
- We identified that the author Li, X [12] presents a taxonomy of basic cyber-attacks in smart grid communication. In this taxonomy, there are four types of attacks: device attack, data attack, privacy attack, and network availability attack. They have different objectives and are often the building blocks of more sophisticated attacks.
- The paper of Bompard, E [13] due to the growing recognition of the importance of power systems in today's society, a classification of malicious threats is presented, there are three malicious threats: physical threat, human threat and cyber threat.
- We identified that the author Laribee, L [14] in his study presented a taxonomy for encoding social engineering attacks, proposes four main dimensions of interest in determining the type and severity of a social engineering attack. The first category is the target, the second category is the type of deception, the third category is the particular resource or target information, and the fourth category is the thrust ploy.

To analyze the results, we have extended a classification of taxonomy types obtained from Alireza, S [7]. The Alireza's classification has three levels of bottom-up abstractions that are asset, services, business. Although most taxonomies rely on asset level, new studies are shifting the focus and moving to the level of service and business. We have included source attacks and external events as two complementary categories. We have added these two new categories, due to the difficulty of matching the classification elements of the taxonomies found and which were difficult to fit into the categories proposed by Alireza, S [7]. Each element of the analyzed taxonomies has been mapped to one of the five categories, the results of this mapping are shown in Table 5.

Table 5. Summary of taxonomies

Author	Domain	Perspective	Taxonomy Items
Kanel et al. (2010)	Enterprise	Asset + Service + Business + External	Sources and events Business objects and dynamics models Risk impacts
Li et al. (2012)	Smart grid	Asset + Attack + Business	Device Attack Data Attack Privacy Attack Network Availability attack
Bompard et al. (2013)	Energy Systems	Attack + Business + External	Physical threat Human threat Cyber threat
Greitzer et al. (2014)	Social engineering	Asset +Attack	Interaction personal Non Interaction personal
Laribee (2006)	Social engineering attacks	Asset + Attacks	Target Type of deception Resource or target information Trust ploy
Herzfeldt et al. (2012)	IT Solutions	Services + Business + External	Environment Provider Risk IT solution Risk Customer Risk
Cebula et al. (2010)	Enterprise	Asset + Service + Business + External	Actions of people Systems and technology failures Failed internal processes External Events
Delmee. (2016)	Enterprise	Asset + Service + Attack+ Business	General information Organization (victim) Threat agent Asset Threat event Business impact Vulnerabilities Control

The analysis indicates that all taxonomies take into account the Asset and Business level, with the exception of taxonomies specialized in social engineering. This is because they are focused on aspects of human interaction. It is noteworthy that specialized smart grid taxonomy, focuses solely on aspects of possible forms of attacks.

It is common in all other taxonomies besides the business aspect, take into account aspects of asset, service and external events. This is due to the need to identify threats to both resources and business activities. External events must be taken into account as they have to be monitored.

The most modern taxonomy, such as Delmee, F. [11], combines the three levels of abstraction with the category of sources of attack. This could indicate that risk taxonomies could be shifting the focus form the asset level to the abstraction levels of services and business.

6 Conclusions

This paper reviewed the cybersecurity taxonomy documents published in the conferences and journals process to understand progress in cybersecurity risk taxonomies. The objective was to identify the studies with respect to risk taxonomies that have been proposed for cybersecurity.

In order to obtain complementary studies, the technique of snowballing has been applied. This has allowed to obtain two additional studies from [8] and one additional study starting from [5],

The results are based on the results of five primary studies identified in the search engines selected for the review, two files identified in the snowballing of Elnagdy, S [8], one paper of Greitzer, F [5] and two primary files identified in the additional search in the Google Academic search engine, which may be a relatively low percentage of the total population of files obtained without applying the inclusion and exclusion criteria.

There are different types of risk taxonomies. Each of these taxonomies have been developed to meet a particular need. We have provided a general view and structure of risk taxonomies. Also, we have developed a new classification of taxonomy types. Based on this classification we have analyzed the taxonomies found and identified that most of them are of particular application and cannot be applied in other domains.

The taxonomies identified only abrogated specific risks such as social engineering risks, financial risks, operational risks, malicious threats and risks taxonomy based in certain specific scenarios.

In the paper of Elnagdy, S [8], it does not present a taxonomy, but it indicates the concepts necessary to be able to define a taxonomy of risks that can be applied in cybersecurity in the cloud.

The most modern taxonomy of cybersecurity risks is Delmee, F [11] study, based on a scenario based risk approach, and is distributed in eight different concepts and fifty-one characteristics.

Using the protocol of this systematic review as a starting point, additional searches from primary studies and new sources of studies will be reviewed in future iterations in order to complete the collection of taxonomies. We believe that this survey and study of risk taxonomies are important, in that it will help explain the different abstraction levels and could led to the development of more comprehensive and effective risk taxonomies.

References

1. ISACA Glosary (2016), https://www.isaca.org/Pages/Glossary.aspx?tid=1784&char=R.
2. ISACA, http://www.isaca.org/Knowledge-Center/Blog/Lists/Posts/Post.aspx?ID=296.
3. Kitchenham, B., Dybå, T., Jørgensen, M.: "Evidence-based software engineering," Proceedings of the International Conference on Software Engineering, pp. 273–281 (2004)
4. Kitchenham, B.: "Guidelines for performing systematic literature reviews in software engineering," EBSE Technical Report EBSE-2007-01, Keele University (2007)
5. Greitzer, F., Strozer, J., Cohen, S., Moore, A., Mundie, D., Cowley, J.: Analysis of Unintentional Insider Threats Deriving from Social Engineering Exploits. IEEE Security and Privacy Workshops, vol 35, 236-250 (2014)
6. Herzfeldt, A., Hausen, M., Briggs, R. O., Krcmar, H.: European Conference on Information Systems ECIS 2012. Developing a risk management process and risk taxonomy for medium-sized it solution providers. Association for Information Systems, Barcelona Spain (2012)
7. Alireza, S., Rouzbeh, B., Cheriet, M. (2016) Taxonomy of information security risk assessment (ISRA). Computers & security, vol 57, 14-30 (2016)
8. Elnagdy, S., Meikang, Q., Keke, G. (2016) Understanding Taxonomy of Cyber Risks for Cybersecurity Insurance of Financial Industry in Cloud Computing. IEEE International Conference on Cyber Security and Cloud Computing, vol 3, 295-300 (2016)
9. Kanel, J., Cope, E., Deleris, L., Nayak, N., Torok, R.: Three key enablers to successful enterprise risk management. IBM J. RES. & DEV, vol 54, 1-15 (2010)
10. Cebula, J., Young L.: A Taxonomy of Operational Cyber Security Risks. Software engineering institute. Recovered from: http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=91013 (2010)
11. Delmee, F.: Graduation research, The structure of a cyber risk a scenario based approach in cyber risk assessment. Utrecht University, Deloitte Nederland (2016)
12. Li, X., Liang, X., Lu, R., Lu, Shen, X., Lin, X., Zhu, H.: Securing smart grid: cyber attacks, countermeasures, and challenges. IEEE Communications Magazine, 50(8):38–45 (2012)
13. Bompard, E., Huang, T., Wu, Y., Cremenescu, M.: Classification and trend analysis of threats origins to the security of power systems. International Journal of Electrical Power & Energy Systems, 50:50– 64 (2013)
14. Laribee, L.: Development of methodical social engineering taxonomy. Master's Thesis, Monterey, CA: Naval Postgraduate School. Amazon Digital Services (2006)

Soft Skills for IT Project Success: A Systematic Literature Review

Carmen Iriarte, Sussy Bayona Orè,
Unidad de Posgrado de la Facultad de Ingenieria de Sistemas
Universidad Nacional Mayor de San Marcos,
carmeniriarteb@gmail.com; sbayonao@hotmail.com

Abstract. IT projects continue to fail at a high rate and this represents losses in both money and opportunity organizations. This leads us to continue studying success factors in projects. The aim of this paper is present a systematic literature review focused on soft skills for the IT project success. The literature search was conducted using the electronic databases and the primary articles from 2010 to 2017. The results show that a high number of studies are theoretical, conceptual or based on experience and there are no models that explain the causal relationship between soft skills and project success. All studies agree about the relevance of the project manager soft skills for projects success. The results show that the most influence soft skills for projects success are communication, leadership and conflict resolution skills.

Keywords: project success, soft skills, systematic review

1 Introduction

Despite the growth in IT investment in the last years, IT projects still fail at an alarming rate [1]. This high percentage of failures suggests the need to continue studying the issue. The literature also suggests that the performance of a project manager contributes to the success of a project [2] [3]. A large body of literature has concentrated mainly on studying the technical skills associated with project managers [4] [5] [6] [7]. The literature on soft skills is much smaller; Nevertheless, Leybourne [8] reports a shift from a technical bias (with its emphasis on project manager technical skills) to project manager behaviors. Studies on soft skills have focused largely on desirable qualities during the selection process of the project manager [9] [10] [11]. However, there is little literature that explicitly studies the link between soft skills and the success of an IT project. To solve this gap, in the present study we will perform a systematic review of literature focused on the link between project manager soft skills and the success of a project. This study contributes to the literature by identifying opportunities for future research in the field of the relationship between soft skills and project success. This study is also useful for managers since it can guide them in their decision-making processes of selection and training regarding the necessary skills in the project managers so that the project ends successfully.

This paper is structured as follows. Section 2 develops the previous literature on the subject. Section 3 presents the research question and the systematic review.

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_14

Section 4 shows the results of the review and discussion of the topic under study. Section 5 shows the limitations of the study. Section 6 presents the conclusions.

2 Background

We will first describe certain concepts that are part of the research topic. First, project management as an area of knowledge that covers this research topic, then as a dependent variable we will describe the concept of project success and then as part of the independent variable we will review the competencies of the project manager as well as the role of the project manager for successful projects.

2.1 Project Management (PM)

Project management is the discipline of initiating, planning, executing, controlling, and closing the work of a team to achieve specific goals and meet specific success criteria [12]. The first challenge for project management is to achieve the project goal, and objectives within the known constraints. The primary criteria or constraints are scope, time, quality, and budget. The secondary challenge, and the most ambitious of all, is to optimize the resource allocation of the inputs needed and integrate them to achieve the predefined objectives. There are many more constraints that depend on the nature of the project, safety, environmental, business opportunity and others of the strategic type.

2.2 Project Success

No clear definition of the concept is found. Talking about project success includes both successful project management and the successful product. There are many studies on the subject, studying components, criteria and factors to achieve a successful project. In addition, this concept is often based on the different perception of each stakeholder depending on the moment the project is founded. Baccarini [13] identifies two components that define a successful project: The success of the final product of the project and the success of the project management. More recently Liebowitz [14] identifies three components that are part of the project management process (planning, management, control), content (the information system as a product of the project) and the context (organizational culture, structure, managerial style, communication Internal, among others), proposing as a key that the context must be managed to ensure the product. It is usual to find that the same project is considered successful by some and a failure by others. Lim and Mohamed [15] explain that a project impacts in a different way to each element of society and each stakeholder (the individual owner, developer, contractor, user, the general public, and so on) has a different perspective. The perception of the stakeholders rather than being a global perspective is due to a perception of the achievement of the objectives of their interest [16].

2.3 Project Manager Competencies

Competencies are a general term widely used by all; however, it has different meanings for different people. A general definition is a set of abilities to do something or intervene in a given subject. In IT context, Bassellier and Blaize Horner Reich [3] study the term of IT competencies and use it as an umbrella that includes everything that impacts in the performance. It includes explicit and tacit knowledge, including the knowledge of knowing who knows what, since this allows one to incorporate the knowledge of others. As tacit knowledge it includes the combination of cognition and experience. In PM context, Crawford and Turner [18] defined ‘project manager competence’ as a combination of knowledge (qualification), skills (ability to do a task), and core personality characteristics (motives, traits, and self concepts) that lead to superior results.

The International Project Management Association's Competence Baseline [19] defines competence as “a collection of knowledge, personal attitudes, skills and relevant experience needed to be successful in a certain function”. A summary of practical and academic views has been developed through Project Manager Competency Development (PMCD) Framework [20] which presents the same three dimensions indicated by [18]: Knowledge, Performance and Personal. Project management knowledge competencies is what project manager knows about project management, project management performance competencies is all that the project manager is able to do or accomplish while applying project management knowledge, project management personal competencies is how a project manager behaves when performing the project or activity (their attitudes and core personality traits).

2.4 Soft skills

There are multiple definitions of soft skills and some synonyms or combinations such as soft competencies, human skills, social skills, and people skills. There is also literature in which it is defined in the field of attitudes as literature in which the concept is mixed with the personality traits. In this Systematic Literature Review (SLR) we will have a broad and comprehensive definition of soft skills as interpersonal qualities, also known as people skills, and personal attributes that one possesses [21]. While hard skills are the technical requirements and knowledge a person should have to carry out a task, soft skills are a psychological term that covers the personality types, social interaction abilities, communication, and personal habits, people believe that soft skills complement hard skills. Soft skills complement the technical skills requirements of a job. According to Goleman [32] the possession and use of soft skills contributes more to an individual’s ultimate success or failure than technical skills or intelligence.

2.5 The Role of Project Manager in Achieving Project Success

We find literature that studies the relevance of the role of project management and the role of the project manager in achieving project success. The project manager has a relevant role in the success of a project. Andersen et al [2] study the relationship between factors of success and actual success of the projects, finding factors that are directly influenced by the project manager and can make a real difference to the outcome of project endeavor. Another topic studied is the role of the project management in achieving the success of the Project. Munns and Bjeirmi [17] raise the importance of distinguishing between the objective of the project and the management of the project and demonstrate that a clear distinction between these two concepts brings a high possibility of project success.

3 Systematic Literature Review

The main objective was to identify previous studies, which investigated soft skills related to project management. In order to detect as much of the relevant literature as possible, we follow a systematic literature review (SLR). According to Kitchenham and Charters [22] the study process has followed 3 stages which are: planning the search, extraction and discussion of results.

In section 3.1 we have the first stage planning which begins with the definition of a protocol that includes the specification of a research question to be addressed in the study and the method that will be followed to carry out the review. The method includes a search strategy that ensures the identification of the largest number of relevant articles possible as well as an explicit definition of the inclusion, exclusion and quality criteria to evaluate each potential paper. In section 3.2 we followed the second phase extraction in which the data obtained are validated, recorded and classified according to their quality. In section 4 we showed a quantitative meta-analysis and discussion of results.

3.1 Research question

The research question posed to be answered in this systematic review is:

RQ1: What are the soft skills of the project manager raised in the literature as the most influential in the success of IT projects?

The present study includes searches from the publishers on technology topics such as IEEE Xplore (Digital Library), Science Direct (Elsevier), EBSCO, ACM (Association for Computing Machinery) and JSTOR. The search terms have been constructed in four steps: 1) identification of keywords from the research question identifying the most appropriate term for the dependent and independent variables, 2) Identification of synonyms and acronyms, 3) terms combination using the "and" and "or" operators and 4) adjustment of the search string according to the syntax of each data source. Search strings used are presented in Table 1.

Table 1. Search strings

Source	Web page	Search string
IEEE	ieeexplore.ieee.org/search	(“soft skills” OR “soft competencies”) AND “project success”
SDE	**www.elsevier.com/solutions/sciencedirect**	**(“soft skills” OR “soft competencies”) AND “project success”**
EBSCO	Web.a.ebscohost.com/ehost	(“soft skills” OR “soft competencies”) AND “project success”
ACM	**dl.acm.org.ezproxy.upc.edu.pe**	**+("competencies" OR "soft skills") +("project success")**
JSTOR	www.jstor.org.ezproxy.upc.edu.pe	((("soft competencies") OR ("soft skills")) AND ("project success"))

Criteria to select an article are presented in Table 2.

Table 2. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
• Include empirical studies of soft skills or soft competencies and their relation with IT project success. • Papers containing keywords that match those defined in the search string. • Papers whose title, summary or content is related to the topic. • Papers in English.	• Duplicated articles. • Papers that are based only on a particular opinion.

A quality assessment question list was constructed. Each selected article will be evaluated according to the criteria of Critical Appraisal Skills Programme [23]: (1) Rigourosity. It is verified that the study has applied a research methodology, (2) Credibility. It is verified that the findings and conclusions are correctly presented and with a complete meaning and (3) Relevance. It means that the study will be useful for the subject of our study. The question list covers the three criteria with seven questions in total. Quality score ranged from 0 to 7. The following scoring system was used to determine the individual question score: Yes (Y) = 1 point, Partial (P) = 0.5 points, No (N) = 0 points. The overall quality score was obtained summing the seven individual question scores. Thus, the total quality score for each paper ranged between 0 (very poor) and 7 (very good). The quality questions and scores obtained from included papers are presented in Table 3.

3.2 Search Execution

Our search resulted in 141 potentially relevant articles (Fig. 1). Of these, 11 publications met our criteria: 2 (ACM), 2 (EBSCO), 3 (Elsevier) y 4 (IEEE). To extract the relevant data of each paper and to standardize the way information has been represented, a data extraction form is designed. The data extraction form shall be filled out with the relevant information (Title, author, publication year, data source, abstract, method, sample size, study type, restrictions, country, period, soft skill name, soft skill definition, category name, category description, value, discussion). Keeping the data will allow more detailed analysis later. Table 3 shows the quality criteria assessment applied.

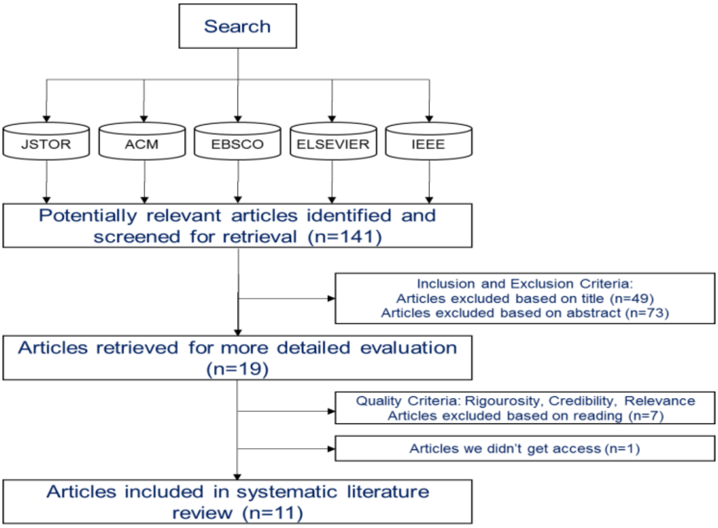


Fig. 1. Search Process. This flow chart shows the process followed during articles selection and quantities found.

Table 3. Quality Criteria

Quality Data	[24]	[25]	[11]	[10]	[26]	[27]	[9]	[28]	[29]	[30]	[14]
Rigourosity											
1 Is the paper based on an empirical research?	Y	P	P	Y	P	P	P	P	P	P	P
2 A clear description of the method was used to analyze data?	Y	Y	Y	Y	P	Y	Y	Y	P	P	P
Credibility											
3 Is there a complete presentation of the findings?	Y	Y	Y	Y	P	Y	Y	P	P	P	P
4 Present the study conclusions based on their findings?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	P
Relevance											
5 Is the paper focus in the study of soft skills?	P	Y	Y	P	Y	P	Y	P	Y	P	Y
6 Does the study primarily analyze project manager soft competencies?	Y	Y	Y	Y	Y	Y	P	P	P	P	P
7 Does the paper study cause and effect for soft skills and project success?	Y	P	P	P	Y	P	P	P	P	P	P
Total Score	**6.5**	**6.0**	**6.0**	**6.0**	**5.5**	**5.5**	**5.5**	**4.5**	**4.5**	**4.0**	**4.0**

4 Results

In the reviewed articles we found that 5 out of the 11 articles do not present scientific studies. They present only a conceptual study of the subject. This small number of published scientific articles reaffirms the above mentioned fact that we are facing a discipline that is still new. Another peculiarity about papers reviewed is that 6 out of the 11 articles that applied experimental design, only two exceeded the number of 100 results analyzed. The other studies that applied the experiment have done so with a very small sample. Fifty percent of the studies that applied the experiment have done it with samples smaller than 35. Most articles have as main topics the following subjects: (1) most valued soft skills during the stage of personnel selection (2) methodological surveys and interviews with managers in order to determine which are the most valued soft skills for them.

4.1 Types of studies

From 11 identified articles, six articles are quantitative studies (55%), one of them is theoretical (9%) and four articles are qualitative (36%) (see Fig.2). Empirical studies were focused on identifying the main factors, but not in explaining the cause and effect relation between soft skills and project success. Only one study presents a simple model between those variables.

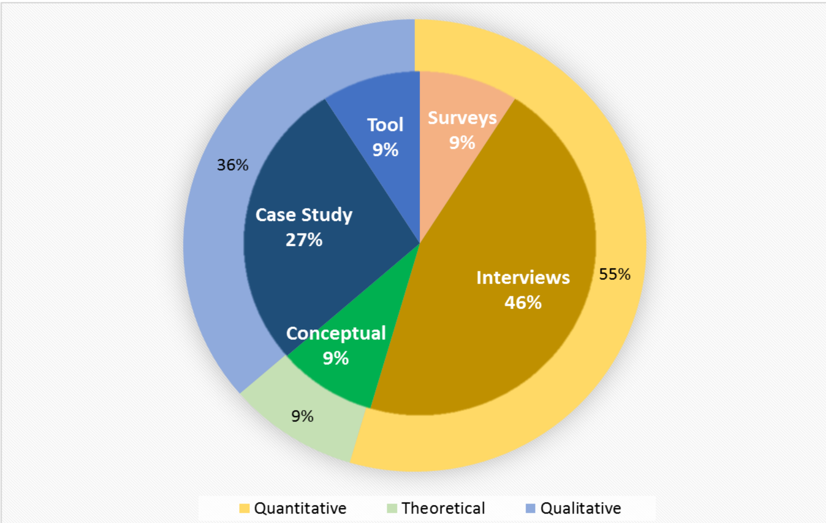


Fig. 2. Types of studies. This graphic shows the percentages of study and investigation types.

4.2 IT Project Manager Soft Skills

We discuss the answers to our research question.

RQ1: What are the soft skills of the project manager raised in the literature as the most influential in the success of projects?

The soft skills more mentioned were: communication, leadership, conflict management, thinking, Innovativeness, change orientation, negotiation, motivation and problem solving (see Table 4).

Table 4. Soft skills to achieve success in IT projects

Soft Skill	[24]	[25]	[11]	[10]	[26]	[27]	[9]	[28]	[29]	[30]	[14]	Fre
Communication		X	X		X	X			X			5
Leadership	X	X	X			X						4
Conflict Management		X			X	X						3
Thinking				X			X					2
Innovativeness		X			X							2
Change Orientation			X		X							2
Negotiation		X				X						2
Motivation	X					X						2
problem solving		X				X						2
Extroversion							X					1
Judgement							X					1
Self-Monitoring					X							1
Attention to detail						X						1
Relationship building						X						1
Self-awareness	X											1
Emotional Resilience	X											1
Intuitiveness	X											1
Interpersonal sensitivity	X											1
Influence	X											1
Conscientiousness	X											1
Collaboration											X	1
Creativity/resourcefulness				X								1
concern				X								1
Ability to get along		X										1
Respectful		X										1
Honest		X										1
Trusting		X										1
Operating within a multicultural environment										X		1

In Table 4 the column with label "Fre" is the frequency that a factor is mentioned by the authors. Most cited soft skill is the communication ability. Skulmoski and Hartman [25] present this soft skill as an ability required in all phases of a project. They say because you are working with others to complete a project, all team members need effective communication skills, and in particular verbal, generating feedback, and listening skills. They need to document what they plan to do and what they have done. One of the purposes of communication is that you need to manage the expectations of the sponsor and the team throughout the project. Sponsors sometimes have unrealistic expectations of what can be done, and technical team members sometimes want to implement perfect or cutting-edge solutions. Another particularity about this ability is that through this skill other skills are potentiated, this is a conclusion from [11] and [26].

In addition, Stevenson and Starkweather [11] emphasize the specific ability to be able to communicate not only with team members, but also with stakeholders and different levels of management. It represents a specific type of communication that should be assessed differently "Project managers need to know the most effective ways to communicate both up and down the organizational chain to a variety of audiences, and how to manage and influence people who don't report directly to them." and interestingly, "Ability to Communicate at Multiple Levels" had the highest factor loading for the Primary Competency Factor.

Creasy and Anantatmula [26] study and explore communication apprehension and take McCroskey [31] definition "an individual's level of fear or anxiety associated with either real or anticipated communication with another person or persons". They posit that a project manager's extent of communication apprehension can affect project outcomes. The ability to communicate was disaggregated by Keil et al. [27] as verbal, writing, listening, and persuasive arguments. Keil et al. ranked the 19 most valued abilities finding verbal ability in position 2, listening position 4 and writing position 6.

The second cited soft skill is the leadership. Müller et al [24] emphasize the growing importance of leadership in increasingly complex roles and tasks and in wider organizations. They also refer to the need for different leadership styles for different types of projects. They emphasize the growing importance of leadership in increasingly complex roles and tasks and in wider organizations. He also refers to the need for different leadership styles for different types of projects. Skulmoski and Hartman [25] find this competency important in all project phases. The leadership competency includes the ability to articulate the business problem, being vision-oriented, political awareness, agility and tact, and decisiveness. Stevenson and Starkweather [11] identified leadership as one of the six critical core competencies. They found leadership as the second most extremely important skill. For Keil et al. [27] leadership is the most important ability for project success. They define leadership as part of the team management skill category and say that competent IT Project managers need not only provide leadership, but also to motivate and empower their team members to successfully execute the project.

The third cited factor is conflict management. Skulmoski and Hartman [25] include it conflict management as a factor that forms part of the Negotiation category. They found this competence is important in all phases, but mainly during the implementation and closure phases due to frequency of problems and disputes in IT projects. Creasy and Anantatmula [26] propose that project managers are more likely to succeed than those who see the conflict as "negative and must be avoided".

4.3 Discussion

There is no unicity among authors in the grouping of soft skills categories. Some of the authors have done their studies by groups of factors, while others have done so by more detailed factors. In order to be able to determine the most valued skills in the literature, only specific factors have been included in the review. The most cited groups of factors in the literature have been the following: people skills [10] [27], personality traits [9] [26] and managerial skills [10]. Skills are defined as the ability to carry out a task. Soft skills include interpersonal skills, social skills and communication skills. Some definitions also include personality traits and attitudes. It seems important to review this definition in more detail and to study whether there is always a correlation between personality traits and soft skills or whether soft skills can be affected by external environmental factors that can change the attitude positively or negatively. Notwithstanding the general concepts and definitions in some cases it is difficult to discriminate a technical skill from a soft skill. Although they are different also there is proximity between them.

The empirical studies were focused on identifying the main factors, but not in explaining the cause and effect relation between soft skills and project success. Only one study presents a model between variables. It means that it is important to find the cause and effect relation between personality traits, attitudes including soft skills, behavior and performance, and finally the project success.

4.4 Limitations of this review

This review has been based on literature referring to IT projects. There are studies that address soft skills for project success in general. Projects in general have not been included. The literature included has been from sources of IT project literature, it has not taken any source related to human resource literature. Articles were selected from 2010 onwards due to the speed and amplitude of changes that have occurred in technology in recent years. The review included articles that addressed both technical and soft skills. The soft skill selection has been done according to the author's understanding following to the definitions established in the background section.

5 Conclusions and Future Work

The studies conclude that personal skills contribute to project success. The most cited soft skills for project´s success are communication, leadership and conflict resolution skills. However, there is no coincidence among the authors about which is the most relevant soft skill. There is no coincidence about categories that group factors. Some studies conclude that the most valued project manager soft skills depend on some project factors such as the type or characteristics of the project or the project phase. In addition to that, there is not cause-effect model that explains how project manager soft skills impact to project success. These gaps suggest future researches.

References

1. Mkoba, E., & Marnewick, C.: IT Project Success: A Conceptual Framework for IT Project Auditing Assurance. Paper presented at the Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists (2016).
2. Andersen, E. S., Birchall, D., Arne Jessen, S., & Money, A. H.: Exploring project success. Baltic Journal of Management, 1(2), 127--147 (2006)
3. Bassellier, G., & Blaize Horner Reich, I. B.: Information technology competence of business managers: A definition and research model. Journal of Management Information Systems, 17(4), 159--182 (2001)
4. Bono, J. E., & Vey, M. A.: Toward understanding emotional management at work: A quantitative review of emotional labor research Härtel, Charmine E. (Ed); Zerbe, Wilfred J. (Ed); Ashkanasy, Neal M. (Ed). (2005)
5. Gale, A.: How to know what: Setting the project management competency agenda. PM Days (1999)
6. Hyväri, I.: Project management effectiveness in project-oriented business organizations. International Journal of Project Management, 24(3), 216--225 (2006)
7. Pinto, J. K., & Slevin, D. P.: Project success: definitions and measurement techniques. Project Management Journal, 19(1), 67--72 (1988)
8. Leybourne, S. A.: The changing bias of project management research. Project Management Journal, 38(1), 61--73 (2007)
9. André, M., Baldoquín, M. G., & Acuña, S. T.: Formal model for assigning human resources to teams in software projects. Information and Software Technology, 53(3), 259--275 (2011)
10. Chipulu, M., Neoh, J. G., Ojiako, U. U., & Williams, T.: A multidimensional analysis of project manager competences. IEEE Transactions on Engineering Management, 60(3), 506---517 (2013)
11. Stevenson, D. H., & Starkweather, J. A.: PM critical competency index: IT execs prefer soft skills. International Journal of Project Management, 28(7), 663--671 (2010)
12. Rose, K. H., A.: Guide to the Project Management Body of Knowledge (PMBOK® Guide)—Fifth Edition. Project Management Journal, 44(3) (2013)
13. Baccarini, D.: The logical framework method for defining project success. Project Management Journal, 30(4), 25--32 (1999)
14. Liebowitz, J., IT Project Failures: What Management Can Learn. IT Professional, 17(6), 8—9 (2015).
15. Lim, C., & Mohamed, M. Z.: Criteria of project success: an exploratory re-examination. International Journal of Project Management, 17(4), 243--248 (1999)
16. Diallo, A., & Thuillier, D.: The success dimensions of international development projects: the perceptions of African project coordinators. International Journal of Project Management, 22(1), 19—31 (2004)
17. Munns, A., & Bjeirmi, B. F.: The role of project management in achieving project success. International Journal of Project Management, 14(2), 81—87 (1966)
18. Crawford, L., & Turner, J.: Developing the project management competence of individuals. Gower handbook of project management, 4, 678--694 (2007)
19. Association, I. P. M., & Caupin, G.: IPMA competence baseline: ICB; Version 3.0: Internat. Project Management Association (2006)
20. Committee, P. S.: Project manager competency development (PMCD) framework. Project Management Institute (PMI), Newtown Square, Pennsylvania (2002).
21. Robles, M. M.: Executive perceptions of the top 10 soft skills needed in today's workplace. Business Communication Quarterly, 75(4), 453--465 (2012).
22. Kitchenham, B., & Charters, S.: Guidelines for performing systematic literature reviews in software engineering Technical report, Ver. 2.3 EBSE Technical Report. (2007)

23. Tore, D., & Torgeir, D.: Empirical studies of agile software development: A systematic review. Information and Software Technology, 50(9-10), 833--859 (2008).
24. Müller, R., Geraldi, J., & Turner, J. R.: Relationships between leadership and success in different types of project complexities. IEEE Transactions on Engineering Management, 59(1), 77--90 (2012).
25. Skulmoski, G. J., & Hartman, F. T.: Information systems project manager soft competencies: A project- phase investigation. Project Management Journal, 41(1), 61--80 (2010)
26. Creasy, T., & Anantatmula, V. S.: From every direction—How personality traits and dimensions of project managers can conceptually affect project success. Project Management Journal, 44(6), 36--51 (2013).
27. Keil, M., Lee, H. K., & Deng, T.: Understanding the most critical skills for managing IT projects: A Delphi study of IT project managers. Information & Management, 50(7), 398--414 (2013).
28. Colomo-Palacios, R., González-Carrasco, I., López-Cuadrado, J. L., Trigo, A., & Varajao, J. E.: I-Competere: Using applied intelligence in search of competency gaps in software project managers. Information Systems Frontiers, 16(4), 607--625 (2014).
29. Frese, R., & Sauter, V.: Improving your odds for software project success. IEEE Engineering Management Review, 42(4), 125--131 (2014).
30. Alias, Z., Zawawi, E., Yusof, K., & Aris, N.: Determining critical success factors of project management practice: A conceptual framework. Procedia-Social and Behavioral Sciences, 153, 61--69 (2014).
31. McCroskey, J. C.: Oral communication apprehension: A summary of recent theory and research. Human communication research, 4(1), 78-96 (1977)
32. Goleman, D. Emotional Intelligence, Bantam Books, New York, NY, 1995

Knowledge Management

Architecture for the integration of Linked Open Drug Data in an Augmented Reality application for mobile devices

Carlos Daniel Flores-Flores[1], José Luis Sánchez-Cervantes[2], Giner Alor-Hernández[1], Lisbeth Rodríguez-Mazahua[1], Luis Ángel Reyes-Hernández[1]

[1]Division of Research and Postgraduate Studies, Instituto Tecnológico de Orizaba
Av. Oriente 9 No. 852. Col. Emiliano Zapata. C.P. 94320, Orizaba, Ver. México,
Télefono/Fax: +52(272) 725 7056
{carlos.daniel.ff, l.a.reyes.h}@gmail.com, {galor, lrodriguez}@itorizaba.edu.mx

[2]CONACYT- Instituto Tecnológico de Orizaba
Av. Oriente 9 No. 852. Col. Emiliano Zapata. C.P. 94320, Orizaba, Ver. México,
Télefono/Fax: +52(272) 725 7056
jlsanchez@conacyt.mx

Abstract. There is an increasing amount of information contained in the knowledge bases belonging to the Linked Open Data (LOD) cloud corresponding to different domains, the exploitation of this information is limited by the ways in which it can be visualized, requiring new forms of visualization so that such information can be understood and used. Augmented Reality (AR) is a technology that allows the interaction of information with the user and also has a significant growth thanks to the proliferation of mobile devices with better processing. This work proposes an architecture for the integration of the datasets belonging to the LOD cloud with AR in an application for mobile devices, which makes it possible for information retrieved from the SPARQL-based query execution on RDF datasets, to be visualized in a way that allows improving the interaction between the user and data retrieved for increasing the understanding of the information shown.

Keywords: AR, architecture, Linked Open dataset, LOD cloud, mobile devices.

1 Introduction

Currently the huge amount of information contained in knowledge bases belonging to the LOD cloud [1] includes different domains. Their exploitation has allowed to obtain information that is presented commonly in tabular form, however, this form of displaying the information is little enriched what makes the user's interaction with it difficult.

Meanwhile AR is a technology of great growth, since the proliferation of the mobile devices, which thanks to their improvements in processing have been able to use

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_15

this technology. AR has allowed the interaction between the information presented as 3D models, videos, images, among other, with the user. The current approaches of AR applications commonly use the client-server architectures.

Considering the above, the main contribution of this paper is providing an architecture for the integration of Linked Open datasets with AR in an application for mobile devices, in order that information obtained from SPARQL-based query execution on the RDF datasets that belongs to the LOD cloud and contains information about medications can be presented to the user by means of AR and in this way improving the interaction, exploitation and understating of such information by the user.

This paper has been structured as follows: The Section 2 presents a set of recent works related to our proposal, organizing them in the works presenting topics of AR and LOD cloud separately and the works that integrate those technologies. In Section 3 the proposed architecture, the layers that integrate it, as well as the components of each layer. Additionally, the Section 4 includes a case study and finally, the Section 5 presents the conclusions and future work.

2 Related Work

There are several papers on the interaction of the users with data obtained from the LOD cloud, as well as the interaction of the users with applications of AR. We have classified the set of works reviewed for this research into two categories: 1) LOD and AR System applied in different domains, and 2) Integrated AR and LOD System.

2.1 Linked Open Data and Augmented Reality System applied in different domains

Similarly, in the medical environment the theory of simulation, AR and the learning based on game were fused for the development of ARISE (Augmented Reality Integrated Simulation Education) [2]. ARISE is based on games for healthcare students, it also contains scenarios for infirmary, medical assistant, infirmary assistant, among other. To test ARISE, prototypes of scenarios were designed and tests were carried out with people from the healthcare field with positive results. It has been considered that ARISE is an emergent, versatile and innovative form of educating the students.

Salmi et al. [3] developed a prototype educational of mobile application called HuMAR (Human Anatomy in Mobile Augmented Reality), whose selected learning topic was the anatomy of the structure of the human skeleton. HuMAR was created to help students improve their learning process. The HuMAR prototype was tested by science students, the objective of these tests was to consolidate the experience of the users from a didactic and technical point of view. Based on test results, it has been concluded that students were satisfied with HuMAR in terms of usability and features.

There is a wealth of medication information available on the Web, these data are not connected to each other, which reduces the ease of obtaining knowledge. For this reason, in [4] it was presented LODD (Linked Open Drug Data) that is a group of works of W3C (World Wide Web Consortium) HCLS IG (Health Care and Life Sciences Interest Group). LODD has researched available drug data and created Linked

Data representations.

Likewise, Kozák et al. [5] identified the sources of structured and unstructured data about medicines on the Web according to the information needs of physicians. Among the structured data sources identified are: Medical Subject Headings (MeSH), DrugBank, among others. While the non-structured data sources identified were Summaries Product Characteristics (SPC) and FDA (Food and Drug Administration) labels. They also created an architecture to integrate these sources, using the principles of Linked Data and methods of NLP (Natural Language Processing). In addition, they developed a Web application that exploits datasets and links obtained with support for clinical decision making.

2.2 Integrated Augmented Reality and Linked Open Data System

The AR browser applications are characterized by the display of Points of Interest (POI) obtained from the execution of queries on spatial databases. The architecture of these applications is composed by three parts: 1) The AR browser; 2) An AR Server, and 3) A POI server that manages and stores the content. In this architecture, the selection and integration of information sources is static and non-scalable, and browsing experience does not support the discovery and exploration of new data. It is for this reason that in [6] it was discussed how LOD can be used to address the current deficiencies in the mobile AR applications, the problems that arise when combining these technologies and the ways to solve them.

In [7] a computational model based on a mathematical archetype was presented, this model combines a sensor-based tracking approach with a mobile RDF processing and management framework for the provision and complementation of Linked Data, with related geographical information, to allow the visualization of POIs in mobile applications interfaces of AR. The applicability of the proposed model has been demonstrated by a proof of concept that retrieves the data of a mountain from a set of sources and shows that data in a live view interface.

The demonstration of an application of Mobile AR focused on tourism that shows digital content with information obtained from different LOD sources was presented by Vert and Vasiu [8]. They designed a three-layer architecture: 1) Usable dataset obtained from DBpedia.org, and the Romanian government's portal; 2) Summary, mapping and integration of the data using LDIF (Linked Data Integration Framework), in addition, an API (Application Programming Interface) was built to query and return data in JSON (JavaScript Object Notation) format, and 3) The application of AR using Javascript libraries like awe.js and three.js. The AR application for mobile devices was also built, which runs on the browser.

The same authors, Vert and Vasiu [9], proposed a model with a set of guidelines for the implementation of a mobile AR tourist application that integrated LOD and the analysis of the utility of the integration of multiple sources of Linked Data information to highlight the problems that are presented and propose solutions. The model for the construction of the mobile application is composed of the stages of identification, modeling, generation, publication, integration and exploitation. The prototype was built, which is a tourist application from Romania for use in the web browser of a mobile device. In addition, it was proposed to improve the integration of data by including an additional dataset, a more detailed analysis, among others.

On the other hand, in [10] it was tried to take advantage of the mobile and semantic technologies like LOD in a collaborative e-learning environment developing the application SmartTourism. The architecture of this application implies: 1) Obtaining information from Open Data and semantic sources; 2) Integrating information from data sources and; 3) Showing information to the tourist through the resulting tour guide service. SmartTourism are allows users to select the city and the language which they want to use, it shows in a map the most important monuments in the city and from each monument it is possible to see their history and images. With the creation of this application it was concluded that the semantic technologies as ontologies and LOD are powerful tools to share information and to generate knowledge about a topic.

Table 1 presents a comparative between the related works and the own one, it indicates the technologies used by each work, it also indicates if the work showed an architecture and the domain for which it was developed.

Table 1. Comparative table between the related work and the present work.

Initiative	Architecture	AR	LOD	Domain
ARISE [2]	X	✓	X	Medical
HuMAR [3]	X	✓	X	Medical
LODD for pharmaceutical research and development [4]	X	X	✓	Medical
LOD for Healthcare Professionals [5]	✓	X	✓	Medical
Exploiting LOD for mobile AR [6]	✓	✓	X	Research
Computational Model for the Integration of Linked Data in Mobile AR Applications [7]	X	✓	✓	Research
LOD4AR [8]	✓	✓	✓	Tourism
Integrating Linked Data in Mobile AR Applications [9]	X	✓	✓	Tourism
SmartTourism [10]	✓	X	✓	Tourism
Architecture for the integration of Linked Open Drug Data in AR applications	✓	✓	✓	Medical

As can be seen in Table 1, the development and research on AR and LOD technologies is increasingly widespread, and occurs in different contexts, with tourism and medicine the most recurrent. An aspect to emphasize after the analysis of the literature that has been realized is that, although there are several projects that have performed the integration of AR with LOD, this integration is limited to the tourist environment and only a few works showed an architecture; reason why an architecture has been designed which is presented in the next section to integrate the RDF datasets of the LOD cloud with AR.

3 Architecture description

The integration architecture of AR and RDF datasets of the LOD cloud, allows us to query datasets that are part of the LOD cloud corresponding to the life sciences domain, more specifically regarding medicines. Through this architecture, users can carry out semantic searches through the recognition of drug makers to obtain information about those drug from Linked Open Drug Datasets [11], [4]. Therefore, the obtained information is more accurate since it is based on the semantic properties belonging to the dataset chosen by the user.

3.1 Architecture description

The architecture shown in Fig. 1 is based on six layers. Each layer contains components, some of them have been divided in subcomponents. The tasks and responsibilities of the application are distributed among its components. The architecture allows scalability and easy maintenance.

Likewise, the architecture is shown in Fig. 1, this helps to understand graphically the structural components that comprise it, their relations, as well as the workflow that will be carried out for the communication of said components.

Each layer has a function explained below:

- **Presentation Layer:** This layer represents the interface between the user and the application. It contains components that make possible the tactile interaction of the user with the application.
- **Integration Layer:** This layer contains the necessary components to send and receive data and thus interact with the presentation layer. This layer includes the selection of the *Endpoint Selector*, as well as the construction of the information that will be shown to the user.
- **Augmented Reality Layer:** It contains the necessary components to process the marker pointed by the user and search for the information about that marker, as well as the corresponding 3D model.
- **Semantic Layer:** This layer contains the components that interpret, integrate and extract the requested data from the results obtained from the queries made to the selected medications dataset belonging to the LOD cloud.
- **Linked Open dataset:** In this section, we have a selection of datasets belonging to the LOD cloud that contain RDF triplets with information about medications, these datasets were selected for this specific architecture.
- **3D Models Repository:** For this architecture, it was considered to have a repository with 3D models that are presented to users through the application actually increases, depending on the recognized marker.

3.2 Components description

The components that are part of each layer of the architecture (Fig. 1) have functions that define their behavior, this is briefly explained below:

- **Mobile Application:** The mobile application contains an intuitive interface that allows the user to access the camera of the mobile device and through it obtain the marker that can be the bar code or QR code of the drug and based on this perform a semantic search. In the same way, it allows to visualize the obtained results as 3D models and medication information obtained from the Linked Open datasets.
- **Dataset Selector:** This component provides users with the option to select one of the Linked Open datasets of selected medications for this architecture such as: DrugBank[12], Diseasome[4], BIO2RDF[13] and DailyMed[4], to mention but a few. In addition, the dataset selector component sends the Uniform Resource Identifier (URI) of the dataset to obtain the correct SPARQL endpoints (namespace).
  - **Endpoint Selector:** This subcomponent obtains and sends the namespace for each dataset to the mobile application so that it sends it to the *AR module* along with the marker.
- **Response Builder:** This component is responsible for receiving the responses from SPARQL-based queries. This module retrieves the useful information of the answers from the selected Linked Open dataset, as well as the 3D model to be displayed, and sends that information to the mobile application to be shown to the user.
- **AR module:** This component is in charge of receiving the Endpoint selected by the user, as well as the marker, sending that information to the recognition component of markers, it is also in charge of sending the information retrieved from the queries to the Linked Open dataset and the 3D model selected to the *Response Builder*. Within this component is the *Vuforia API*:
  - **Vuforia API:** Allows the application to process the 3D models, markers and information to be shown to the user.
- **Marker Recognition:** This component is responsible for receiving the marker captured by the camera and processed by the *AR module*, to obtain from it the keyword which will be sent to the *Data Query Manager* for the queries to the selected datasets and the *3D Model Selector* to find the required model. It will also receive the information retrieved from the *Linked Open Data Extractor* component and send it to the *AR module*.
- **3D Model Selector:** This component is responsible for receiving the keyword to search, and based on it, it searches and selects the 3D model to be displayed and sends it to the *AR module* to be processed.
- **Data Query Manager:** Depending on the source selected in the *Dataset Selector* component, this component will retrieve the information contained in the LOD cloud dataset. This component has a subcomponent:
  - **Semantic Linked Data Query Processor:** This subcomponent executes the process to obtain the information contained in the datasets, the thrown process implies the execution of SPARQL-based queries including the keyword obtained in the Marker Recognition component.
- **Linked Open Data Extractor:** This component aims to analyze the information stored in the dataset previously selected of the LOD cloud, and in case it is necessary to extract it. This component has a subcomponent:

  - o **Data Extractor:** With the dataset selected to extract the information and the Semantic LD *Query processor* component executed, this subcomponent is responsible for extracting the appropriate information about the drug to be searched.
- **Linked Open dataset – RDF:** The Linked Open dataset contains structured data information from different resources on the Web where SPARQL-based queries will be applied. For this case only datasets with medications information are used.
- **3D Model Repository:** A 3D model is a communication tool for compression, it is a three-dimensional representation of a real object in a virtual space.

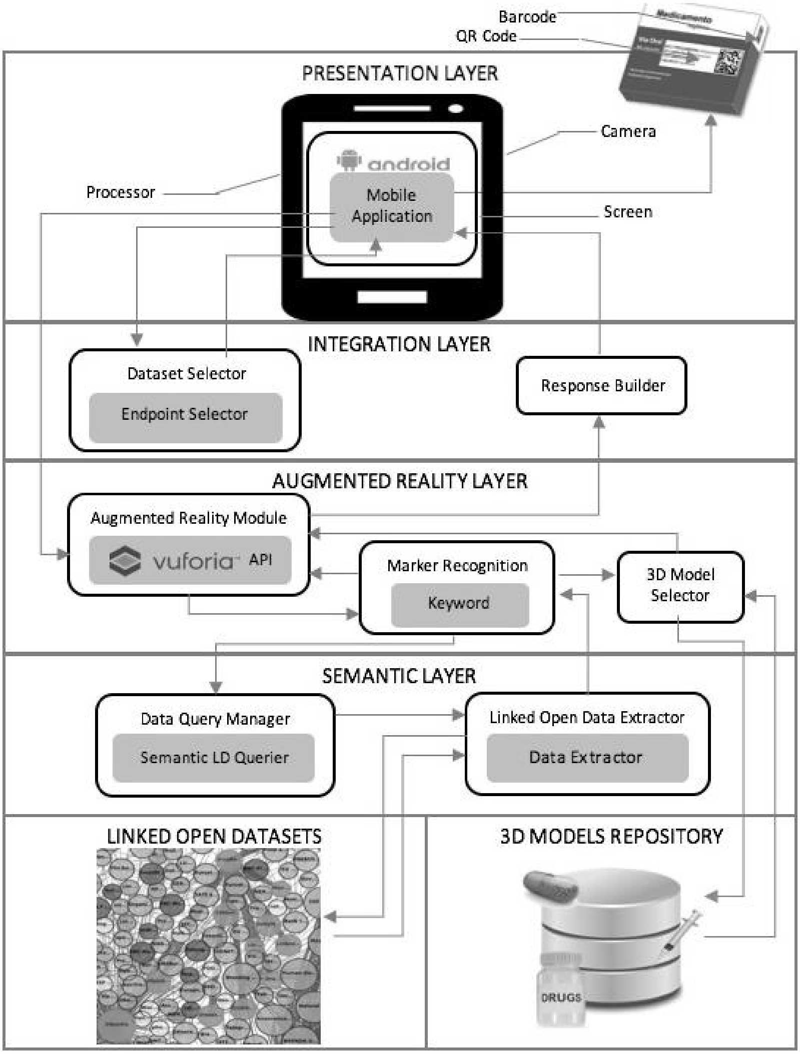


Fig. 1. Integration architecture of AR and RDF datasets of the LODD cloud.

3.3 Workflow description

The interrelationships among components and subcomponents of the architecture shown in Fig. 1 define the workflow for the LOD cloud integration process with AR. Users have the possibility of using a mobile device to recognize a marker belonging to a medicament and retrieve information about it from a selected dataset, and that information retrieved is displayed to the user by complementing it with a 3D model so that the user can interact with it. The architecture workflow is briefly described below:

1. By using the mobile application, the user will select through the *Dataset Selector* the endpoint that the application allows, from which the user wants to obtain the information. The *Dataset Selector* will send the URI of the selected dataset returned to the application.
2. After the endpoint is selected, the device camera will be activated and the user will have to point it to the marker contained in the medicine box.
3. This marker will be received by the *AR module* and the *Vuforia API* and will be sends to the *Marker Recognition* module.
4. The Marker Recognizer will analyze the marker and get the keyword to search from it, the keyword will be sent to the *Data Query Manager* along with the URI of the selected endpoint.
5. With the URI of the selected dataset and the keyword to search, the *Data Query Manager* component launches a process to retrieve the information stored in the selected datasets in the *Dataset Selector* component.
6. During the process of querying and retrieving data from the datasets, the *Linked Open Data Extractor* component analyzes the information retrieved and, if relevant, extracts it.
7. When the information is extracted the results are sent to the *Marker Recognizer*, these results send a keyword to the component selector of 3D models, and likewise send the retrieved information to the *AR module*.
8. The *3D Model Selector* with the keyword commands to search the repository of 3D models.
9. The 3D Models Repository returns the selected model to the *3D Model Selector*.
10. The *3D Model Selector* sends the 3D model to the *AR module*.
11. The *AR module* with its *Vuforia API* prepares the 3D model together with the retrieved information to be presented to the user and sends it to the *Response Builder* component.
12. The *Response Builder* component is responsible for preparing and giving the latest details to the 3D model and retrieved information and sending them to the Mobile Application.
13. Finally, the *Mobile Application* shows the 3D model and the recovered information superimposed on the marker pointed by the user so that he or she can interact with them.

The Fig. 2 depicts a UML sequence diagram with the interaction among components belonging to architecture and the previously described workflow.

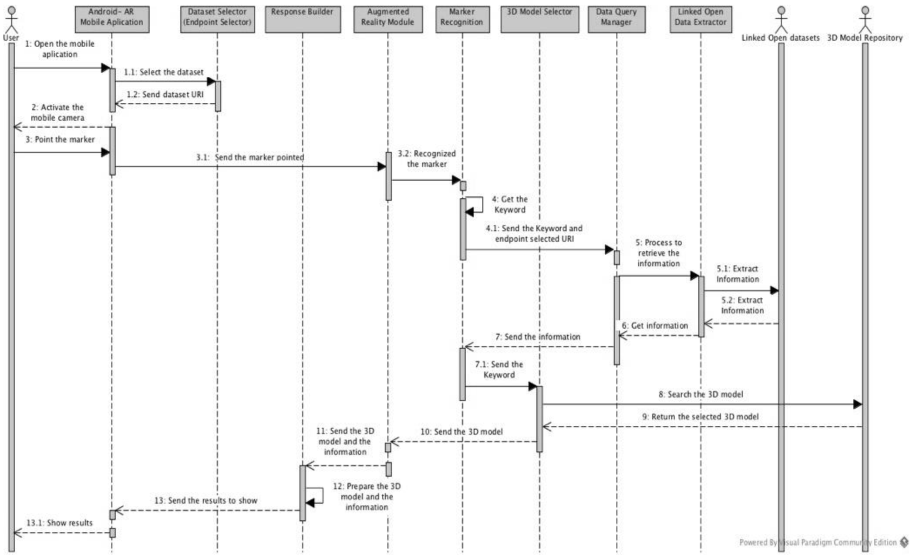


Fig. 2. UML sequence Diagram of the Proposed Integration Architecture's Workflow.

In the workflow of the architecture, the extraction of drug information is carried out in the Semantic Layer using the *Data Query Manager* and *Linked Open Data Extractor* components. Is important to mention that in the analysis of the literature we did not find an architecture designed for native applications of mobile devices which integrates AR with the support for loading 3D models and that it obtains information from triplestores of the medical domain available in the LOD cloud through the SPARQL-based queries execution. On the other hand, this architecture allows to be configured easily for its use in other domains such as tourism, government, cross domain, geography and social networking, to mention but a few, through the loading of new markers, 3D models and the generation of SPARQL-based queries corresponding to required domain.

The use of *Vuforia API* for AR is based on some of its benefits, such as the easy creation of natural markers using images, a cloud database to store such markers, as well as a better recognition of markers [14] [15]. Additionally, *Vuforia API* provides an easy integration with the Unity development platform, which makes the developing applications for different devices will be faster and loading 3D models easier. For the implementation of the architecture the minimum requirements of software and hardware with which the mobile device must count, is the Android Operating System in its version 5.1, with a camera of 5 MP of 30 fps. with autofocus, 1GB RAM, WiFi connectivity, also integrated accelerometer and gyro and at least a 4-inch display for better viewing of information.

4 Case Study: Search a drug's information in an RDF dataset of the LOD cloud

For this case study, a search for information about a specific drug is performed on an RDF dataset of the LOD cloud on medications. The RDF dataset used is DrugBank, although it is possible to use other RDF dataset of drugs.

1. Let us suppose that a medical student needs to know different aspects of the drug Ibuprofen as are its indications, chemical formula, among other aspects.
2. Although it has a large number of pages with quality information about the drug, the student is not able to retain the information about the drug for a long time, so he has to constantly review that information.
3. Therefore, the student search for other options that help him obtain reliable information about the drug and also allow him to interact with this information in order to improve the retention of the same.

Our solution alternative is based on the proposed integration architecture for the development of a mobile application that integrates AR with LODD cloud in order to provide the user, in this case the medical student, an intuitive way to improve the understanding the information of the drug consulted. The user would point with the mobile device's camera to the marker present on the medicine packaging, this can be its bar code, from recognition of the marker would get the keyword to search in the RDF dataset DrugBank for more information on the drug.

Listing 1 depicts the SPARQL-based query, which is executed in the RDF dataset DrugBank and it allows retrieving information about the drug Ibuprofen in which case its ID for searching in the RDF dataset DrugBank is DB1050.

Listing 1. SPARQL-based query to retrieving information Ibuprofen.

```
PREFIX drug:  <http://wifo5-04.informatik.uni-
mannheim.de/drugbank/resource/drugs/>
SELECT DISTINCT ?name ?description ?indication ?state
?foodInteraction ?chemicalFormula
WHERE {
  drug:DB01050 rdfs:label ?name .
  drug:DB01050 drugbank:description ?description .
  drug:DB00415 drugbank:indication ?indication .
  drug:DB01050 drugbank:state ?state .
  drug:DB01050 drugbank:foodInteraction ?foodInteraction
.
  drug:DB01050 drugbank:chemicalFormula ?chemicalFormula
.
}
```

Fig. 3 shows the results retrieved by the SPARQL-based query which are information about the drug Ibuprofen. It must be mentioned that it is possible to obtain more information about the drug, however, only some of them are represented.

name	description	indication	state	foodInteraction	chemicalFormula
"Ibuprofen"	"A nonsteroidal anti-inflammatory agent with analgesic properties used in the therapy of rheumatism and arthritis. [PubChem]"	"For treatment of infection (Respiratory, GI, UTI and meningitis) due to E. coli, P. mirabilis, enterococci, Shigella, S. typhosa and other Salmonella, nonpenicillinase-producing N. gononhoeae, H. influenzae, staphylococci, streptococci including streptoc"	"Solid"	"Avoid alcohol."	"C13H18O2"
"Ibuprofen"	"A nonsteroidal anti-inflammatory agent with analgesic properties used in the therapy of rheumatism and arthritis. [PubChem]"	"For treatment of infection (Respiratory, GI, UTI and meningitis) due to E. coli, P. mirabilis, enterococci, Shigella, S. typhosa and other Salmonella, nonpenicillinase-producing N. gononhoeae, H. influenzae, staphylococci, streptococci including streptoc"	"Solid"	"Take with food to reduce irritation."	"C13H18O2"

Fig. 3. SPARQL-based query results about Ibuprofen.

As presented in the Fig. 3 the results are retrieved in tabular form, so that in our solution alternative it is intended that the interaction between these results and the user be improved by the integration of 3D models that represent some of these results recovered. Figures 4a-c, depict how some of the results retrieved would be represented by 3D models.

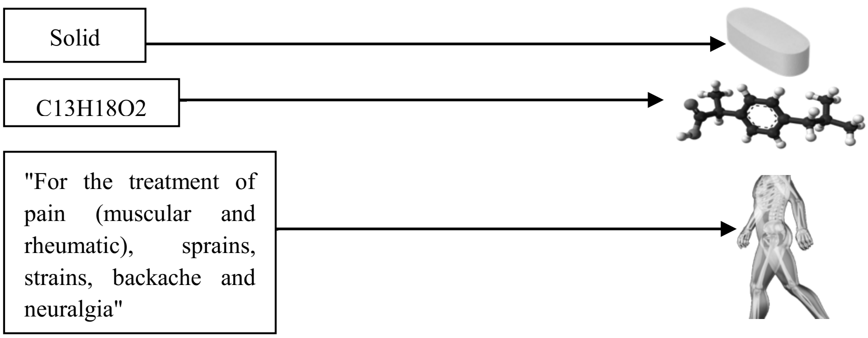


Fig. 4a, 4b and 4c. Different recovered results represented by 3D models.

The proposed solution pretends that 3D models representing results, as shown in Figures 4a-c, are presented to the user by utilizing AR, enhancing the user's interaction with the information so that by presenting the information in this way it is easier for the user to understand, learn and retain information about medicines and in this way expanding the possibilities of learning through the use of technologies that are practically within the reach of all like a smartphone. Another aspect to emphasize is that RDF datasets are intended to be used with information from sources fed by scientists and doctors, which makes them information of great quality and reliability, reason why the users can be assured that it is reliable information.

5 Conclusions and Future Work

The information obtained from knowledge bases belonging to the LOD cloud is presented in tabular or graphical form and, although this way of presenting the data is easy to read for most users, it does not improve the interaction between them and the information obtained, therefore, AR can be used as an alternative to solve this interaction, for which we analyzed several works related to the LOD cloud, AR and the integration of these technologies, based on this analysis we can conclude that these technologies complement one another. On the one hand, AR helps to improve users' visualization and interaction with information retrieved from the RDF dataset of the LOD cloud queries, while retrieved information helps to overcome constraints on without growth and static information of AR applications.

As a future work, we have contemplated the development of the mobile application of AR to show the usefulness of the architecture, for the development of this application we must construct the AR module for the recognition of markers as well as to present the 3D models, similarly it will be constructed the module for the consultation and retrieval of information through queries to the RDF datasets of the LOD cloud, this module will be based on the proposed architecture.

6 Acknowledgements

Authors are grateful to the National Technological of Mexico for supporting this work. This research paper was also sponsored by the National Council of Science and Technology (CONACYT) and the Secretariat of Public Education (SEP) through PRODEP.

References

1. “The Linking Open Data cloud diagram.” [Online]. Available: http://lod-cloud.net/. [Accessed: 19-Feb-2017].
2. K. J. Carlson and D. J. Gagnon, “Augmented Reality Integrated Simulation Education in Health Care,” Clin. Simul. Nurs., vol. 12, no. 4, pp. 123–127, 2016.
3. S. Salmi, J. Ab, M. F. Shiratuddin, K. W. Wong, and C. L. Oskam, “Utilising Mobile-Augmented Reality for Learning Human Anatomy,” in Procedia - Social and Behavioral Sciences, 2015, vol. 197, pp. 659–668.
4. M. Samwald et al., “Linked open drug data for pharmaceutical research and development,” J. Cheminform., vol. 3, no. 1, pp. 19–24, 2011.
5. J. Kozák, M. Nečaský, J. Dědek, J. Klímek, and J. Pokorný, “Linked Open Data for Healthcare Professionals,” in Proceedings of International Conference on Information Integration and Web-based Applications & Services, 2013, pp. 400–409.
6. V. Reynolds and M. Hausenblas, “Exploiting linked open data for mobile augmented reality,” W3C Work. Augment. Real. Web, vol. 1, pp. 1–6, 2010.
7. S. Zander, C. Chiu, and G. Sageder, “A Computational Model for the Integration of Linked Data in Mobile Augmented Reality Applications,” in Proceedings of the 8th International Conference on Semantic Systems, 2012, pp. 133–140.
8. S. Vert, B. Dragulescu, and R. Vasiu, “LOD4AR: Exploring Linked Open Data with a Mobile Augmented Reality Web Application,” in Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272, 2014, pp. 185–188.
9. S. Vert and R. Vasiu, “Integrating Linked Data in Mobile Augmented Reality Applications,” in Proceedings of Information and Software Technologies: 20th International Con-

ference, International Conference on Information Science and Technology (ICIST) 2014, Druskininkai, Lithuania, October 9-10, 2014., G. Dregvaite and R. Damasevicius, Eds. Cham: Springer International Publishing, 2014, pp. 324–333.

10. A. M. Fermoso, M. Mateos, M. E. Beato, and R. Berjón, “Open linked data and mobile devices as e-tourism tools. A practical approach to collaborative e-learning,” Comput. Human Behav., vol. 51, pp. 618–626, 2015.
11. A. Jentzsch, M. Samwald, and B. Andersson, “Linking Open Drug Data,” I-Semantics ’09 Proc. Int. Conf. Semant. Syst., pp. 3–6, 2009.
12. D. S. Wishart et al., “DrugBank: a comprehensive resource for in silico drug discovery and exploration,” Nucleic Acids Res., vol. 34, no. suppl_1, pp. 668–672, 2006.
13. F. Belleau, M. A. Nolin, N. Tourigny, P. Rigault, and J. Morissette, “Bio2RDF: Towards a mashup to build bioinformatics knowledge systems,” J. Biomed. Inform., vol. 41, no. 5, pp. 706–716, 2008.
14. D. Amin and S. Govilkar, “Comparative Study of Augmented Reality Sdk’s,” Int. J. Comput. Sci. Appl., vol. 5, no. 1, pp. 11–26, 2015.
15. A. Serrano, “Herramientas de desarrollo libres para aplicaciones de Realidad Aumentada con Android. Análisis comparativo entre ellas,” Universidad Politécnica de Valencia, 2012.

An Architecture based in Voice Command Recognition for faceted search in Linked Open Datasets

Betia Lizbeth López-Ochoa[1], José Luis Sánchez-Cervantes[2], Giner Alor-Hernández[1], Ma. Antonieta Abud-Figueroa[1], Beatriz A. Olivares-Zepahua[1], Lisbeth Rodríguez-Mazahua[1]

[1]Division of Research and Postgraduate Studies, Instituto Tecnológico de Orizaba
Av. Oriente 9 No. 852. Col. Emiliano Zapata. C.P. 94320, Orizaba, Ver. México,
Télefono/Fax: +52(272) 725 7056
betializbeth@gmail.com, {galor, lrodriguez}@itorizaba.edu.mx,
{mabud, bolivares}@ito-depi.edu.mx

[2]CONACYT- Instituto Tecnológico de Orizaba
Av. Oriente 9 No. 852. Col. Emiliano Zapata. C.P. 94320, Orizaba, Ver. México,
Télefono/Fax: +52(272) 725 7056
jlsanchez@conacyt.mx

Abstract. Faceted browsers have become a popular interface paradigm, since they combine the visualization of data that are part of a graph with data filtering techniques. On the other hand, NLP (Natural Language Processing) allows the use of everyday or natural language to interact with computer systems. LOD (Linked Open Data) cloud is the collection of datasets published in Linked Data format and covers a large number of domains, but the interaction with them is intended to be exploited by experts. Because of the problematic raised it is proposed the creation of an integration architecture for Linked Open datasets in a faceted browser with recognition of voice commands, so that through the NLP, on voice commands issued by the user, SPARQL queries will be generated to perform the search and navigation among the data stored in datasets available in the LOD cloud.

Keywords: Faceted navigation, Linked Data, Natural Language Processing, Voice recognition.

1 Introduction

LOD cloud [1] is the union of structured datasets published in Linked Data format, relationships between entities are found in it and allows making new discoveries of related information. It covers various domains such as: geography, government, social networks, life sciences, among others [2]; however, current mechanisms for retrieving semantic information are intended to be used by expert users, making the access to such information difficult for inexperienced users.

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_16

On the other hand, Web applications commonly incorporate interfaces that are limited to the capture of data through keyboard or touch screens, making the interaction become complicated and unnatural [3].

The main contribution of this paper is the creation of an architecture for Linked Open datasets in a faceted browser with recognition of voice commands, so that through the NLP, on voice commands issued by the user, SPARQL (SPARQL Protocol and RDF Query Language) queries will be generated to perform the search and navigation among the data stored in datasets available in the LOD cloud. Through the use of facets the users are provided with a list of results by which they will navigate to other data related to their search. This proposal will also give users the ability to discover additional information within Linked Datasets that are part of the LOD cloud.

We consider that the development of an architecture based on voice command recognition for faceted search in Linked Open Datasets will provide the following benefits: a) Obtaining information and systematically navigate large amounts of information that cannot be processed without automatic support; b) Unify information from triplestores linked through their properties (predicates), and c) Voice commands will be used as the main means of interaction between users and the faceted browsers.

This paper is structured as follows: Section 2 describes recent work related to our proposal. The layers and components of the proposed architecture are described in Section 3; Section 4 presents a case study for the search and faceted navigation on Linked Datasets corresponding to Life sciences domain, through voice commands. Finally, conclusions and future directions are addressed in Section 5.

2 Related Works

There are several works that studied how to obtain information from different Linked Open Datasets through queries, some of these works obtained encouraging results using multimodal interfaces for an easy and natural interaction of the users with the applications. The following, describes a set of initiatives related to the proposed work.

The semantic search satisfies a wide range of information retrieval needs compared to traditional full-text search. Based on the above problem, Tablan et al. [4] presented Mímir, an open source semantic search framework for interactive information seeking and discovery. Mímir performs an integrated semantic search over text, document structure, linguistic annotations and formal semantic knowledge, it also supports complex structural queries, as well as basic keyword search.

An architecture for interfaces based on NL (Natural Language) and techniques of automatic translation of statistics SMT (Statistical Machine Translation) was presented by Revuelta-Martínez et al. [5] to translate the sentences obtained from voice commands issued by the user in a formal query language that is sent to the database manager to retrieve the information.

The use of smartphones and tablets to access the Web has increased, but the reduced size of the screen and keyboards has caused interaction with graphical interfaces to be complicated. In this sense, in [6] a framework that is used to develop systems of adaptive oral dialogue was presented. Such framework combined knowledge representation, NLP, user modeling and intelligent information retrieval for facilitating personalized access to information through speech.

Heck et al. [7] proposed an open-source conversational system that combined the power of browser interfaces with multimodal entries and data extracted from search records. Two input modes have been considered for interacting with the browser, Web page interfaces and their elements: 1) Speech and hand gestures, and 2) The combination of both. Microsoft's Kinect sensor has been used for voice recognition and motion detection

Paredes-Valverde et al. [8] presented a NL interface that allows non-expert users to access databases of semantic knowledge, through NL queries. NLP, POS (Parts of Speech Tagging), lemmatization, NER (Named Entity Recognition), synonymous expansion and semantic annotation techniques were used to obtain semantic information of the questions. The interface consists of four tasks: a) Pre-processing of the knowledge base; B) Processing of the NL of the user's question; C) Generation of SPARQL-based queries, and d) SPARQL queries execution for the results recovery.

With the aim of achieving that users indicate by voice commands to a computer what they want to search, using any number of sentences, in [9] an approach to design and develop a semantic Web search system using NL was presented. In such an approach, an ontology set were used to store domain structure and data, as well as to describe the meaning of user queries.

Paredes-Valverde et al. [10] presented ONLI (Ontology-based Natural Language Interface) for DBpedia. User through voice commands, expresses a question, the question processing module performs the NLP of the question to obtain the semantic information of its elements, the elements that have been obtained are consulted in the knowledge base, after that the information is organized in the question model, then the search and response module manages the ambiguities and obtains the possible responses from the knowledge base, finally, these are organized by relevance and are shown to the user.

For communication between a human and a computer it is necessary that the latter has communication capabilities, to achieve such capabilities. Hence, Serón & Bobed [11] presented a built-in conversational agent system that exploits the knowledge provided by Linked Data to assist users in their search tasks, and returns only relevant results for a Mechanical domain, such results were presented as text, image or video formats.

In [12], MEANS, an automatic response system to medical questions with the combination of NLP techniques and semantic Web technologies, was presented, which allows a deep analysis of questions and documents. The system responds to questions in the English language that have been expressed through NL. NLP techniques were applied to analyze the source documents used to obtain the answers.

Although there are several works related to the proposed work, with the help of the comparative presented in Table 1, it is highlighted that most of the analyzed initiatives handle several modalities for the recovery of information such as speech, text, on-screen strokes, gestures, speech, to mention but a few, or through them they developed oral dialogue systems, however, not all works retrieve knowledge base information from the LOD cloud, but they use relational databases, multimedia or files of different formats.

Table 1. Comparative summary of multimodal search systems, semantic query and analyzed search engines.

Initiative	Modality	Context	Data source	NLP
Mímir [4].	Web Application	Environmental science and research	LOD	Yes
Multimodal interaction for information retrieval using NL [5].	Web application with interface based on NL	Transport	Relational DB	Yes
Framework to develop systems of adaptive oral dialogue [6].	Mobile application with voice recognition interface	Transport	Relational DB	Yes
Multimodal Conversational Search and Browse [7].	TV navigator with recognition of voice commands and gestures.	Not specified	Datasets from speakers	Yes
NL interface to access databases of semantic knowledge [8].	Web application with interface based on NL	Music	Linked Data	Yes
Semantic Web Search Using NL [9].	Web application with interface based on NL	Transport and Tourism	Knowledge base	Yes
Ontology-based system for querying DBpedia using NL paradigm [10].	Interface based on NL	Education	LOD cloud	Yes
Conversational agent system that exploits Linked Data [11].	Web application with multimodal interface with voice command recognition	Mechanics	Linked Data	Yes
Medical question-answering system [12].	Web application with voice command recognition	Medical	Data base	Yes
Integration Architecture based in Voice Command Recognition for faceted search in Linked Open Datasets.	Web application with voice command recognition	Medical	LOD cloud	Yes

The most of the works consume data on a single domain, different of medical, with the exception of the MEANS system, however, its corpus of responses does not include LOD cloud. After analyzing the literature, it is concluded that we do not find any work that covered the characteristics such as the search of information in the LOD cloud through voice commands, faceted navigation of data, integration of data from different triplestores and discovery of additional information in the Web, corresponding to the proposed work.

3 Architecture

The proposed architecture allows the querying of datasets that are part of the LOD cloud, the users will be able to perform semantic search and navigation using voice commands. This will grant only relevant results to the user, since through the use of facets the results are filtered in an iterative way.

3.1 Architecture description

The proposed architecture is based on layers in order to facilitate its organization and maintenance, each layer is integrated by components, and some of these in subcomponents which carry out tasks and functionalities that generate workflows among them. Fig. 1 depicts the architecture, its components, and the relationships among them to obtain a better understanding of it.

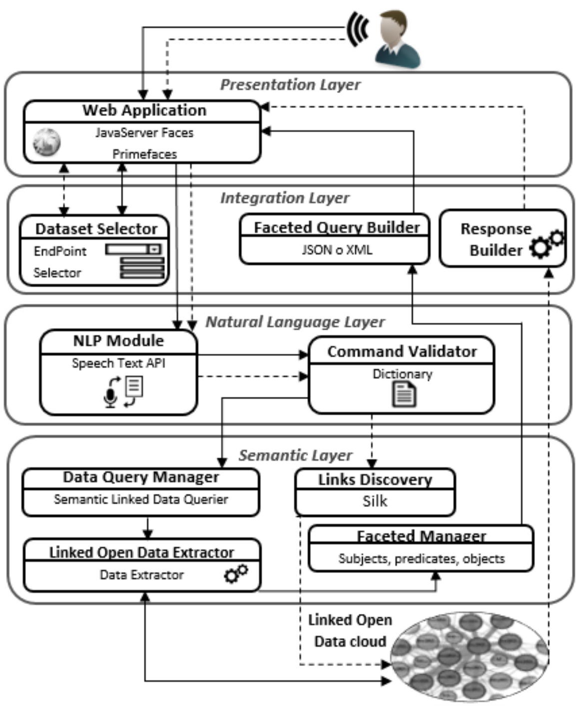


Fig. 1 Architecture based on Voice Command Recognition for faceted search in Linked Open Datasets.

The general function of each layer of the architecture is briefly described below:

- **Presentation Layer:** This layer represents the Web interface that exists between the user and the application, within it there are components that allow the interaction of the user with the application, such interaction is not limited only to the touch type, but also includes recognition of voice commands.
- **Integration Layer:** This layer has the necessary components to make the communication between *Presentation Layer* with *Semantic Layer* and the *Linked Open Data*.
- **Natural Language Layer:** In this layer, there is communication with the *Presentation Layer* through the *NLP Module* component, where the transformation of the voice commands to text is performed to be later processed by the *Command Validator* component that has communication with the *Semantic Layer*.
- **Semantic Layer:** This layer takes care of the integration of the components that interpret the queries that come from the *Natural Language Layer* and extracts the results obtained from the selected dataset before starting the search, such results are sent to the *Integration layer* for prior processing before being presented to the user.

Despite the LOD cloud is not considered as a layer of the architecture, it is depicted as an external component that, although it is not developed and managed by us, it is an essential component which allows the consumption of its data through its SPARQL Endpoints and the SPARQL-based query execution [1].

3.2 Components Description

The components in each layer have specific functions that determine their behavior and are explained briefly below:

- **Presentation Layer:**
  - **Web Application:** This application provides a GUI (Graphical User Interface) developed through the frameworks JavaServer Faces and Primefaces, the latter for the responsive design of the page. The GUI provides the voice commands given by the user, also the search and faceted navigation is performed, optionally the discovery of information is performed, and the results obtained are shown to the user.
- **Integration Layer:**
  - **Dataset Selector:** Through this component the user has the option of choosing, via voice commands, a Linked Open dataset available in the dataset dictionary. This component also sends the URI (Uniform Resource Identifier) of the datasets in order to get the appropriate SPARQL Endpoints, that is, the namespace.
    - **Endpoint Selector:** This subcomponent obtains and sends the namespace for each dataset to the *Data Selector* component.
  - **Faceted Query Builder:** This component uses a document in XML or JSON format built by the *Faceted Manager* component, extracts and processes the information so that the obtained facets are displayed in the *Web Application* GUI.

- **Response Builder:** This component receives the responses from SPARQL-based queries executed, it retrieves the useful information from the responses and creates a document with the result list related to the user search, such document will be presented in XHTML format to the user.

• **Natural Language Layer:**

- **NLP Module:** Its main function is to obtain the relations among the relevant words contained in the commands emitted by the user and the semantic relations between these words and the properties contained in the dictionary stored in the *Command Validator* component, to later send the commands like keywords to the *Data Query Manager* component.
- **Command Validator:** In this component, the dictionary of terms is stored, this is built with information about datasets that belong to determinate domains e.g., some of the datasets of the Life sciences domain included in the dictionary are: PubMed, Bio2RDF, BioPortal, to mention but a few. The dictionary also contains keywords that allow you to perform special actions such as "discovery" for discovery of additional information, "back" or "restart" for star a new search.

• **Semantic Layer:**

- **Data Query Manager:** This component initiates a process to retrieve the information in the datasets that involves the SPARQL-based query execution, including their properties that will allow navigation among more datasets through the discovery of links in the Web. When the component cannot retrieve the properties of the datasets it gets the list of properties, then the Semantic Linked Data Querier subcomponent executes a process to obtain the information stored in the datasets.
- **Linked Open Data Extractor:** The function of this component is to analyze the information stored in the LOD cloud and if at any time such information is necessary, it is extracted. The Data Extractor subcomponent extracts the appropriate information once the *Data Query Manager* module has determined what information is to be obtained from the Linked Data source and the *Semantic Linked Data Querier* subcomponent has been executed.
- **Faceted Manager:** This component is responsible for the creation of facets corresponding to the results obtained after the extraction process of information, later it will create an XML or JSON file that will be sent to the *Faceted Query Builder* component.
- **Links Discovery:** Once at least one user search has been performed, this component allows discovering information related to the search keywords provided by the user as voice commands, such functionality is possible with the use of the Silk Framework that will be responsible for discovering the links related with the search keyword.

• **Linked Open Datasets:** Although it is not an element of the proposed architecture, Linked Open dataset interacts with architecture as an external element. The Linked Open dataset contains the structured data information of different resources on the Web where the SPARQL-based queries will be applied.

3.3 Workflow description

The workflow for searching through keywords and using the selected facets is similar, in the first type of search users give keywords to perform their search, while in the second the user selects some facet of the results obtained during the search. For a better understanding, the workflow of the architecture is briefly explained below:

1. Through the Web application the user makes a request based on the HTTP (Hypertext Transfer Protocol) by entering a keyword or a selected facet.
2. Once the voice command has been received in the user interface, the *NLP Module* component performs the conversion of the voice command to text so that it is sent and analyzed using the *Command Validator* component that will determine if the command corresponds to a valid dataset and is considered a keyword, if so the dataset predicates are selected from dataset and predicate dictionaries.
3. When entering a keyword using voice commands or choosing a facet, the *Dataset Selector* component sends the URIs of the datasets to the *EndPoint Selector* subcomponent to obtain the appropriate SPARQL EndPoints.
4. The *Data Query Manager* component executes a process for retrieving the information stored in the selected datasets.
5. During the data retrieval process, the *Linked Open Data Extractor* component analyzes the information obtained and, if relevant, it is extracted by this component.
6. When the information is extracted, the *Faceted Manager* component obtains the results and creates a document in XML or JSON format with the information coming from the LOD cloud, such results will represent the facets through which users can navigate, the document is sent to the *Faceted Query Builder* component.
7. Based on the results stored in the XML or JSON document, the *Faceted Query Builder* component generates the facets that are displayed in the GUI of the *Web application* and through which the user can search for more information through voice commands or chooses some facet with results.
8. Finally, facets and results are displayed to the user via the GUI in the *Web application*.
9. Additionally, the user can perform related information discovery by repeating steps 1-3, the *Command Validator* component sends the keywords to the *Links Discovery* component, which will be responsible for carry out a mapping among the keywords and similar items included in DBpedia as nucleus of the LOD cloud [13] in order to discovery information related to user's search.
10. The *Builder Response* component receives the information obtained by the *Links Discovery* component, it prepares a list with the results that will be shown to the user through the GUI in the *Web application*. The user can choose some of the results obtained and the link pages will be opened externally to the *Web application*.

4 Case Study

For this case study, the domain selected within the LOD cloud is Life Sciences and as RDF dataset we chose BioPortal [14], a repository of biomedical ontologies.

SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) was considered as source of information, however, it is allowed to use other information sources only if they are in Linked Data format. Suppose that a pregnant patient was recently diagnosed with hypertension and is interested in knowing information related to the disease and its condition.

When the application opens a list of Linked Open dataset is shown and it requests to select one, which in this case is BioPortal, then it will issue the voice command "BioPortal", later the user issues a voice command to start navigation, it is validated that this first command is in a medical dictionary, in such case, the first SPARQL query is generated according to the search criteria, as already mentioned the data source for this case will be SNOMED CT so the command that initiates navigation is "SNOMED". Suppose that the patient issues the first command "Hypertension" and it exists in the dictionary, then a list of results is returned through which the user can navigate.

Suppose the patient observes that there is information about her pregnancy-related disease in the results and gives the voice command "Hypertension complicating pregnancy", then the SPARQL query shown in Fig. 2 is generated according to the search criteria and two results are obtained, where now the patient should give voice command as one of the elements deployed.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
  SELECT DISTINCT *
  FROM
<http://bioportal.bioontology.org/ontologies/SNOMEDCT>
  FROM
<http://bioportal.bioontology.org/ontologies/globals>
  WHERE {
    ?x rdfs:label ?label .
     FILTER CONTAINS(?label, ``Hypertension complicat-
ing
                         pregnancy'')
  }
```

Fig 2. SPARQL-based query executed in the scenario of Case Study

With the process described above the patient will navigate through the results obtained from SPARQL queries through voice commands, this way she will continue to navigate iteratively until she finds the results that satisfy her search or grant a special voice command as "restart" to start a new search. Fig. 3 shows a navigation graph that helps to better understand the aforementioned navigation flow.

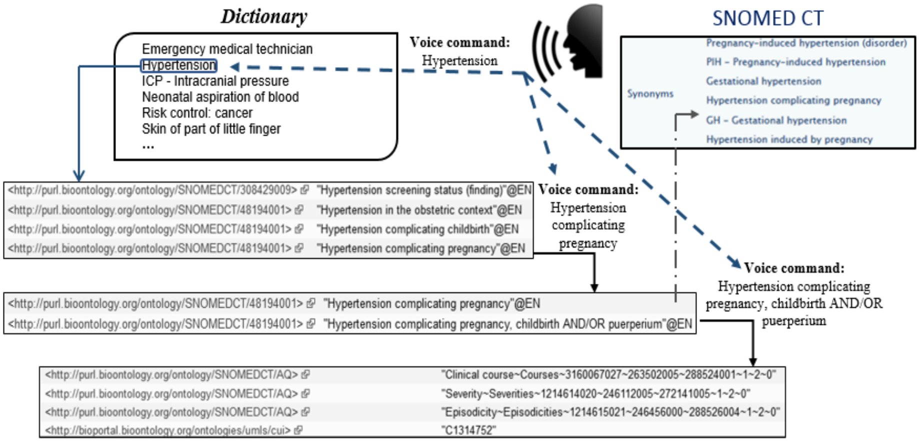


Fig 3. Navigational graph of the search through voice commands

The previous figure shows the navigation that is done through voice commands issued by the user and how they govern the navigation between the data contained in the RDF dataset, since the SPARQL queries are generated according to them. It is also shown that some of the results of the queries are associated with a series of synonyms, that is, terms that link to the same information.

The case study presented shows that it is possible to perform search and navigation in an RDF dataset, so that with the use of the architecture proposed in this work it is intended to develop a Web application that allows showing the results of the previous queries through facets and allows the user to navigate among them through the use of voice commands.

5 Conclusions and Future Work

The current semantic information retrieval mechanisms are intended to be used by expert users, on the other hand, Web applications usually integrate interfaces where normally the data capture is given through keyboard or touch screens, so the interaction is little natural and intuitive.

Based on the problems raised, we have proposed the creation of an architecture based on voice command recognition for faceted search in Linked Open Datasets, which benefits from NLP to reduce the gap between systems to navigate and explore semantic data available in the LOD cloud and users. The architecture is made up of four layers: Presentation, Integration, NL and Semantic. Each layer contains different components and the relationships between them give rise to the workflow. In the same way, a case study was presented for searching and faceted navigation in Linked Data that responds to the domain of the Life Sciences, through voice commands, said case study exemplified the usability of the proposed architecture.

As future work, we have contemplated to develop three modules; the first is a faceted navigation module in knowledge bases of the LOD cloud of medical domain and other domains through the recognition of NL, the second module will allow exporting the obtained results in the common formats including Spreadsheets, PDF, JSON and XML, and the third module will allows saving faceted navigation history of the user search to optimize the processes of search, faceted navigation and links discovery in the LOD cloud. With the development of these modules we pretend provide an alternative of solution to reduce the limitations of interaction with the systems that exploit the semantic data available in the LOD cloud. With the development of the proposed architecture the following benefits will be obtained: a) Obtaining information and systematically navigate large amounts of information that cannot be processed without automatic support; b) Unify information from triplestores linked through their properties (predicates), and c) Voice commands are used as the primary means of interaction between users and browsers with facets.

Some of the challenges expected through the development of the faceted browser are to provide reliable information to the user, because the sources of information are not managed by us, a study of Linked Open datasets will be done to determine which ones will be used to provide useful and reliable information. Likewise, it is contemplated that the developed browser be integrated to a third-party system that, like the present work, are still in development and belongs to the medical domain, therefore a challenge is to guarantee a correct integration.

Acknowledgments

Authors are grateful to the National Technological of Mexico for supporting this work. This research paper was also sponsored by the National Council of Science and Technology (CONACYT) and the Secretariat of Public Education (SEP) through PRODEP.

References

1. A. Andrejs and J. McCrae, “The Linking Open Data cloud diagram,” 2017. [Online]. Available: http://lod-cloud.net/.
2. L. F. Sikos, Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data. Apress, 2015.
3. J. Polowinski, “Widgets for faceted browsing,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5617 LNCS, no. PART 1, M. J. Smith and G. Salvendy, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 601–610.
4. V. Tablan, K. Bontcheva, I. Roberts, and H. Cunningham, “Mímir: An open-source semantic search framework for interactive information seeking and discovery,” Web Semant. Sci. Serv. Agents World Wide Web, vol. 30, pp. 52–68, 2015.
5. A. Revuelta-Martínez, L. Rodríguez, I. García-Varea, and F. Montero, “Multimodal interaction for information retrieval using natural language,” Comput. Stand. Interfaces, vol. 35, no. 5, pp. 428–441, 2013.
6. D. Griol, J. M. Molina, and Z. Callejas, “A proposal for the development of adaptive spoken interfaces to access the Web,” Neurocomputing, vol. 163, pp. 56–68, 2015.

7. L. Heck et al., “Multimodal Conversational Search and Browse,” IEEE Work. Speech, Lang. Audio Multimed., pp. 96–101, 2013.
8. M. A. Paredes-Valverde, R. Valencia-García, M. Á. Rodríguez-García, R. Colomo-Palacios, and G. Alor-Hernández, “A semantic-based approach for querying linked data using natural language,” J. Inf. Sci., vol. 42, no. 6, pp. 851–862, 2016.
9. I. Habernal and M. Konopík, “SWSNL: Semantic Web Search Using Natural Language,” Expert Syst. Appl., vol. 40, no. 9, pp. 3649–3664, 2013.
10. M. A. Paredes-Valverde, M. Á. Rodríguez-García, A. Ruiz-Martínez, R. Valencia-García, and G. Alor-Hernández, “ONLI: An ontology-based system for querying DBpedia using natural language paradigm,” Expert Syst. Appl., vol. 42, no. 12, pp. 5163–5176, 2015.
11. F. J. Serón and C. Bobed, “VOX system: a semantic embodied conversational agent exploiting linked data,” Multimed. Tools Appl., vol. 75, no. 1, pp. 381–404, 2016.
12. A. Ben Abacha and P. Zweigenbaum, “MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies,” Inf. Process. Manag., vol. 51, no. 5, pp. 570–594, 2015.
13. S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, “DBpedia: A Nucleus for a Web of Open Data,” in The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007. Proceedings, K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 722–735.
14. N. F. Noy et al., “BioPortal: ontologies and integrated data resources at the click of a mouse,” Nucleic Acids Res., vol. 37, no. suppl_2, pp. W170–W173, 2009.

Engineering Organizational Absorptive Capacity for Effective Knowledge Transfer

Orlando Lopez-Cruz[1] and Nini Johanna Garnica[2]

[1] Pontificia Universidad Javeriana, Carrera 7 42-27 p.7, Bogotá D.C. 110311, Colombia
[2] Nacional de Seguros, Calle 94 11-30 p.4, Bogotá D.C. 110221, Colombia
saudiel.lopez@javeriana.edu.co; nini.garnica@nacionaldeseguros.com.co

Abstract. The absorptive capacity (ACAP) construct has been defined as an organizational dynamic capability to identify, assimilate, transform and apply external knowledge in a productive way. While many researchers have focused in conceptual refinements of this construct, this research emphasizes on computer simulations to design actual ACAP in organizations. This paper suggests a lack of design of the ACAP conceptual models inferred from the reviewed literature and a new implementable design of the ACAP construct is proposed. An agent-based system implements the model of ACAP by completing previous proposals taking into account practical issues of knowledge flow both inside the organization and in its interaction to its environment. This new design sheds light upon an actual implementation of the ACAP construct able to support effective knowledge absorption at organizations.

Keywords: Absorptive capacity, knowledge management, engineering organizations, technological innovation, agent-based systems.

1 Introduction

Knowledge is a central concept in the knowledge-based view of the firm (KBV) [1-3] as well as in organizational learning theory [4, 5]. These approaches recognize knowledge as a critical asset to produce competitive advantage [6]. The KBV addresses the modes of knowledge in organizations whether codified or not (explicit or tacit) [7, 8]. KBV seeks to explain performance of the firm [2] on the basis of knowledge and, therefore, the way appropriable knowledge makes competitive advantage sustainable. Besides, KBV considers knowledge an intangible and difficult to imitate resource. Knowledge may be intra-organizational and extra-organizational transferred. In addition, KBV acknowledges individual shrewdness and abilities as hard to transfer (tacit) knowledge and keeps together operational knowledge and learning [9]. Accordingly, identification of the means of exploiting transferred knowledge in organizations when seeking competitive advantage and productivity becomes relevant: Innovation and strategic flexibility may conduct to competitive advantage in a dynamic market. Technology and technological knowledge provide the foundation for competitive advantage based in innovation [10].

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_17

This research introduces (i) the design of an ACAP construct able to describe both knowledge acquisition-exploitation, and (ii) results from an agent-based computer simulation of the ACAP model.

The remaining part of the document is structured as follows: previous studies on ACAP are analyzed, then the research methodology conducted and main findings. Finally, a proposed design for ACAP construct is introduced.

2 On the absorptive capacity (ACAP) construct

The ACAP construct [11] refers to an organizational capability [12-14] consisting of three dimensions: capacity to (i) identify or recognize, (ii) assimilate, and (iii) apply or exploit knowledge available in the organization environment. The ACAP construct was originally defined as an ability based on previous knowledge [11] to recognize the value of new information, assimilate it and apply it to commercial purposes. This proposal of ACAP is highly conceptual and, therefore, it is a loosely implementable model.

2.1 The birth and development of the ACAP construct

Literature [15] identifies three stages of the research on ACAP: (i) emergence of the ACAP construct (1989-2001): ACAP was developed in management science based on descriptive works [11, 16], (ii) conceptual foundation and establishment of a research domain stage (2002-2007), the ACAP construct is conceptually refined [17-20], extending research to organization theory, and (iii) stage of research domain consolidation (2007-now).

Results of the studies on ACAP may be classified as follows: identification of the components of ACAP [19, 21-23], the role of contextual conditions and other external factors regarding innovation in organizations [24], and, even, applications [25] [26], design of measurement instruments [27], experimental research [28] and their relationship to technology transfer [29].

2.2 ACAP and organizational routines

After twelve years from the seminal paper [11], ACAP was importantly re-conceptualized [20] as a set of organizational routines and processes to (i) acquire, (ii) assimilate, (iii) transform, and (iv) exploit knowledge to produce an organizational dynamic capability [12]. In other words, the ACAP construct is evoked as a dynamic capability that broadens organizational abilities to get and sustain competitive advantage [6] to generate organizational changes in a strategic manner.

When comparing this proposal [20] to the original construct [11], ACAP appears as a (i) dynamic capability, (ii) identifies components of ACAP as organizational routines, and describe their roles [30], and (iii) identifies conditions under which components of ACAP are able to create value. These allow to search for explanations on the differences between two organizations operating in the same economical sector, and, the most important issue, to prescribe those differences. However, the absence of details in the re-conceptualization prevents any implementation (operationalization) of the construct.

A third re-conceptualization of ACAP [19] returns to the original idea of ACAP consisting of (i) identify or recognize knowledge value, then proposes two parallel processes (ii) acquisition and (iii) assimilation before completing (iv) knowledge transformation in order to (v) exploit (apply or benefit from) knowledge. In addition, this re-conceptualization includes a feedback loop making coherent the structure and dynamics of ACAP as an organizational capability, that may not be easily captured by analytic (linear) methods.

2.3 ACAP and organization capabilities

An organizational capability is a high level routine that, by means of organizational specific processes and resources, enable senior management a set of options to make decisions to produce meaningful results of a particular kind. Usually, these results are strategic goals that use available know-how (knowledge) and non-specific organizational resources [31, 32].

Organizational capabilities may be classified as either operational or dynamic capabilities. An operational capability is a set of competences and abilities determining the effectiveness of an organization to perform missionary daily activities. In contrast, a dynamic capability is the ability to integrate, build, and reconfigure operational capabilities [14], as well as internal and external competences to address fast changes in the environment [33]. Dynamic capabilities addresses organizational change. Therefore, they allow to look for a proper-contextualized answer to fluctuating environments, and an implementation of a course of action of the organization revealing evolutionary features of organizations.

Accordingly, the components of ACAP –organizational routines [30] – are grouped into two sets: potential absorptive capacity and realized absorptive capacity. Potential ACAP organizational routines are knowledge acquisition and knowledge assimilation. Realized ACAP organizational routines are transformation and exploitation. The aim of routines of potential ACAP (PACAP) is to foster organizational receptiveness to external knowledge, which is loosely equivalent to the dimension “recognize knowledge value” [11]. Routines of realized ACAP (RACAP) are aimed to leverage organizational learning based on absorbed knowledge.

Researchers of the first decade of the 21st century suggest that these organizational routines provide systemic, structural, and procedural mechanisms allowing organizations to exploit knowledge by chance and, even so, get long range benefits, despite barriers such as “appropriability” (i.e. institutional dynamics, such as intellectual property, affecting organizational capabilities to protect their rights to benefit from new products or processes). Knowledge appropriability is in inverse relation to the organizational propensity for investment in research and development (R&D) [18].

However, this model [20] exhibits ambiguities [19], for instance the omission for routines that in the original proposal are just components of the construct “recognition of the value of external knowledge". Literature warns [19] that linear analytical models [11, 20] do not describe properly the complexity of the causal relationships of ACAP that explain dynamics of ACAP through organizational learning processes: exploratory, transformative and exploitation [18].

A new model of ACAP [19] (i) resumes the component of knowledge value recognition [11] including it as one of its components, (ii) extends the scope of incidence (either positive or negative) of the social mechanism of integration [17] to all the organizational routines by theorizing contingency factors, (iii) includes into the set of contingency factors power relationships influencing both knowledge assessment routines and knowledge exploitation routines, (iv) recognizes complex relationships between the dimensions of assimilation and transformation, (v) redefines transformation routines, and (vi) includes feedback loops characterizing dynamics and complexity of the phenomenon.

From this point of view, an organization may be conceived as a knowledge repository. An organization may be modeled as an instrument to transfer and develop knowledge not just next to other related organizations (i.e. not just create or apply knowledge, which is an static view), but exposed to a knowledge flow. Hence ACAP seems to be the best suited construct to explain the way an organization may be able to actually acquire knowledge and make it productive. This is important because prior studies indicate that knowledge transfer depends on the level of ACAP as well as the complexity of the knowledge being transferred and social aspects in turbulent environments [9].

A structure that targets knowledge management [34] deploys in four organizational learning processes: (1) knowledge acquisition, (2) information distribution, (3) information interpretation, and (4) organizational memory. This process structure resembles and complements the absorptive capacity (ACAP) construct.

3 Methodology

To accomplish the aim of this research, the ODD protocol [35] was used to separate modeling from implementation details and computer programming languages [36]. In this way a conceptual model of an agent-based system was completed. In that model agents are organizations, knowledge 'flow' represented by 'technological artefacts' as well as 'people' external to the organizations, and organizational barriers to access knowledge. These describe the situational domain and the model specifications after the phases (I) Overview, (II) Design concepts and (III) Details.

Knowledge is understood and represented as organizational routines. Each organization in the model has its own set of organizational routines. An organization has full access to 'technological artefacts' and 'people' routines in its neighborhood (they have the capability to identify knowledge in the knowledge flow) in order to decide either to "capture" the knowledge in the external flow or not. In addition, an organization has a 'knowledge base' which is a list of tasks to perform their activities. Knowledge decay in data and, therefore, disappears from the model.

Verification of the model included code verification in terms of (i) error detection and debugging in the computer program, and (ii) evaluation that the implemented program corresponds to its specifications ion ten conceptual model. This leads to evaluate if the conceptual model has been implemented in the programming language. This was done before the validation process [37]. Validation of agent-based systems is a difficult task when tackled in the same way as scientific traditional models [38-40]. In this research, validation was accomplished by subject experts [41] of health and insurance organizations.

4 Configuration of the Organizational Absorptive Capacity

A previous study [42] identifies knowledge as the main element involved in organization dynamics, confirming that intangible elements are the main elements to generate organization identity and sustainable competitive advantage [43]. In order to take advantage of this and avoid the flaws of previous models, we proposed a re-design of the ACAP construct. While some ACAP models [11, 19] lack of an explicit specification of the interrelationship of the organization and its environment, the design introduced in this research paper considers environmental elements as determinants of the components of the construct. Furthermore, the aim of this new model is not to describe but to prescribe the way organizations may implement ACAP. This is done to provide organizations with assistance to design their own means to identify interactions between them (organizations) and the environment where knowledge is taken from.

In this re-design, several organizational routines were identified as the elements causing knowledge to flow and be seized or not seized. From this on, organizational routines are organized in classes or categories: Identification and recognition, acquisition, assimilation, adoption/adaptation, acceptance, transformation, exploitation and innovation. Finally, they were assembled in a unique artefact by means of both internal and external (environmental) knowledge flows (Fig. 1).

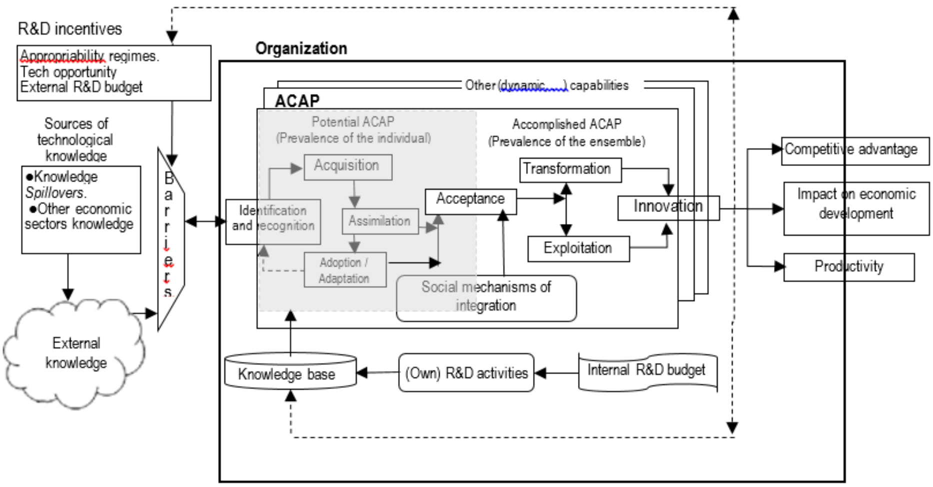


Fig. 1. A re-design of the construct absorptive capacity ACAP.

4.1 Information, knowledge and ACAP

Knowledge is different from information. Knowledge can be obtained by means of cognition and interpretation processes, only [44]. Learning organizations acquires knowledge by means of adaptive learning – assuming environmental changes but making minor changes –or not at all- in mental models (paradigms)- and via generative learning –which is creative and produces changes in the organizational "mental" model [45]. Knowledge in organizations is verified in relation to structures and organizational routines, since it is not derived from information available to any particular member of the organization but emerges as a property of the learning system materialized in people (i.e. human capital) [46]. Besides, it is shaped by the interaction between different learning processes that constitute the organization [47].

Tacit knowledge is embodied partially in habits and organizational routines. This makes tacit knowledge irreducible to information [44] and not able to be transmitted in a codified form. On the other hand, economic agents depend on tacit knowledge [48] and, in turn, all the abilities and decisions depend on tacit knowledge. Hence the importance of having a construct that allows a more real or concrete implementation.

The proposed design of the construct (Fig. 1), acknowledges ACAP as a complex construct [49]. ACAP is integrated by different, mutually interacting components which in turn allows the organization interact with its environment. Its components are routines (potential and accomplished ACAP) to identify and recognize knowledge value from organizational environment, acquisition, assimilation, adoption/adaptation, acceptance, transformation, exploitation and innovation (Fig. 1).

In this model, ACAP is recognized as a dynamic capability (one of the many organizational dynamic capabilities). This design resumes the distinction between potential ACAP (Potential ACAP or PACAP) to represent the domain where individual action prevails over collective action, and achieved ACAP (Achieved ACAP or AACAP) where collective action and knowledge prevails over individual ones.

The role of interactions with the environment (Fig. 1) - some of them feedback as a result of knowledge transformation and exploitation- is to capture the real alteration of appropriability regimes, because of knowledge exploitation, by means of patenting or copyrights on knowledge that has been internally created. Besides, this design captures the internal feedback to the knowledge base which is altered by own R&D budget (internal budget) and it is altered by internal R&D activities, if exist.

The main contribution of the design of this construct, that herein is being introduced, is the operationalization feasibility when implementing (and to implement) each set of organizational routines. Furthermore, this design includes several feedback loops (dashed lines from right to left in Fig. 1), representing different knowledge elements altering ACAP in itself (as the directed axis crossing from adoption/adaptation and identification and recognition) or inside the organization border (as the directed axis from innovation to knowledge base) or the organization and its environment (as the directed axis running from innovation to incentives for R&D). To the best of our knowledge, no other model is closer to represent dynamic nature plus inherent feedback loops of ACAP. This may ease computer simulation models development.

4.2 ACAP implementation

The ACAP computational model was implemented in NetlLogo 6.0.1 [50] on a Windows 10® microcomputer operating system. An organizational routine (OR) was implemented as an ordered list of tasks 1 thru n, $n \in N$ [51]. An OR consists of an ordered k-uple of tasks ($1 \leq k \leq n$). Let t_i ($1 \leq i < j < k < l \leq m$; i, $m \in \mathbf{N}$) a minimal unit of actions. The list of tasks [t_i, t_j, t_k, t_l] is an organizational routine integrated by four tasks: t_i, t_j, t_k, t_l,; i< j< k< l. In general, each agent in the knowledge flow exhibits an organizational routine (a list) [t_1, t_2,… t_n] where *n* is the size of the list.

The organization agent (blue shuttle) is commanded by the user using the left-right buttons to guide the shuttle thru the flow of persons, artifacts (and barriers) (Fig. 2).

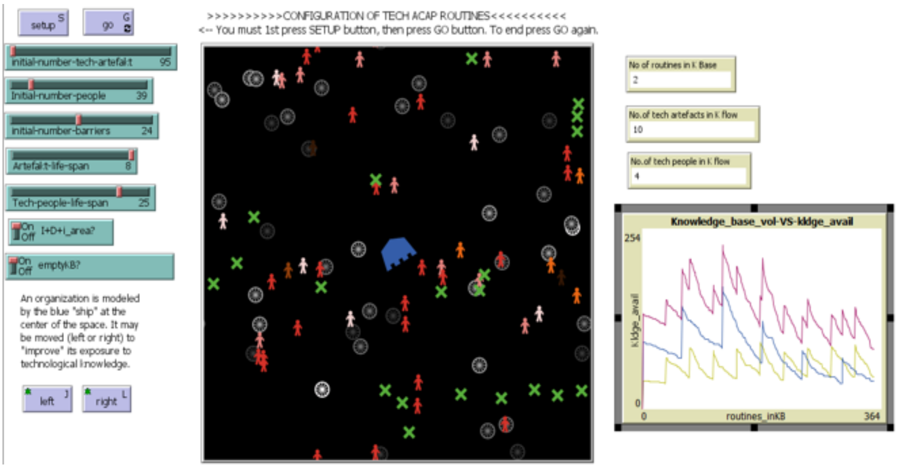


Fig. 2. A screenshot of the computer simulation of the ACAP re-design.

5 Results and analysis

This ACAP implementation allowed experimentation with different configurations of the ACAP of an organization. Each execution may conclude in different results, but the relevance was in different behaviors as a result of a change one or more initial parameters (sliders) controlling knowledge flow and barriers. In addition, lifespan make sense to users to represent different business cycles of different ICT and medical products. In addition, lifespan allowed to represent 'knowledge' obsolescence in organizations. This is usually the case when, for instance, a person knows how to use computer command-line operating systems as UNIX, iRMX, CP/M or MS-DOS and then has to change to GUI operating systems.

When users concentrated in acquiring technological artefacts based knowledge, under some circumstances the number of people lowered in the flow of knowledge. In some runs people in the knowledge flow disappeared. Experts believed this may be a result of the obsolescence of people (knowledge people holds) and there is not an 'update' of knowledge inside organizations. Because this model does not represent organizations such as universities and research centers experts considered this a plausible result. A symmetrical result happens when the user focused in gathering people. This represented an extreme context where there is no technological industry or international commerce. This generates a technological lag for the organization and its neighborhood or contextual economic sector.

This results allows the representation of technological lag but, also, the absence of (organizational or national) research and development policies that tend towards to create incentives to the creation of new products (or services) able to be incorporated to the knowledge flow. This is another way to say: innovation.

According to the (expert) users knowledge lifespan and the balanced focusing in acquiring 'technological artefacts' and (trained) 'people' does not guarantee any success of the organization in its economic context, which seemed to be the behavior of the computational model in some executions. In addition, an excess in knowledge barriers may degrade a knowledge flow at an early stage.

6 Conclusion

Addressing the ACAP construct from the knowledge-based view (KBV) has shown the evolution of the construct from a conceptual dynamic capability of organizations deployed in three dimensions: (i) identify or recognize, (ii) assimilate, and (iii) apply or exploit knowledge available in the environment (Cohen & Levinthal, 1990), to a construct that is able to describe and operationalize available knowledge (Fig. 1) in order to create competitive advantage and improve organizational productivity.

There are two contributions of this research. The re-design of the ACAP construct in terms of components with (internal and external) feedback representing knowledge as organizational routines, and the implementation of the model in an agent-based computer system. The later evidences that the re-design of the ACAP proposed in this research is actually implementable not just as a reference for statistical calculations but as a usable artefact of people in organizations. Also, in contrast to previous models aiming to describe what happens to ACAP, the benefit of this model (Fig. 1) is the capability to actual prescribe components in the construct.

7 Further research

There are many opportunities to continue the development of this results. At first, an assessment of the completeness of the ACAP design (fig. 1) may be conducted to determine if this construct represents all organizational routines that innovative organizations perform.

In addition, further computational development may be conducted to enrich the model including the role of universities and research centers to ‘reproduce’ incorporated knowledge in people and technological artefacts that eventually will flow as a part of the knowledge flow where organizations navigate.

The implementation of organizational routines may be widen to include other conceptual categories of organizational routines which understand them as (i) set of rules or standard procedures and (ii) as trends that link to previously acquired and adopted behaviors that may be triggered by the proper stimuli and contexts.

7 Acknowledgments

We want to pay special thanks to people at Instituto Geofísico, Pontificia Universidad Javeriana (Bogotá D.C.) and to Carlos I. Delgado at RadioXenter Ltd. who validated the simulation model. We are also grateful to Nelson Obregón Ph.D., Pontificia Universidad Javeriana, for his support. Finally, we want to acknowledge the blind reviewers for his comments and suggestions to improve this paper.

REFERENCES

1. Eisenhardt, K.M. and F.M. Santos, *Knowledge-based view: A new theory of strategy.* Handbook of strategy and management, 2002. **1**: p. 139-164.
2. Grant, R.M., *Toward a knowledge- based theory of the firm.* Strategic management journal, 1996. **17**(S2): p. 109-122.
3. Nickerson, J.A. and T.R. Zenger, *A knowledge-based theory of the firm—The problem-solving perspective.* Organization science, 2004. **15**(6): p. 617-632.
4. Argyris, C. and D.A. Schön, *Organizational learning: A theory of action perspective*. Vol. 173. 1978: Addison-Wesley Reading, MA.
5. Levitt, B. and J.G. March, *Organizational learning.* Annual review of sociology, 1988: p. 319-340.
6. Porter, M.E., *Competitive strategy : techniques for analyzing industries and competitors : with a new introduction*. 1998, New York: Free Press.Traducción a castellano- Estrategia Competitiva : Técnicas Para El análisis De Los Sectores Industriales y De La Competencia. México D.F.: Compañía Editorial Continental, 1998.
7. Nonaka, I. and H. Takeuchi, *The knowledge-creating company : how Japanese companies create the dynamics of innovation*. 1995, New York: Oxford University Press.
8. Nonaka, I., R. Toyama, and A. Nagata, *A firm as a knowledge-creating entity: a new perspective on the theory of the firm.* Industrial and corporate change, 2000. **9**(1): p. 1-20.
9. Dhanaraj, C., et al., *Managing tacit and explicit knowledge transfer in IJVs: the role of relational embeddedness and the impact on performance.* Journal of International Business Studies, 2004. **35**(5): p. 428-442.

10. Teece, D.J., *Profit in form of technological innovation.* Research Policy, 1986. **15**(6).
11. Cohen, W.M. and D.A. Levinthal, *Absorptive capacity: a new perspective on learning and innovation.* Administrative science quarterly, 1990: p. 128-152.
12. Helfat, C.E., et al., *Dynamic capabilities: Understanding strategic change in organizations*. 2009: John Wiley & Sons.
13. Teece, D. and G. Pisano, *The dynamic capabilities of firms: an introduction.* Industrial and corporate change, 1994. **3**(3): p. 537-556.
14. Teece, D.J., G. Pisano, and A. Shuen, *Dynamic capabilities and strategic management.* Strategic management journal, 1997. **18**(7): p. 509-533.
15. Leydesdorff, L.A. *The Triple Helix, Quadruple Helix,…, and an N-tuple of Helices: Explanatory Models for Analyzing the Knowledge-based Economy?* 2011.
16. Cohen, W.M. and D.A. Levinthal, *Fortune favors the prepared firm.* Management science, 1994. **40**(2): p. 227-251.
17. Easterby-Smith, M., et al. *Absorptive capacity in practice: an empirical examination of Zahra and George's model*. 2005. Citeseer.
18. Lane, P.J., B.R. Koka, and S. Pathak, *The reification of absorptive capacity: A critical review and rejuvenation of the construct.* Academy of management review, 2006. **31**(4): p. 833-863.
19. Todorova, G. and B. Durisin, *Absorptive capacity: valuing a reconceptualization.* Academy of management review, 2007. **32**(3): p. 774-786.
20. Zahra, S.A. and G. George, *Absorptive capacity: A review, reconceptualization, and extension.* Academy of management review, 2002. **27**(2): p. 185-203.
21. González, I., *Modelo teórico de capacidad de absorción, innovación organizacional y emprendimiento.* Informe de Investigaciones Educativas, 2014. **27**(1): p. 290-340.
22. Nishiyama, T., K. Ikeda, and T. Niwa, *Technology Transfer Macro-Process: A Practical Guide for the Effective Introduction of Technology*, in *ICSE 2000*. 2000.
23. Vega, J., A. Gutiérrez, and I. Fernández de Lucio. *An analytical model of absorptive capacity*. in *DRUID Summer Conference*. 2007.
24. Fabrizio, K.R., *Absorptive capacity and the search for innovation.* Research Policy, 2009. **38**(2): p. 255-267.
25. López-Paniagua, I., et al., *Clases prácticas: Una herramienta esencial en la enseñanza de las ingenierías en el marco del Espacio Europeo de Educación Superior.* Dyna, 2011. **86**(5): p. 523-530.
26. Volberda, H.W., N.J. Foss, and M.A. Lyles, *Absorbing the Concept of Absorptive Capacity: How To Realize Its Potential in the Organization Field.* 2009.
27. Marín, A., D. Laureiro, and C. Forero, *Innovation patterns and intellectual property in SMEs of a developing country.* Galeras de Administración, 2007. **17**.

28. Filgueiras Sainz de Rozas, M.L., M. Castro Fernández, and I. Rafull Suárez, *Determinación de la capacidad de absorción: estudio de caos en la empresa GEYSEL.* Ingeniería Energética, 2013. **XXXIV**(3): p. 175-185.
29. López-Cruz, O., *Transferencia de tecnología informática: entorno colombiano.* Revista de Tecnología - Journal of Technology, 2010. **9**(1): p. 19-23.
30. Bobrow, D.G. and T. Winograd, *An overview of KRL, a knowledge representation language.* Cognitive Science, 1977. **1**(1): p. 3-45.
31. Teece, D.J., *Technological transfer by mutlinational firms: the resource cost of transferring know-how.* Economic Journal, 1977. **87**(June): p. 242-261.
32. Winter, S.G., *The satisficing principle in capability learning.* Strategic Management Journal, 2000. **21**(10-11): p. 981-996.
33. Zollo, M. and S.G. Winter, *Deliberate learning and the evolution of dynamic capabilities.* Organization science, 2002. **13**(3): p. 339-351.
34. Huber, G.P., *Organizational learning: The contributing processes and the literatures.* Organization science, 1991. **2**(1): p. 88-115.
35. Grimm, V., et al., *A standard protocol for describing individual-based and agent-based models.* Ecological modelling, 2006. **198**(1): p. 115-126.
36. Grimm, V., G. Polhill, and J. Touza, *Documenting social simulation models: the ODD protocol as a standard*, in *Simulating Social Complexity*. 2013, Springer. p. 117-133.
37. Sankararaman, S. and S. Mahadevan, *Integration of model verification, validation, and calibration for uncertainty quantification in engineering systems.* Reliability Engineering & System Safety, 2015. **138**: p. 194-209.
38. Klügl, F. *A validation methodology for agent-based simulations*. in *Proceedings of the 2008 ACM symposium on Applied computing*. 2008. ACM.
39. Parunak, H.V.D., R. Savit, and R.L. Riolo. *Agent-based modeling vs. equation-based modeling: A case study and users' guide*. in *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. 1998. Springer.
40. Xiang, X., et al. *Verification and validation of agent-based scientific simulation models*. in *Agent-Directed Simulation Conference*. 2005.
41. Niazi, M.A., A. Hussain, and M. Kolberg. *Verification and validation of agent based simulations using the VOMAS (virtual overlay multi-agent system) approach*. in *MAS&S at Multi-Agent Logics, Languages, and Organisations Federated Workshops (MALLOW)*. 2009. CEUR-WS.
42. Lopez-Cruz, O. and N. Obregon Neira, *Diseño de la capacidad de absorción en las organizaciones: propuesta de un nuevo constructo y literatura*, in *Congreso Nacional e Internacional en Innovación en la Gestión de Organizaciones, Abril, 2016*, F.R. Santoyo, Editor. 2016, Universidad Central: Bogotá. p. 222-237.
43. Calix, C.G., H.P. Vigier, and A.E. Briozzo, *Capital intelectual y otros determinantes de la ventaja competitiva en empresas exportadoras de la zona norte de Honduras.* Suma de Negocios, 2015. **6**(14): p. 130-137.

44. Tecuci, G., *Cooperation in knowledge base refinement*, in *Machine Learning: Proceedings of the Ninth International Conference ML92*, D. Sleeman and P. Edwards, Editors. 1992, Morgan Kauffmann. p. 445-450.
45. Senge, P., *The Fifth Discipline: The Art & Practice of The Learning Organization*. 1990, New York: Double Day Dell Publisher.
46. Rodríguez, A.S.D., R.C. Aguirre, and L.S. Ordóñez, *Clima organizacional, migración tecnológica y apagón analógico en la televisión ecuatoriana. .* Suma de Negocios, 2015. **6**(14): p. 138-146.
47. Dosi, G. and L. Marengo, *Toward an evolutionary theory of organizational competencies*, in *Evolutionary concepts in contemporary economics*, R.W. England, Editor. 1993, University of Michigan Press: Ann Arbor.
48. Bareiss, E.R., B.W. Porter, and K.S. Murray, *Supporting strat-to-finish development of knowledge bases.* Machine Learning, 1989. **4**: p. 259-283.
49. Morin, E., *Introducción al pensamiento complejo*. 2011, Barcelona (España: Gedisa.
50. Wilensky, U., *NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling*. 1999, Northwestern University: Evanston, IL.
51. Miller, K.D., B.T. Pentland, and S. Choi, *Dynamics of performing and remembering organizational routines.* Journal of Management Studies, 2012. **49**(8): p. 1536-1558.

Decision-Support Platform for Industrial Recipe Management

Edrisi Muñoz[1], Elisabet Capón-García[2], Mirna Muñoz[1] and Patricia Montoya[3]

[1] Centro de Investigación en Matemáticas A.C., Jalisco S/N, Mineral y Valenciana 36240, Guanajuato, México
[2] Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 1, Zurich 8049, Switzerland
[3] Instituto Tecnológico de Zacatecas, Carretera Panamericana S/N crucero a Guadalajara, Zacatecas, Zac.
{edrisi.munoz, mirna.munoz@cimat.mx, elisabet.capon@chem.ethz.ch, ilosth3aven@gmail.com}

Abstract. The effective and timely use of enterprise wide optimization models requires robust and reliable data acquisition systems to extract model relevant parameters and to drive the proposed enterprise wide coordination strategies. This work aims to create a knowledge-based platform for systematic and standardized management of the general, site and master recipes within the process industry. The platform allows the creation of a master recipe ready for the production planning and scheduling, as well as for the process management. As a result, recipe management functionalities are supported by a traceable and reliable system.

Keywords: Decision-support system, recipe management, enterprise processes, knowledge management.

1 Introduction

In the 21st century, industry faces new challenges, many of which are concerns shared by society as a whole: the impact of climate change, aging population, availability of safe drinking water, food scarcity and cost, and the security of energy supplies. With these challenges come tremendous business opportunities for the chemical industry, which is uniquely placed to help in developing solutions through the creation of products that improve the quality of life, health, productivity and safety. For many years, companies have developed management information systems to support the end users to exploit data and models, with the final objective of decision-making. Nowadays, global competition has made some of these decisions, which are related to certain manufacturing characteristics, such as economic efficiency, product quality, flexibility or reliability, essential for the viability of the enterprise [1]. Specifically, an increasing number of efficient solution technologies are currently available to improve the profitability of industrial production. However, one main challenge consists of establishing efficient information sharing within

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*, Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_18

industrial production management systems in order to take advantage of such solution technologies. Indeed, a widely accepted and simple procedure of general applicability for direct process data exchange is still missing. In this sense, production recipes represent a structured piece of information, which can be shared at all decision levels. Production recipes contain the information to be extracted for process documentation, fast product design, development and integration in batch plants. Nevertheless, it is difficult for any new solution to gain wide acceptance within industry as companies already use their established interface. From an acceptance point of view, the ISA-88 standard could act as basis for a neutral data-exchange platform, thus avoiding costly custom-tailored software interfaces.
The objective of this work is to create the recipe management functions in order to provide a decision support system for enterprise recipe management. Therefore, this work relies on ANSI/ISA-88 standard for automation and control [8], which provides a model encompassing the entire scope of manufacturing activities for the integration of plant information.

2 State-of-the-art

The concept of decision support system, DSS, was introduced from a theoretical point of view, in the late 1960s. [3] defines a DSS as a computer information system that provides information in a specific problem domain using analytical decision models as well as techniques and access to databases, in order to support decision makers in making decisions effectively in complex and ill-structured problems. Thus, the basic goal of a DSS is to provide the necessary information to the decision maker, in order to help him to get a better understanding of the decision environment and the faced alternatives.

Decision-making has significantly improved in the last two decades through the rapid progress of information technology and computer science. Typically, the phases of the decision-making process overlap and blend together, with frequent looping back to earlier stages as more is learned about the problem, as solutions fail, and so forth. Figure 1 describes what probably comes to be a more customarily used model of the decision-making process in a DSS environment. A first step is the recognition of the problem or opportunity. After the problem is recognized, it is defined as a term that facilitates the creation of the model. Some authors state that emphasis must be placed in the next two steps: the model development and the alternatives analysis. After that, the choice is made and implemented. As final step and if necessary, a new recognition should be done. Obviously, no decision process is this clear-cut in an ill-structured situation [4].

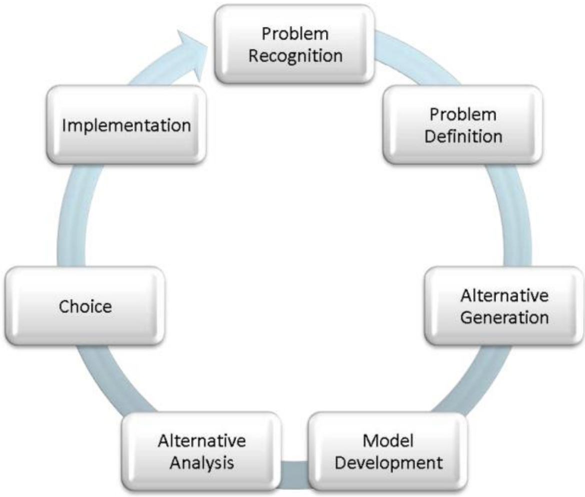


Fig. 1 Decision making process cycle.

There has also been a huge effort in the DSS field for building a Group Support System (GSS) or Collaboration Support System in order to enhance the communication-related activities of team members engaged to computer-supported cooperative work. The communication and coordination activities of team members are facilitated by technologies that can be characterized along the three continua of time, space, and level of group support [4][5]. Teams can communicate synchronously or asynchronously; they may be located together or remotely; and the technology can provide task support primarily for the individual team member or for the group's activities. These technologies are utilized to overcome space and time constraints that burden face to face meetings, to increase the range and depth of information access, and to improve group task performance effectiveness, especially by overcoming "process losses".

The combination of decision theory with the new knowledge and the powerful tools offered by computer science and information technology, leads to the development of new types of information systems able to support decision makers and improve the decision-making process. Basically, the efforts in supporting the whole decision-making process focused on the development of computer information systems providing the support needed. Initially, two types of systems were developed: i) decision support systems (DSS), and ii) expert systems (ES).

Although these two approaches were very promising, their implementation revealed several problems. To overcome those problems without missing the advantages of both ESs and DSSs, a new type of intelligent system called a knowledge-based DSS (KBDSS) has been proposed. The basic characteristic of this new approach is the integration of ES technology with models and methods used in the decision support framework, such as mathematical programming methods, multicriteria decision aid methods, and multivariate statistical methods.

In the 21st century, World Wide Web and telecommunications technologies can be expected to result in organizational environments that will be increasingly more global, complex, and connected. [6], following [7], suggests that DSS researchers should embrace a much more comprehensive view of organizational decision-making (see Figure 2) and develop decision support systems capable of handling much "softer" information and much broader concerns than the mathematical models and knowledge-based systems have been capable of handling in the past. The primary difference between Figure 2 and typical decision models in a DSS context is the development of multiple and varied perspectives during the problem formulation phase. [7] suggests that perspectives be developed from organizational (O), personal (P) and technical (T) positions. In addition, ethical and aesthetic factors are considered as well. The mental models of stakeholders with various perspectives lie at the heart of the decision process, from defining what a problem is, to the analysis of the results of the problem.

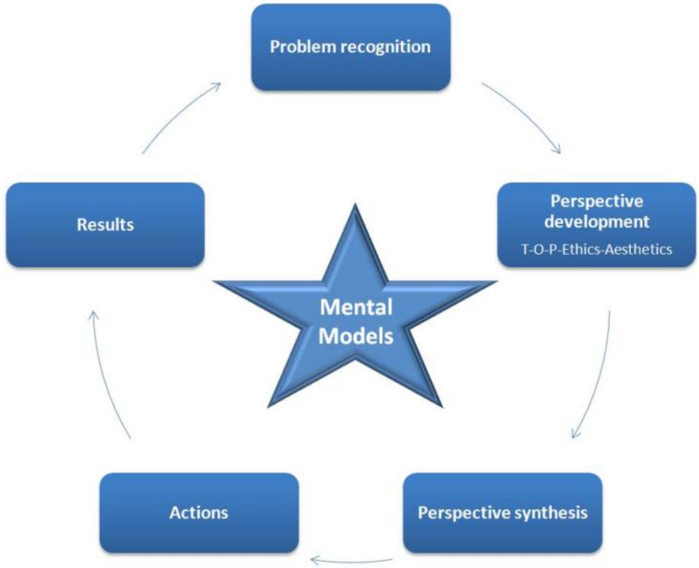


Fig. 2 A new decision paradigm for decision support systems [6].

3 Methodology

As presented in the previous section, Decision Support Systems (DSS) are information technology solutions that can be used to support complex decision-making and problem solving. DSS are defined as "aid computer systems at the management company level that combine data and sophisticated analytic models for supporting decision-making" [8]. Classic DSS design comprises components for (i) sophisticated database management capabilities with access to internal and external data, information, and knowledge; (ii) modeling functions accessed by a model management system; (iii) simple user interface designs that enable interactive queries, reporting, and graphing functions; and, (iv) optimization by mathematic algorithms and or intuition/knowledge. Much research and practical design effort has been conducted in each of these domains [4]. This work combines the elements of classical DSS and extends them a step forward towards KBDSS using a knowledge based model that supports communication and the formalization of the domain through semantic relationships.

As for recipe management, it can be generally divided in four stages: development, application, storage and share. In this context, a software platform, which embeds the activities of each stage, has been developed as described next.

Development: it refers to the generation of the different types of recipes. Specifically, the recipe management function as described in ANSI/ISA 88 standard contains the elements for managing the three upper levels of recipes, namely general, site and master recipes, and defines the procedural elements used in the recipe procedures. Thus, different elements for recipe creation, such as formula, header, and procedural logic, among others, are guided by the platform.

Application: it refers to the correct use of recipes for their implementation in a specific plant. Thus, the format for the information integration of the recipe within plant system is improved and effectively guided by the platform.

Storage: it refers to the correct creation of recipe database, allowing a reliable and intelligent manner for control, storage and validate effectively guided by the platform.

Share: it refers to the recipe sharing among different enterprises, facilities or areas guided by the platform.

The platform development process follows the design pattern MVC (model-view-controller) [9]. This pattern defines roles of the MVC-objects as well as the interactions among them. In this work the "model object" is represented by the ontological model [10], previously defined, which is codified in web ontology language (OWL) format. Next, "view object" regards to the graphical user interface, which mainly interacts with the final user. Finally, in "controller object" the main decision functions of the platform are programmed in Java and Jython languages due to their capacity and the flexibility, respectively.

On the other hand, the platform development is guided by software prototyping model. Besides, the model is reinforced with the addition of software requirement identification and management based on best practices provided by ISO 29110 [11], PMBOK [12], CMMI [13] and TSPi [14]. Those additional practices allow the methodology having a better control and monitoring activities for any possible project change along its development cycle. The complete development method comprises seven steps (Figure 3), which are described next.

1. Identify requirements. In this step, requirements are identified, as well as, information gathering for domain understanding based on the ontological model.
2. Identify potential technology. A systematic revision related to tools for ontological model management is done. Next, a selection and configuration of tools must be done, for their utilization in the cycle development.
3. Develop initial prototype. In this step, the prototype functions are defined based on the previously requirement's identification.

4. Codify prototype. In this step, the functions are programming following the prototype design.
5. Test. In this phase, the prototype must be test by the development team, as well as, the customer. Thus, if an improvement is detected it should be documented as change request to be analyzed.
6. Review & improve of prototype. In this step, the improvements previously defined are analyzed in order to making the corresponding changes in the prototype.
7. Final platform. Finally, when the prototype fulfills the customer's performance expectations, the final product is achieved.

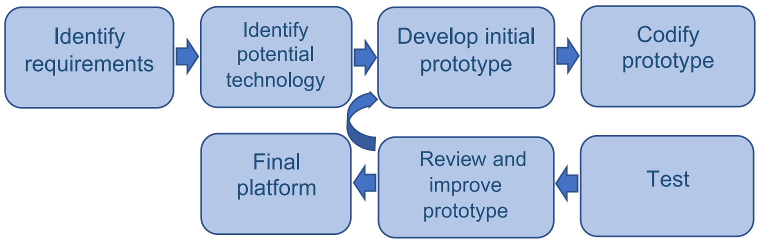


Fig 3. Flow diagram of decision support platform development steps

4 Results

The decision support platform allows a more efficient and user-friendly use of the implementation and management of the recipes within an ANSI/ISA-88 standard environment. In this sense, [10] created a knowledge-based model formalizing the physical, process, procedural and recipe models described in the standard, allowing the exploitation of data end information of the recipe model approach. As a result, the user interaction with process data is improved; data protocols (from ANSI/ISA-88) and the recipe management methods are standardized. Additionally, information can be easily stored in intelligent databases relying on ontological structures. In order to illustrate the developed platform, a case study demonstrating how the functionalities are transformed in the user interface is presented. In particular, the case study tackles the modification, archiving and maintenance of the recipe procedural elements. On the one hand, Figure 4 presents the steps in the algorithm for element modification and element maintenance. In both cases, the name and description of the recipe procedural element (building block) is provided and a semantic based search is conducted in the repository of recipe building blocks. After selecting the desired building block, either the status of the version is updated (for the maintenance case) or changes are performed in a duplicated building block (for the modification case). In both cases, the modified version is saved and the repository of building blocks is updated.

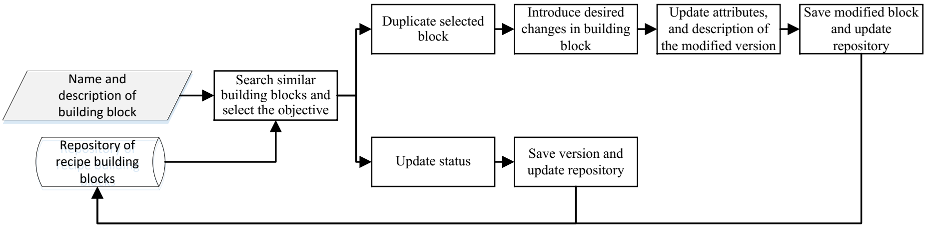


Fig 4. General steps for modification (upper path) and maintenance (lower path) of recipe procedural elements according to ANSI/ISA-88.

On the other hand, Figure 5 presents the prototype of the graphical user interface for the equipment requirement, which is a part of the recipe procedural element. The user interface relies on the theoretical scheme presented in Figure 4, which improves user recipe management tasks. Indeed, the proposed platform semantically defines and contains data and information allowing semantic-based searches automating recipe data requirements response. Figure 5 shows four main blocks, namely a, b, c and d, structuring the user interface in the decision support platform. Thus, 5a shows a common action buttons and commands for edition, tools, view, etc. 5b shows the domain class tree that guides the user through recipe development. 5c shows data and information resume regarding a specific class or instance. Finally, 5d presents a digitalized version of an automatically developed structured recipe.

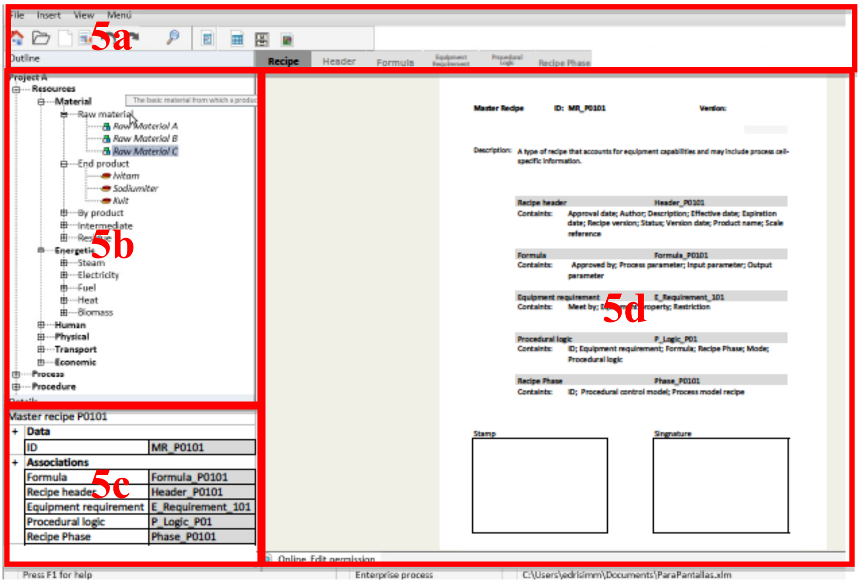


Fig 5. Screenshot of the user interface created for decision support.

5 Conclusions

This work presents a decision support platform for conducting recipe management activities relying on the ANSI/ISA-88 standard guidelines. The resulting framework allows a systematic management of the general, site and master recipes as well as the associated recipe procedural elements. Several algorithms are proposed, and semantic based searches are programmed in Jython in order to automate this process and help the process engineer to reach adequate recipes for the process management activities. Besides, the platform is a guideline for the automation of recipe documentation creation, storage, control and share improving the recipe management process activities within industry. Overall, this framework stands for a traceable and reliable system for recipe management, and represents a step towards including artificial intelligence based tools at the plant level.

As part of future work, data and information structures developed in this work stand for the basis for the automation of standard operating procedures (SOPs) and development of good manufacturing practices (GMPs) at process industry. On the one hand, in the batch process industry, there are written instructions (recipes) that document the way the unit operations are performed in the production plants; namely, they define the SOPs. Besides, at food and drug process industry GMPs describe the methods, equipment, facilities, and controls for producing processed, as the minimum sanitary and processing requirements for producing safe and wholesome food and drug. Thus, an integrated and smart management environment is pursued for improving decision-making tasks and the overall performance in process industries.

References

1. Venkatasubramanian, V., C. Zhao, G. Joglekar, A. Jain, L. Hailemariam, P. Suresh, P. Akkisetty, K. Morris, and G. Reklaitis.: Ontological informatics infrastructure for pharmaceutical product development and manufacturing. Computers and Chemical Engineering 30, 1482–1496. (2006)
2. International Society for Measurement and Control.: Batch control part 1 models and terminology (1995)
3. Klein, M. R. and L. B. Methlie.: Knowledge-Based Decision Support Systems: With Applications in Business (2nd ed.). New York, NY, USA: John Wiley & Sons, Inc. (1995)
4. Shim, J. P.,Warkentin, Merrill, Courtney, F. James, Power, J. Daniel, Sharda, Ramesh, Carlsson, and Christer.: Past, present, and future of decision support technology. Decis. Support Syst. 33 (2), 111–126. (2002)
5. Alavi, M. and P. G. W. Keen.: Business teams in an information age. Inf. Soc. 6 (4), 179–95.) (1989)
6. Courtney, J. F.: Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS., Decision Support Systems. 31 (1), 17–38. (2001).
7. Mitroff, I. I. and H. A. Linstone.: The unbounded mind: Breaking the chains of traditional business thinking. Business Horizons 36 (5), 88–89 (1993)

8. Simon, F. and T. Murray.: Decision support systems. Commun. ACM 50 (3), 39–40. (2007)
9. A. Leff; J.T. Rayfield. Web-application development using the Model/View/Controller design pattern Enterprise Distributed Object Computing Conference, 2001. EDOC '01. Proceedings. Fifth IEEE International. (2001)
10. Muñoz, E., Espuña, A., & Puigjaner.: L. Knowledge Management Tool for Integrated Decision-making in Industry: Technology for ontology development aimed at integrated decision support systems in process industries. LAP LAMBERT Academic (2012)
11. ISO/IEC. Systems and software engineering — Lifecycle profiles for Very Small Entities (VSEs) —Part 5-6-2: Systems engineering — Management and engineering guide: Generic profile group: Basic profile. TR 29110-5-6-2. (2014)
12. PMI, A Guide to the Project Management Body of Knowledge, Fifth ed., Pennsylvania: Project Management Institute, Inc. (2013).
13. CMMI Product Team.: CMMI for Development, Version 1.3 (CMU/SEI-2010-TR-033). Retrieved from the Software Engineering Institute, Carnegie Mellon University (2010)
14. Humprey, W.: The Team Software Process, Cargenie Mellon, Software Engineering Institute. Technical report CMU/SEI-2000-TR-023, ESC-TR-2000-023 (2000)

SmartLand-LD: A Linked Data approach for Integration of heterogeneous datasets to intelligent management of high biodiversity territories

Nelson Piedra[1] and Juan Pablo Suárez[1],
[1] Universidad Técnica Particular de Loja, Ecuador
{nopiedra, jpsuarez}@utpl.edu.ec

Abstract. A crucial issue in the development of smart sustainable cities and territories is the use of standards and approach's to ensure interoperability – so that equipment and systems produced by different vendors work together seamlessly – and to reduce costs through economies of scale. In this work, we investigate the core challenges faced when consuming multiple data sources for smart management of territories of high biodiversity using the Linked Data approach. We design a framework to achieve better data interoperability and integration by republishing real-world data into linked data: the SmartLand-LD framework. SmartLand-LD will work as a flexible and distributed ecosystem of autonomous and heterogeneous data sources (in content, technology, storage and structure) that being interoperated will facilitate access and the mixture of an extraordinary range of information. In order to fully achieve the benefits of SmartLand Initiative, it is crucial that information not only be available, but also is put into a context in order to create new knowledge and enable a diversity of applications. The information about resources is published in a way that allows interested human users and software applications to find and interpret them.

Keywords: SmartLand, Open Data, Linked Data, Semantic Web, Interoperability, RDF. URI.

1 Introduction

The convergence of information and communication technologies is producing changes in the management of urban and rural environments. What is more, these territories are different from any other reality that has been experienced until the present day. During great part of the 20th century, the idea that a territory could be considered smart was merely a matter of science fiction. However, the view that a city or territory could become smart, or sensitive, converts this idea into a new reality, namely thanks to the expansion of connectivity, mobile computing, advances data sciences, the proliferation of all types of gadgets linked to the Internet of Things, the possibilities of executing analytics of vast volumes of heterogeneous data, advances made in artificial intelligence, machine learning and predictive models.

Building an ecosystem for the integration and interoperability of data is a complex task. The technology, society and policy interfaces should worked in harmony, prioritizing needs, enabling integration by breaking barriers of existing information

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_19

silos and delivering purposeful technology. In this way, Linked Data (LD) approach provides a global environment for describing the objects, relationships and their significant properties. Therefore, this approach reduces duplication of effort when describing resources and their attributes, and fosters the creation of a global information graph encompassing all the information needed to perform complex queries and actions [2].

In this work, the authors present an approach for intelligent management of territories of high biodiversity: the SmartLand Initiative. SmartLand is about the smart management of the interactions between people and territories. The SmartLand initiative promotes the coherent and effective implementation of the three vast and ambitious objectives of the Convention on Biological Diversity: the conservation biological diversity, the sustainable usage of their components, and the creation of proposals that seek the just and equitable distribution of the benefits deriving from the usage of the territory's resources [1]. Authors describes a general framework for the publication of data that were collected from heterogeneous datasets, that is, by following the Linked Data design issues. The first section describes the vision the smart management of cities and territories; the following section presents the organization of the SmartLand initiative. Next section describes the role of Semantic Web (SW) and Linked Data (LD) as a way to reach semantic data interoperability. In the next section describes the proposed framework for publication of linked data from some existing datasets, and one way to define and reuse vocabularies that can be used to represent Linked Data. Finally, last section covers the respective conclusions and future areas of research.

2 Smart Management of Territories

2.1 The intelligent management of urban areas

Cities are the nucleus of consumption of global resources. The number and the proportion of the urban population will grow in the future. It is estimated that global cities occupy approximately 2% of the planet's land surface, consume between 60-80% of energy, and are responsible for 75% of carbon emissions. In addition, many of the decisions carried out by inhabitants of the cities directly affect biodiversity and eco-systems. In light of this, the current trend is to seek to create intelligent cities, or Smart Cities. In last two decades, the concept of smart city [6] has become more and more popular in scientific literature and international policies. The main reason seems related to the principal role of cities and territories in the social and economic aspects of people worldwide, and in the huge impact on environmental sustainability [7]. The idea of smart cities refers to the promotion of more "intelligent", sustainable and inclusive cities through the promotion of appropriate technological innovations [8]. To understand the concept, it is important to recognize why cities and territories are considered the key element in strategic plans for the future [9][10][11]. The citizen should be in the center of attention. However, computing, information science and engineering play a crucial role in the innovation and creation of new public as well as in developing the concept of smart management.

2.2 The intelligent management of bio-diversity territories

The interaction processes between citizens and bio-diversity potentially leads to conceptual gaps that should be closed if we wish to make progress towards the creation and consolidation of smart cities and territories. The complexity of its management requires interdisciplinary focuses based on research, development, innovation and public policies. In this context, some of the contemporary challenges in research are focused on the development and maintenance of habitable, sustainable and resilient territories, as well as on the solution of challenges within the context of sustainability, eco-systems and bio-diversity. The smart management of a territory (which we will refer to as SmartLand in this paper) refers to the promotion of more sustainable, “intelligent”, use of biodiversity and fair and equitable sharing of the benefits arising from the use of biodiversity resources, through the promotion and usage of appropriate technological innovations that take action for smart management of interaction between humans and bio-diverse territories.

3 About the SmartLand Initiative of UTPL

ICTs, open data, IoT, Semantic Web, smart cities, smart lands and collaboration with all key stakeholders including citizens, researchers, governments and communities will be key to the achievement of all the United Nations’ Sustainable Development Goals (SDG). A crucial issue in the development of smart sustainable cities and territories is the use of standards and approach’s to ensure interoperability – so that equipment and systems produced by different vendors work together seamlessly – and to reduce costs through economies of scale.

The complexity of the management of extensive territories requires interdisciplinary focuses based on research, development and technological innovation. In light of these challenges, the Universidad Técnica Particular de Loja (UTPL) has promoted an institutional initiative called **SmartLand** (see http://smartland.utpl.edu.ec). For UTPL, the concept of smart management within the field of the administration of a territory is strictly linked with multifaceted research and management of mega diverse ecosystems.

SmartLand seeks to create indicators that facilitate planning by means of multiple time scales, increasing the possibility of making territories more intelligent in the long term, that is, by means of continual reflection during the short term. In order to improve the understanding of the indicators, the initiative promotes strategies and technological tools that are necessary for the collection, interoperability, integration, data analytics, visualization, and preservation of large collections of data and information.

Highlights of Initiative. Since 2014, SmartLand executed 108 projects with approximately 300 researchers involved (84 international) and 2000 students; consolidated monitoring systems for weather, hydrology, seismology, wildlife, mental health and university orientation; in addition, five observatories on socio-environmental conflicts, tourism, strategic communication, family entrepreneurs and territory & economy.

Initiative Organization. SmartLand comprises multidisciplinary research-action programs, in 12 working packages executed in three provinces encompassing Coast, Andes and Amazonia ecosystems from Southern Ecuador. The southern region is the smallest of Ecuador territory, but the most bio-diverse. The Social innovation approach links professors, students and identified beneficiaries to improve public and private decision-making process and natural resources management. Work packages help contribute to the successful completion of the initiative's core objectives via research projects proposed by researchers from different knowledge domains. Each work package is composed by research projects that improve the scientific understanding of the environment, i.e. with the aim of adding value to social, biological, environmental and infrastructural indicators. The work packages in which the initiative is focused are: Patrimonial, cultural, tourism and recreation goods; Biodiversity and integrity of the eco-system; Cartography and geomorphology; Climatology; Education: indicators of quality and coverage; Energy and telecommunications; Infrastructure and transportation; Hydric resources and water quality; Public health; Production, entrepreneurship, innovation and economic indicators; Society, human mobility and values; Sustainable usage of biodiversity. The work packages transversely share a set of strategic technological objectives for information management, which are reflected in the availability of an open web data architecture which encompasses collection, data preservation, reusing, analysis, predictive modeling and data visualization. Each research group works on its own research questions, while sharing four Strategic Objectives (SO).

SO-1. Consolidate a system for the preservation of knowledge that guarantees free access. SmartLand promotes the creation of a technological infrastructure, which from a distributed focus, helps collect, monitor, and reuse data and information of high heterogeneity in various fields of knowledge: biodiversity, ecosystems, economics, territory, entrepreneurship, society, values and identity, among others.

SO-2. Improve our understanding of the indicators and their permanent monitoring. SmartLand promotes a strategy of sustainable collection of data, which is based on networks of sensors and monitoring systems of long-term indicators.

SO-3. Improve the management of territories through the efficient use of resources, the preservation of their heritage and sustainable use. SmartLand seeks to exploit, model, visualize, and monitor various variables, as well as develop probabilistic sceneries, virtual territorial ordinances, and enhance the analysis of decision-making.

SO-4. Strengthen the capacities and potentialities of citizens. SmartLand promotes the responsible participation of the citizenry by strengthening the identity and the value of human beings with the environment.

4 Data Interoperability and Semantic Web

4.1 Challenge: Building an ecosystem for global interoperability of data

The SmartLand "semantic architecture" must ensure the interoperability of different types of data and systems (See Fig. 1). Interoperability involve reaching agreements, sharing strategies and technologies that integrate and interoperate all the elements of information that make up the scope of the initiative. Silos includes different data collection mechanisms, such as on-site sensor networks, survey field data processes, data collection modules, collection models and processing of the raw data, data distribution systems and services, data prediction modules, decision making support tools and other data visualization products that are key for the end users.

Therefore, the following challenges must be addressed: (i) heterogeneity in the representation of data and semantics, which results in a variety of data models and formats (ii) The heterogeneity of the systems that manage the data and that support various modes of access and consultation. (iii) The challenge of processing the great volumes of data and the solution of many problems in society, business and academia; (iv) The loss of the interpretive context or the distortion of the meaning of the data when they influence other semantic borders in which the meanings, the terminology and the vocabulary are different: (v) Veracity, diversity of scales, privacy, bad usage and/or abusage of published material, among others.

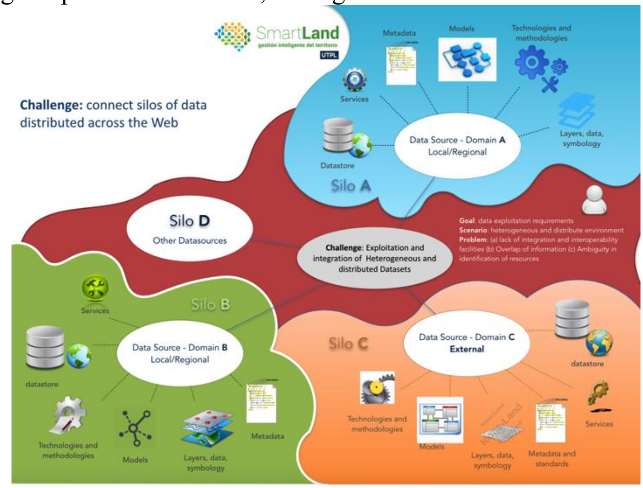


Fig. 1. The SmartLand challenge: Connect silos of data.

4.2 Semantic Web and Linked Data to Connect Silos Across The Web

The SW, a movement led by the World Wide Web Consortium (W3C), is envisioned as a decentralized global information space for sharing machine-readable data and meanings with a minimum of interoperability and integration costs. The SW promotes Linked Open Data (LOD) and encourages institutions to publish, share and cross-link their data using the web. This vision presents four main challenges: the global identification of resources, the use of formal ontologies as a way of specifying content-specific agreements for the sharing and reuse of knowledge among human beings and software entities, the distributed modeling of the world with a shared data model, and the infrastructure where data, schemas and formal ontologies can be published, shared, found and re-used in supporting knowledge sharing activities. According to Tim Berners-Lee, the inventor of the Web and the founder of the LOD project, SW not only seeks to publish data on the Web, but also aims to create links between data, which would help to explore the Web of data – for both human beings and machines [12]. With linking data, when we consult a resource, we may access other information resources published on the Web [14]. Linked Data provides the opportunity for publishers to contextualize their data, adding richness and depth. Tim Berners-Lee proposed four design principles to linking data by means of the Web [13]: 1) Use URIs (Uniform Resource Identifiers) as names for things; 2) Use HTTP URIs so that people can look up those names; 3) When someone looks up a URI, provide useful information, using the standards (RDF*[1], SPARQL[2]); 4) Include links to other URIs so that they can discover more things.

4.3 The Challenges of Linking Data

In order to use the Linked Data approach in SmartLand-LD, for data interoperability and data integration in a distributed, autonomous and heterogeneous environment, we needed to address the following challenges:

(1) Use open data and open platforms to mobilize collective knowledge. Open data have the potential to help communities understand, evaluate, and implement technology solutions tailored to specific community needs and objectives. Developing a long-term technology innovation strategy and a participative policy could be achieved using open data as an effective policy instrument through to digital infrastructure. For SmartLand initiative to effectively function, and for society to reap the full benefits from scientific endeavors, it is crucial that science data be made open and machine-readable.

(2) Re-publish existing data to semantically linked data. Establish the semantics and ontologies that must be leverage (re-used) in a particular domain. In addition, since several ontologies will be used, it will be necessary that the concepts of different ontologies be aligned and related to each other.

(3) Satisfy the 'Cool URIs' requirement of the SW, such that they improve the interoperability between independent systems, and to make the URIs as consistent and persistent as possible.

(4) Make the potentially 'linkable data' really link together and enable complex queries in a distributed and global data-space. The essence of LD is to allow data to connect to each other.

(5) Enable a way to serve data and semantics in a programmable web-compliant way so that more users can use them. The majority of Linked Data are currently served through SPARQL endpoints, which require users to be familiar with the domain ontologies to query the data.

(6) Implement a procedure to track the provenance of Linked Data published. Since a linked data publication process will publish existing data sources into Linked RDF data, there is a need to trace where the data came from, and what tools were used for such a transformation.

(7) Define the use of licenses. All data or collections of data that are shared should be published with a clear and explicit statement of the wishes and expectations of the publishers with respect to re-use and re-purposing of individual data elements, the whole data collection, and subsets of the collection.

5 The SmartLand-LD Approach

This section presents the SmartLand semantic framework architecture that has been designed based on the requirements defined in previous section. The underlying concept of SmartLand is like a constellation of components of various scales and types (both digital and analogue), which are connected by means of multiple networks, and that provide continual data with respect to environmental, biological,

[1] The Resource Description Framework, (RDF) is a family of specifications from the W3C, which was originally designed as a model of data for metadata. RDF is a syntax for representing and modeling data and semantics in the Web RDF breaks every piece of information down in triples: Subject – a resource, which may be identified with a URI; Predicate – a URI-identified reused specification of the relationship; and, Object – a resource or literal to which the subject is related.

[2] SPARQL is a standardized language for querying RDF data.

social, and economic factors of the movement of people and materials, i.e. in terms of the dynamics of the ecosystem, and decision making, and the physical and social makeup of the region. The overall architecture of the framework is presented in Fig. 1. As we can notice, at the lower level there are existing repositories. The metadata of these repositories are harvested and stored in the SmartLand repository (http://ambar.utpl.edu.ec), which is located in the middle level of the architecture. Moreover, in the middle level of the architecture there are components for Linked Data Publication. Finally, in the upper level of the architecture there is different user interfaces (e.g. a searching mechanism for accessing the RDF triple store, a recommender system for recommending suitable and appropriate indicators). The next section elaborates on the components of the architecture in more detail.

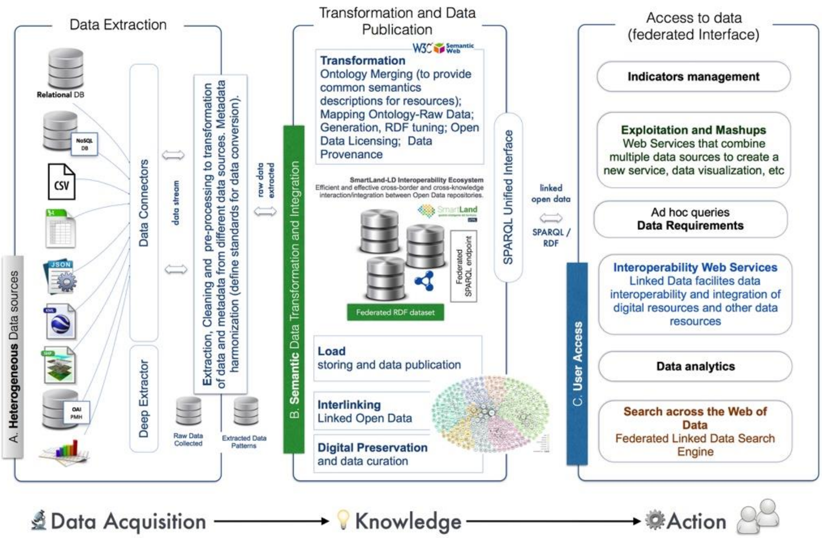


Fig. 2. **SmartLand-LD Overview:** Framework for data integration of SmartLand Datasets. On the SmartLand-LD all information has to be expressed as statements about resources or things. Resources (such as places, people, species, products, process, ideas and concepts such as ontology classes) are identified by URIs [RFC3986]. All the URIs related to a single real-world object should also be explicitly linked with each other to help information consumers understand their relation. This modeling approach is at the core of RDF. RDF provides all necessary framework to create meaningful and comprehensive (both to humans and web engines) data and to share them among different systems and with LOD universe.

5.1 Components

The SmartLand-LD Framework consists of a set of pluggable components. These components are organized as data extraction, data transformation and data output components. Since the data coming to the component may come from different source which commonly are of disparate systems resulting in different data formats, SmartLand-LD uses processes based on ETL methodology to make use of the data. These processes are extraction, transformation and loading. The figure below shows the schematic architecture of Linked Data applications that implement the crawling/data warehousing pattern. The figure 1 highlights the flow of the data interoperability process that is currently supported by SmartLand-LD. At present time, we have implemented the following components:

5.1.1 Data Collector

The data collector is a core component of SmartLand-LD that collects and stores information from different sets of data. The component provides one central point for data collection across database servers and applications from SmartLand projects and monitoring systems. This component gathers data from several sources and utilizes these data to serve as vital information for the SmartLand-LD. This collection point can import data from a variety of primary sources and is not limited to structured data. These data will be used to support informed actions. The component stores the collected data in http://ambar.utpl.edu.ec a data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data. The data collector uses CKAN [5] that enables to adjust the scope of data collection and to manage the data that are collected by setting different data retention periods. The data collector supports dynamic tuning for data collection and is extensible through its API. Whenever a new data is detected, the program automatically does its function to update and transfer the data to the ETL process.

5.1.2 Data Extraction

Data extraction process is a very important aspect of SmartLand-LD. Data extraction is a process that involves retrieval of all format and types of data patterns. In order to meet this requirement various tools have been developed to handle this process. The data extraction process in general is performed through the data connector's module, with customized functionality for each type and nature of data sources. The extraction phase entails retrieving (extracting) raw data from a given data source using the appropriate plug-in, depending on the technology of the source. Data patterns are imported into an intermediate extracting system before being processed for data transformation where they will possibly be padded with metadata before being exported to another stage in the LD transformation and publication work flow. The extracted data are added to a relational database, whose flexibility is ensured with schemas in the form of triplets (tables with three columns that will store subjects, predicates and objects). Data extraction either runs constantly or on a user-defined schedule. Since the component needs to do other processes and not just extracting alone, have been implemented programs that repetitively checks on many new data updates. This way, the code just sits in one area of the data collector sensing new updates from the data sources. These data will be further used for semantic transformation.

5.1.3 Global Resource Identification

With the increasing exchange of metadata, different identifier systems will clash in a data ecosystem. Any type of additional persistent identifier is useful to linking and fetches more contextual Information [3]. In SmartLand-LD, each identifier is a URIs. A URI is a string of characters used to identify the name of a resource. Such identification enables interaction with representations of the resource over a network, typically the World Wide Web, using specific protocols. Schemes specifying a concrete syntax and associated protocols define each URI. The resolution is the key mechanism enabling a system to locate and access the identified object or information related to it on the Web. To resolve a URI means either to convert a relative URI reference to absolute form, or to dereference a URI or URI reference by attempting to obtain a representation of the resource that it identifies. The 'resolver' component in document processing software generally provides both services.

5.1.4 SmartLand-LD Core Vocabularies and Ontologies

Regardless of whether information is in people's heads, in physical or digital documents, or in the form of factual data, it can be linked. A Core Ontology or Vocabulary is a simplified, reusable, and extensible data model that captures the fundamental characteristics of a data entity (properties, relations and constraints) in a context-neutral way. Core vocabularies can be used in interoperability agreements for enabling information systems, developed independently, to exchange information, thus making it possible for information to travel across borders and domains.

Good quality description metadata can improve the discoverability and integration of open datasets. Core Vocabularies help describe data entities by defining their components. They are designed to become the foundation of new, context-specific vocabularies to be used for exchanging public sector information. They can also be used as pivotal vocabularies for mapping and aligning other vocabularies. In this way, as part of the SmartLand-LD implementation process, an attempt to map heterogeneous metadata standards, encodings, code lists and also definition of elements and thus achieve interoperability was made. Besides the common vocabularies: RDF, RDFS, XSD, SKOS, OWL, PROV, VoID the RDF Data Cube vocabulary is used to describe multi-dimensional statistical data, and SDMX-RDF for the statistical information model. The vocabularies used in SmartLand-LD as core vocabularies are:

DCAT, an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Smartland initiative. DCAT is a W3C recommended specification that can be used for homogenizing metadata of datasets hosted on different projects and allows for querying them using a uniform vocabulary. Its basic use case is to enable cross-data portal search for data sets and make data better searchable across borders and sectors. This can be achieved by the exchange of descriptions of datasets among datasets. By using DCAT to describe datasets in data catalogues, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogues. It further enables decentralized publishing of catalogues. DCAT is widely used as a metadata specification to connect geospatial and statistical metadata across datasets in SmartLand and related projects using interoperability programs.

SDMX, The Statistical Data and Metadata eXchange was designed to develop and use more efficient processes for exchange and sharing of statistical data and metadata among international organizations and individual members.

RDF Data Cube. The Data Cube vocabulary is a core foundation that supports extension vocabularies to enable publication of other aspects of statistical data flows or other multi-dimensional data sets. The model underpinning the Data Cube vocabulary is compatible with the cube model that underlies SDMX.
DCMI Metadata Terms, an authoritative specification of all metadata terms maintained by Dublin Core Metadata Initiative (DCMI) -- elements, element refinements, encoding schemes, and vocabulary terms (the DCMI Type Vocabulary).
Geodata from a global perspective: ISO 19115:2003, ISO 19139 and 19115:20141 are standard of the International Organization for Standardization (ISO), defines how to describe geographical information.
FOAF, a project devoted to linking people and information using the Web. FOAF integrates three kinds of network: social networks of human collaboration, friendship and association; representational networks in factual terms, and information networks that use Web-based linking to share independently published descriptions of this inter-connected world.
VoID, is mainly used in Linked Data applications to express an RDF dataset. VoID helps in data discovery and cataloguing as well as helps consumers finding the right data on the web.
PROV, provides basis for representing provenance information about the entities, activities, and actors involved in producing a data item for assessing trust, quality and reliability of the data item; using classes - Entity, Activity, Agent, Role, Time, Usage and Generation.

5.1.5 Linked Data Conversion and Publication

Linked Data approach is used for integrating data housed in different types of sources, in different formats, both structured and unstructured. We put forward a set of components for that, inspired by the work of Heath & Bizer [14]. The goal of the transformation and publication component is to harmonize, cleanse, and prepare for storing on LD metadata extracted. The first step is to design and assign persistent URIs to the data and describe it in RDF using commonly-agreed vocabularies. In the case of structured data (e.g. stored in relational databases, CSV, Excel or XML files), RDF wrappers, APIs have been adapted and used for transforming it to RDF and publishing it on the Web in structured sources. Whereas for unstructured data, entity extractors and text mining techniques have been employed for discovering the data entities to be published as Linked Data. The SmartLand-LD, transformation pipeline consists of the following steps:

Data Preprocessing - Disambiguation and data cleansing. This component filters data according to different quality assessment policies and provides for remove uncertainty of meaning and fusing data according to different conflict resolution methods [4]. This includes ridding extracted data and metadata of any additional information that will not be included in your published data sets.

Data Modeling. Choose appropriate vocabularies for extracted dataset from existing ones. This include, reuse established vocabularies and additional models to ensure smooth data conversion to RDF. An expressive mapping module was created to translate data and metadata extracted into a consistent, target vocabulary (metadata harmonization for data conversion).

Converting extracted data to RDF. Vocabulary mappings are expressed using an own mapping language. The language provides for simple transformations as well as for more complex structural transformations (1-to-n and n-to-1) and property value transformations such as normalizing different units of measurement or complex string manipulations. Data visibility and usability can vastly improve through interlinking with other information sources.

Specify licenses. Provide a clear license for each dataset convert to RDF.

Link data to other datasets. While data is being represented in RDF, it should be linked to other data coming from trusted sources, in order to provide context and enrich it's meaning. Regarding the linking process, we have used Silk Link Discovery Framework[3] for discovering the same real-world entity, owl:sameAs links between our dataset and external datasets, i.e., World Bank Linked Data, GeoNames and DBpedia; and an own application for validating the links discovered by Silk. In order to specify the

[3] SILK: http://www4.wiwiss.fu-berlin.de/bizer/silk/

condition that must hold true for two entities to be considered a duplicate, the user may apply different similarity metrics, such as string, date or URI comparison methods, to multiple property values of an entity or related entities. An identity resolution component discovers URI aliases in the input data and replaces them with a single target URI based on user-provided matching heuristics. For each set of duplicates that have been identified by Silk, SmartLand-LD replaces all URI aliases with a single target URI within the output data. In addition, it adds owl:sameAs links pointing at the original URIs, which makes it possible for applications to refer back to the original data sources on the Web. If the SmartLand-LD input data already contains owl:sameAs links, the referenced URIs are normalized accordingly.

Publication. Once the extracted data has been transformed into RDF, these data needs to be store and publish in a triplestore. In our scenario we decided that the converted data are load in a Virtuoso RDF triplestore via the Virtuoso Loader. For the metadata information we have used the vocabulary of Interlinked Datasets (VoID[4]) for expressing the metadata about each dataset. For provenance tracking, SmartLand-LD employs the W3C Provenance Model (Prov[5]). After Linked Data is converted, linked and published, any system capable of running semantic queries (using the SPARQL query language) can extract and re-use the relevant information.

Promote data reuse. Given only a URI, machines and human users should be able to retrieve a description about the resource identified by the URI, and links to related data from the Web. Such a look-up mechanism is important to establish shared understanding of what a URI identifies. Machines should get RDF data and humans should get a readable representation, such as HTML. The communication between human beings or machine users, and the RDF store is realized in the form of SPARQL queries. In addition, since one of the requirements was to ensure that the concepts are interlinked at either one of the endpoints, each endpoint had to include each observation's reference area as well as its interlinked concept.

6 Conclusions and Future Works

In this work we have presented how Linked Data can be used to lower the high barriers of integrating in a large, heterogeneous and distributed data ecosystem. LD is a set of design principles for sharing machine-readable data on the Web. Linked data offers a number of advantages for: enables for semantic interoperability; data integration with small impact on legacy systems; enables creativity and innovation through context and knowledge-creation. SmartLand is committed to the creation of an ecosystem of data that supports decision-making based on the creation of social, academic and technical bridges that facilitate the exchange and opening of scientific and public data using Web Semantics and LOD.

SmartLand-LD can be used to connect information pieces from different data sources -mitigating the data interoperability problem. This allows us to collect and linking the data that is needed to create relevant recommendations for decision-making, as well as increasing the value of data in general. To demonstrate the validity of our approach we have incorporated this approach to the SmartLand initiative with Linked Data. Preliminary evaluations show a significant improvement in data reuse and integration. SmartLand-LD facilitates data integration and enables the interconnection of previously disparate dataset. The increased (re)-use of data triggers a growing demand to improve data quality. Through crowd-sourcing and self-service mechanisms, errors are progressively corrected. The availability of linked open data gives rise to new services offered by the academia, civil society, public and/or private sector. The reuse of data in SmartLand applications leads to considerable cost

[4] VoID: http://semanticweb.org/wiki/VoID
[5] PROV: http://www.w3.org/TR/prov-primer/

reductions. SmartLand-LD offers a platform that integrates and facilitates interoperability, re-usage and preservation of collated data, the integration of monitoring systems, analytic functionalities of vast volumes of information, and the generation of predictive models that support decision making processes for the smart management of a territory. The selection of indicators to be monitored should therefore focus on sustainability and its continual attainment over time. The element of quality is essential when understanding changes in the behavior of indicators over time, and in evaluating the relations/interactions among indicators, and also of creating predictive models. As future research we will continue the framework implementation with bigger data sets from different domains and expand the geographical area of interest.

References

1. Balmford, A., Bennun, L., Ten Brink, B., Cooper, D., Côté, I. M., Crane, P., ... & Gregory, R. D. (2005). The convention on biological diversity's 2010 target. Science, 307(5707).
2. W3C (2011) Library Linked Data Incubator Group: Use Cases. Report. 25 October. Available at: http://www. w3.org/2005/Incubator/lld/XGR-lld-usecase-20111025/. 2011.
3. COAR. COAR Roadmap. Future Directions for Repository Interoperability. Comparison of TRAC Checklist and PLEDGE Policy List. 2015
4. Piedra, N., Chicaiza, J., Lopez-Vargas, J., & Caro, E. T. (2016, October). Guidelines to producing structured interoperable data from Open Access Repositories. In Frontiers in Education Conference (FIE), 2016 IEEE (pp. 1-9). IEEE.
5. CKAN. The open source data portal software. Available at: http://ckan.org
6. Albino, V. Berardi, U. and Dangelico, R. M. "Smart Cities: Definitions, Dimensions, Performance, and Initiatives," Journal of Urban Technology, vol. 22, no. 1. 2015.
7. Mori, K. Christodoulou, A. Review of sustainability indices and indicators: Towards a new City Sustainability Index (CSI), Environmental Impact Assessment Review, Vol. 32, No. 1, pp. 94-106. 2012.
8. Hollands, R.G. Will the real smart city please stand up?, City: analysis of urban trends, culture, theory, policy, action, Vol. 12, No. 3, pp. 303-320. 2008.
9. Nam, T. Pardo, T.A. Conceptualizing Smart City with Dimensions of Technology, People, and Institutions, Proc. 12th Annual International Conference on Digital Government Research. 2011.
10. Deakin, M. Al Waer, H. From intelligent to smart cities, Intelligent Buildings International, Vol. 3, No. 3, pp. 140-152. 2011.
11. Dirks, S. Gurdgiev, C. Keeling, M. Smarter Cities for Smarter Growth: How Cities Can Optimize Their Systems for the Talent-Based Economy. Somers, NY: IBM Global Business Services. 2010.
12. Berners-Lee, Tim; Hendler, J.; Lassila, Ora. The Semantic Web, Scientific American, 284(5), 34-43. 2001.
13. Berners-Lee, Tim. Linked Data-Design Issues. 2006.
14. Heath, Tom & Bizer, Christian. Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, 1(1), 1-136. Morgan & Claypool. 2011.

Software Systems, Applications and Tools

Usability analysis: Is our software inclusive?

Hans Guerrero[1], Vianca Vega[2]

[1] Escuela de Ingeniería, Universidad Católica del Norte, Chile
{hgb003@alumnos.ucn.cl}
[2] Departamento de Ingeniería de Sistemas y Computación, Universidad Católica del Norte, Chile
{vvega@ucn.cl}

Abstract. Usability is considered as one of the most important quality attributes, being a fundamental aspect in all software products. It is even more crucial in those systems that are designed for a wide variety of users, called by different authors as inclusive products. It can be observed that the traditional requirements elicitation process does not consider the special needs of people with disabilities. Not even in the proposals of inclusive systems are special techniques included. In this context, the present article seeks to provide a broad perspective on different factors influencing usability identified in several investigations focused on software products for users with sensory disability. A comparative chart is presented showing the strengths and weaknesses of several models, guides and frameworks designed to satisfy the needs of this group of users. Finally, several metrics of usability requirements are proposed, quantifying the process of requirements capture by means of usability factors in inclusive systems.

Keywords: Usability, Inclusive Design, Requirements Elicitation, Sensory Disability.

1 Introduction

Nowadays, software products are a part of people's everyday life, contributing to their quality of life. However, it has been detected that a majority of the systems are not oriented to satisfying the needs of people with disabilities, mainly of the sensory type (sight and hearing). Thus, one of the great challenges of software engineering is integrating users with disabilities from the early stages of a software product.

According to the World Health Organization (WHO), the concept of disability is defined as an umbrella term covering impairments, activity limitations, and participation restrictions, being a complex phenomenon that reflects the interaction between the characteristics of the body of a person and the characteristics of the society in which the person lives [1]. Hence, disability is a consequence of the functional impairments of a person that prevent them from successfully interacting with a product or service in a given context [2]. Statistics provided by the WHO show that there are 285 million people with visual impairment worldwide [3] and 360 million people with hearing loss [4]. In Chile according to the 2012 Census, the population with one or more disabilities is 2.119.316 people, which corresponds to the 12,7% of the total inhabitants of the

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_20

country. Regarding sensory disability, in Chile, 890.569 inhabitants (5,35%) suffer from blindness or sight impairment even when using glasses, and 488.511 inhabitants (2,94%) are deaf or have hearing impairment even when using hearing aids [5].

Assistive technology (AT) is a generic term that corresponds to devices created for people with motor and sensory impairments that are used for assistance, adaptation and rehabilitation [6]. AT promotes more independence by enabling people with disabilities to carry out tasks that implied a great difficulty due to the interaction with the required technology [7]. ATs are commonly created for a specific number of users according to the type of disability, being difficult to holistically cater to all the users in one technology. Nevertheless, there are products and services that are designed for responding to the needs of a wider public, these are called inclusive designs [8]. Keates [9] defines the concept "universal access" or "inclusive design" as those settings, services or interfaces that work for people of all ages and abilities. According to the British Standards Institution, it is defined as the design of conventional products and/or services accessible and usable by as many people as it is reasonably possible, without the need of a special adaptation or a specialized design [10]. An inclusive software is a product that must be accessible and usable by all people (with disabilities or not), this has to be taken into account from the requirements specification until the validation of the products.

Accessibility and usability are two topics that are commonly analyzed separately, while other research indicates that accessibility and usability should be adequately integrated [11] [12]. Usability is related to how a software allows a group of users achieve specified goals in a specified context of use, on the other hand, accessibility describes the degree to which a product, device or service is comprehensible to as many people as possible [7]. In inclusive software, usability is considered as a critical attribute, due to the emergence of problems because of different factors that prevent the product from being totally usable.

In this research investigates on usability requirements elicitation for inclusive software product. The article is divided in four sections. The first section presents different predominant usability factors in inclusive software. In the second section presents a comparative chart with strengths and weaknesses of different works oriented to satisfying the needs of people with disabilities. The third section, metrics of usability requirement are detailed. Finally, conclusions and further work will be mentioned.

2 Usability factors

Definitions: Usability is considered as one of the most important aspects of software quality, being an attribute that verifies the ability of the software to be understood, operated and attractive to a wide range of users.

One of the most formal definitions of usability is established by the ISO-9241-11 standard that defines it as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use [13]. Effectiveness is understood as the accuracy and completeness with which users accomplish their goals. Efficiency is defined as the existing relationship between resources and time employed. Satisfaction is defined as the absence or presence of attractiveness, acceptance or comfort in the use of the product.

The ISO 9241-11 standard has been commonly used to assess the usability in a variety of software, for instance, Zitkus [14] applies the principles of this standard in an evaluation of websites in blind users. Here it is pointed out that effectiveness is related to successfully accomplished goals, efficiency depends on the tasks carried out correctly within the variables of time and degree of difficulty (measured by the number of errors made) and, on the other hand, satisfaction is related to the responses given to a questionnaire evaluating structure, navigability, search system and information. Leopirini [15] considers three levels of relevance for usability factors in disabled users. The first level considers effectiveness, the second efficiency, and the last one satisfaction. The authors argue that inability to meet these criteria could lead to users being unable to accomplish their tasks.

Nielsen [16] defines usability as a quantitative and qualitative measurement of the design of a user interface, grouped into five key factors: learnability, efficiency, memorability, errors and satisfaction.

The other formal definition of usability is established by the ISO 9126-1 standard that defines it as the capability of a software product to be understood, learned, used and attractive to the user, when used under specified conditions [17]. Subsequently, the ISO 9126 standard was replaced by the ISO 25010 standard [18] that kept the definition of usability subdividing it in: appropriateness recognizability, learnability, operability, user error protection, user interface aesthetics and accessibility.

The literature does not provide a concrete definition of usability, leaving to the researcher's judgment the formalization of the usability factors relevant to the study. This research seeks to present principles that must be analyzed in the elicitation of requirements, following the ISO 25010 standard, mainly in appropriateness recognizability, that is, the degree to which users recognize whether a software product is appropriate for their needs [18].

Usability in inclusive systems. Inclusive or universal design considers the design of products and services to be accessible and usable by everyone, regardless of their degree of disability, cultural background, age, among other characteristics [19]. The principles of universal design [20] are seven: (1) equitable use, (2) flexibility in use, (3) simple and intuitive use, (4) perceptible information, (5) tolerance for error, (6) low physical effort, (7) size and space for approach and use.

Digital inclusion has been a concept used as a strategy for ensuring that people have access to digital technologies and skills to use them [21].

Usability in inclusive systems has been considered as a critical attribute. Cano [22] evaluates different formative or serious games by means of a usability test, emphasizing effectiveness, efficiency, satisfaction, emotions and learning. Navarrete [23] considers in the design certain attributes such as ease of use, effectiveness, efficiency and satisfaction, related to the accomplishment of the users' goals [24]. Holzinger [25] points out that it is necessary to add the acceptance factor.

Among the models of inclusive systems that are related to the usability factors is the QUIM model ("quality in use integrated measurement") proposed by Seffah [26]. The

QUIM model is composed by eleven representative factors based on several standards and models from literature. The model adapts these factors and expresses them in criteria to then "map" these criteria into metrics. The eleven factors are: (1) efficiency, (2) effectiveness, (3) productivity, (4) satisfaction, (5) learnability, (6) safety, (7) trustfulness, (8) accessibility, (9) universality, (10) usefulness, (11) acceptance.

These factors are influenced by a specific category of users, tasks or settings. In Mesiti [7] an analysis on the QUIM model is carried out, showing that productivity and learnability are useful in the design of settings where the user is an active producer of knowledge; accessibility, universality and acceptance are useful for the possibility of designing interactive settings for users with disabilities; and usefulness is utilized to measure the ease of using the system, providing a representation regarding the real needs of the users.

The proposal of the technology acceptance model for inclusive education (TAM 4 IE) proposed by Silvia and Vilela in [27], is a model composed of five factors: subjective perception, perceived usability, perceived usefulness, future expectations and facilitating conditions. The aspect of perceived usability considers the following factors: learnability, memorability, accessibility and aesthetic. Satisfaction is considered as a subjective perception.

It has been detected that most of the works are not oriented to assessing usability in the requirements phase. This study has considered the primordial factors for the analysis in the capture of information from the users, such as learnability, understandability, compliance, effectiveness and satisfaction.

3 Related work

This section will present different models, frameworks and guides regarding inclusive systems, highlighting strengths and weaknesses from the perspective of requirements elicitation. Table 1 shows the results of the investigation.

The research has allowed developing a broad perspective on inclusive systems. Among the disadvantages of the studies described above, it is observed that most of them do not relate the users' needs with usability factors. In addition, a lack has been detected in metrics definitions in the requirement elicitation phase in inclusive systems. On the other hand, it has been found that it is necessary to formalize the process of requirement elicitation in the design of inclusive systems.

4 Usability metrics

Usability is an attribute that can be evaluated qualitatively or quantitatively. Usability engineering is an approach whose objective is to develop a software product taking into account a set of measurable characteristics.

Good [35] highlights that the specification of the measurable characteristics of usability is crucial to determine the requirements of usability of a product, or to measure if the completed product meets the requirements or not.

Table 1. Comparative chart of research in inclusive systems

Contribution	Advantages	Disadvantages
Keates and Clarkson framework [28]	Identifies users' desires and aspirations. Subsequently determines users' needs.	It does not provide metrics to validate if the product is really inclusive. It does not relate requirements with usability.
CEN / CENELEC Guide [29]	It allows considering the needs based on different clauses, addressing problems of usability and accessibility.	The guide does not provide mechanisms to relate factors of usability in quantitative form.
HAAT Model [30]	It relates users' needs to the activities of their daily living.	It does not formalize the process of capture of information from the early stages of the model. It does not include metrics.
CAT Model [31]	It categorizes user's requirements to analyze user's information.	It does not consider usability factors within the context of the model.
USERfit Model [32]	In the user analysis phase, it allows considering functional implications and necessary actions to counter impediments. It allows integrating the characteristics with existent standards in literature.	It does not include metrics in the user analysis. It does not address usability factors in depth.
Goodman's model [33]	It defines the problem and then details the user requirements. It integrates elicitation techniques in the process of analysis and data collection.	It does not consider metrics. It does not relate requirements with aspects of software quality.
Inclusive software development model [34]	Tasks and their attributes are identified by the development and design team.	Metrics are not proposed in the development cycle of the model. It does not include usability in the phase of task decomposition, it is only integrated in the evaluation of the model.

In inclusive systems, it is important to bear in mind two criteria: the requirements that define it and its merit when judged against those requirements [28].

Hersh [31] mentions two kinds of measurement for inclusive systems: (1) objective measurements, such as measures of system performance and services of human-assisted technology, like the time that the user takes to read a dataset or find a specific website using a screen reader; (2) subjective measurements, such as user satisfaction evaluations about the device and services, and other subjective measures regarding any change in quality of life.

Padda [36], whose thesis provides usability metrics based on different usability criteria, proposes one of the most important contributions in usability metrics. Based on these metrics, models have been proposed that relate metrics with software usability attributes [26].

Recent research refers to software usability metrics according to the ISO/IEC 9126-1 standard [37], grouped according to different factors: learnability, operability, understandability, attractiveness and usability compliance.

Currently, there are few works that refer to metrics that deal directly with attributes in the requirements elicitation process, however, there is an automated tool called Automated Requirement Measurement (ARM) that was developed by NASA's Software Assurance Technology Center (SATC) [38]. This instrument provides metrics to assess the quality of a requirements specification document.

ARM is an automated tool that seeks to help write requirements correctly by searching line by line of the document of requirements for specific words or phrases that SATC has identified as quality indicators. The result of ARM analysis are three reports, the first is a summary that counts the total number of times that each quality indicator appears in the requirements document; the second report is detailed, specifying the location of each statement identified by the tool in the source file and a copy of the specified statement; the third report is on weak phrases, it locates and specifies the indicators that have been considered as phrases that weaken the specifications.

In the Table 2 present different usability metrics - proposed by the authors - for requirements elicitation of inclusive software considering the factors: learnability [7, 16, 18, 26, 27], understandability [37], effectiveness [13, 23, 26], compliance [37], and satisfaction [13, 16, 22, 23, 26, 27].

5 Open challenges

The main challenge identified from the results of this research is the need to design a guide for elicitation of usability requirement for the formalization of the requirement of an inclusive software product. The present works shows that there is a lack of guidelines that support the development team for the qualitative and quantitative categorization of software usability.

Another challenge is the validation of the proposed metrics. This process will allow assessing objectively and subjectively usability in the early stage of the development of inclusive software. The investigation seeks to give a solution to the diverse needs that users with sensory disabilities could have, allowing the analyst or requirement engineer to focus on the problems, tasks and activities of the users, approving the use of different inclusive software.

Table 2. Usability metrics for the requirements elicitation of inclusive software.

Factor	Metrics	Parameter A	Parameter B	Formula
Learnability	Requirements familiarity	Number of requirements previously known by the eductor.	Total number of requirements.	A/B
	Requirements consistency	Number of requirements changed by the user	Total number of requirements.	1 - A/B
Understandability	Requirements integration	Number of requirements described.	Total number of requirements.	A/B
	Requirements demonstration	Number of requirements demonstrated by the user.	Total number of requirements.	A/B
	Function of understandable requirements	Number of interface requirements that refer to what the user understands.	Total number of requirements.	A/B
Effectiveness	Function of effectiveness of requirements	Number of requirements that could be successfully realized.	Total number of requirements.	A/B
	Requirements termination	Number of pending requirements (with missing information).	Total number of requirements.	A/B
Compliance	Level of compliance of the requirements	Number of requirements that could be correctly applied.	Total number of requirements.	A/B
	Function of compliance	Number of requirements valued for the design.	Total number of requirements.	A/B
Satisfaction	Level of satisfaction of the requirements	Degree of user satisfaction after applying the technique of requirements elicitation.	-	Weighting of parameter A.
	Level of satisfaction of the information provided	Degree of satisfaction of the eductor with respect to the information provided by the user.	-	Total weighting of parameter A.

6 Conclusions and future work

This study has examined the aspect of usability, considered by literature as one of the most important attributes of software quality. The research has been oriented to respond to how this property influences software products for people with disabilities. The article has looked into usability factors, giving as a result several principles that must be emphasized by inclusive systems, that is, products and/or services that are used by a wide range of users.

An analysis of the research related to usability requirements has been carried out by means of a comparative chart, showing advantages and disadvantages of each investigation regarding the process of capture of information from people with disabilities. Based on the results, different metrics of usability for inclusive software have been created, from the perspective of requirements elicitation. This is because this aspect is analyzed in the phase of development and evaluation of a software product. This will allow assessing in a quantitative way the usability requirements in inclusive systems by means of the factors of learnability, understandability, effectiveness, compliance and satisfaction of the requirements.

The analysis exposed has determined that to elicit the requirements of inclusive software it is necessary firstly to know the usability needs of the users, identify the actions (tasks) that are commonly carried out and find interaction problems with similar software. Having the requirements defined by different stakeholders, it is fundamental to carry out a categorization and measurement of the factors and criteria of usability, thus formalizing the attribute of usability in the requirements specification.

Recently, work has been done in the design of a guide of usability that will allow analysts to investigate about the attribute of usability from the early stage of a software product. In addition, research has been done about different techniques of elicitation that can be addressed in inclusive systems according to the user's kind of disability.

As future work, the usability guide worked with the metrics discussed in this article will be validate. In addition to applying elicitation techniques that have not been considered in the design of an inclusive software. The work carried out will be a contribution to software engineering, contributing socially to a wide range of users and fostering the development of new software products.

References

1. World Health Organization (WHO): Disabilities, http://www.who.int/topics/disabilities/en.
2. Keates, S: Design for the Value of Inclusiveness. Handbook of ethics, values, and technological design. Springer Netherlands (2014).
3. World Health Organization (WHO): Visual impairments and blindness, http://www.who.int/mediacentre/factsheets/fs282/en/.
4. World Health Organization (WHO): Deafness and hearing loss: http://www.who.int/mediacentre/factsheets/fs300/en/.
5. SENADIS (Servicio Nacional de la Discapacidad), http://www.senadis.gob.cl.
6. International Organization for Standardization. Assistive products for persons with disabilities – classification and terminology. DS/EN ISO 9999:2007(E). Genova (2007).
7. Mesiti, M., Ribaudo, M., Valtolina, S., Barricelli, B., Boccacci, P., Dini, S.: Collaborative Environments: Accessibility and Usability for Users with Special Needs. In Community-Built Databases: Research and Development. Springer, Heidelberg (2011).
8. Department of trade and industry: A study on the difficulties disabled people have when using everyday consumer products. In Government Consumer Safety Research, London (2000).

9. Keates, S: Designing for accessibility: a business guide to countering design exclusion. Lawrence Erlbaum, Mahwah (2007).
10. British Standard Institution: BS 7000-6: 2005 design management systems. Managing inclusive design. Guide. London (2005).
11. Aizpurua, A., Harper, S., Vigo, M.: Exploring the relationship between web accessibility and user experience. Int. J. Human–Computer Studies. 91, 13-23 (2016).
12. Petrie, H., Kheir, O.: The relationship between accessibility and usability of websites. In: Proceeding of the SIGCHI Conference on Human Factors in Computing Systems, New York, USA, pp. 397-206. ACM, New York (2007).
13. International Organization for Standardization: Ergonomic requirements for office work with visual display terminals – part 11: guidance on usability ISO 9241-11. Genova (1998).
14. Zitkus, E., Brigatto, A., Ferrari, A., Bonfim, G., Carvalho, I., Reis, T., Medola, F., Paschoarelli, L: Accessibility and usability of websites intended for people with disabilities: a preliminary study. In: Marcus A. (eds.) Design, User Experience, and Usability: Novel User Experiences DUXU 2016, LNCS, vol. 9747, pp. 678-688. Springer, Cham (2016).
15. Lepirini, B., Paternò, F.: Criteria for usability of accessible web sites. In: Carbonell N., Stephanidis C. (eds.) Universal Access Theoretical Perspectives, Practice, and Experience, UI4ALL 2002, LNCS, vol. 2615, pp. 43-55. Springer, Heidelberg (2003).
16. Nielsen, J.: Usability Engineering. Morgan Kauffmann, San Diego (1993).
17. International Organization for Standardization: Software process: improvement and practice ISO/IEC 9126-1. Genova (2001).
18. International Organization For Standardization: System and software quality requirements and evaluation, Square ISO/IEC 25010: 2011. Genova (2011).
19. Beyene, W.: Realizing inclusive digital library environments: opportunities and challenges. In: Fuhr, N., Kovács, L., Risse, T., Nejdl, W. (eds.) Research and Advanced Technology for Digital Libraries TPDL 2016, LNCS, vol. 9819, pp. 3-14. Springer, Cham (2016).
20. Principles of universal design, the center for universal design, North Caroline State University: https://www.ncsu.edu/ncsu/design/cud/about_ud/udprinciplestext.htm.
21. Real, B., Berlot, J., Jaeger, P.: Rural public libraries and digital inclusion: issues and challenges. Inormation Technology and Libraries. 33(1), 6-24 (2013).
22. Cano, S., Alghazzawi, D., Muñoz, J., Fardoun, H., Collazos, C., Bustos, V.: Applying the information search process model to analyze aspects in the design of serious games for children with hearing impairment. Universal Access in the Information Society (2017).
23. Navarrete R., Luján-Mora S.: Bridging the accessibility gap in Open Educational Resources. Universal Access in the Information Society (2017).
24. Petrie, H., Bevan, N.: The evaluation of accessibility, usability and user experience. In: The Universal Access Handbook, pp. 10-20, CRC Press (2009).
25. Holzinger, A., Searle, G., Wernbacher, M.: The effect of previous exposure to technology on acceptance and its importance in usability and accessibility engineering. Universal Access in the Information Society 10 (3), 245-260 (2011).
26. Seffah, A., Donyaee, M., Kline, R., Padda, H.: Usability measurement and metrics: A consolidated model. Software Quality Journal, 14(2), 159-178 (2006).
27. Silva, S., Vilela, L.: Technology acceptance evaluation by deaf students cionsidering the inclusive education context. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) Human-Computer Interaction – INTERACT 2015, LNCS, vol. 9296. Springer, Cham (2015).
28. Keates, S., Clarkson, P: Countering design exclusion: bridging the gap between usability and accessibility. Universal Access in the Information Society, 2(3), 215-225 (2003).

29. CEN/CENELEC Guide 6: Guidelines for standards developers to address the needs of older persons and persons with disabilities: ftp://ftp.cen.eu/BOSS/Reference_Documents/Guides/CEN_CLC/CEN_CLC_6.pdf.
30. Cook, A., Hussey, S.: Assistive technology: principles and practice. 4th edn. Mosby Inc., St. Louis, USA (2015).
31. Hersh, M., Johnson, M.: Assistive technology for visually impaired and blind people, chapter 1: Disability and assistive technology systems. Springer, London (2008).
32. Abascal, J., Arrue, M., Garay, N., Tomás, J.: Userfit tool. A tool to facilitate design for all. In: Carbonell, N., Stephanidis, C. (eds.) Universal Access Theoretical Perspectives, Practice, and Experience, UI4ALL 2002, LNCS, vol. 2615, pp. 141-152. Springer, Heidelberg (2002).
33. Goodman, J., Langdon, P., Clarkson, P.: Providing strategic user information for designers: methods and initial findings. In: Clarkson, J., Langdon, P., Robinson, P. (eds.) Designing Accessible Technology, pp. 41-51. Springer, London (2006).
34. Bonacin, R., Calani, M., Rodrigues, M.: An agile process model for inclusive software development. In: Filipe, J., Cordeiro, J. (eds.) Enterprise Information Systems, ICEIS 2009, LNBIP, vol. 24, pp. 807-818. Springer, Heidelberg (2009).
35. Good, M., Sine, T., Whiteside, J., Geroge, P.: User-derived impact analysis as a tool for usability engineering. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 241-246. Boston , USA (1986).).
36. Padda, H.: QUIM map: A repository for usability/quality in use. Master's thesis, Degree of Master of Computer Science at Concordia University (2003).
37. Rochimah, S., Rahmani, H., Laili, U.: Usability characteristics evaluation on administration module of academic information system using ISO/IEC 9126 quality model. In: Imternational Seminar on Intelligent Technology and its Applications (ISITIA), pp. 363-368, Surabaya, Indonesia (2015).
38. Carlson, N., Laphante, P.: The NASA automated requirement tool: a reconstruction. Innovations in Systems and Software Engineering, 10(2), 77-91 (2014).

Software testing education through a collaborative virtual approach

Juan P. Ucán Pech[1], Raúl A. Aguilar Vera[1], Omar S. Gómez[2],

[1]Universidad Autónoma de Yucatán, Mérida 97000, Yucatán, México,
[2]Escuela Superior Politécnica de Chimborazo, Riobamba 060106, Chimborazo, Ecuador,
{juan.ucan, avera}@correo.uady.mx, ogomez@espoch.edu.ec

Abstract. The use of collaborative virtual environments facilitates communication, coordination and cooperation within a group of people. In the field of Information and Communications Technology (ICT), collaborative virtual environments have been developed to support the learning of programming. With the aim of fostering software testing education by alternative means, we present findings from the use of a collaborative virtual environment used for software testing training. Particularly, we report an empirical study (controlled experiment) which assesses the effectiveness of detecting defects in instrumented programs with and without the support of an collaborative virtual environment (CVE). The CVE was used as part of a programming course at under-graduate level. The results suggest an equivalent effectiveness in defect detection for both types of participants, those who used the CVE and those who worked in a traditional manner. It was observed that with regard to effectiveness, for this type of task (the detection of defects), working virtually through a CVE versus working in a tradition manner (at the same time and in the same place) yields equivalent results. When collaborative work cannot be done in a traditional way, the use a CVE is an alternative approach equally effective for collaborating and learning on software testing.

Keywords: Software Testing Education, Computer-supported Collaborative Learning, Computer-supported Cooperative Work, Experimentation in Software Engineering.

1 Introduction

Education in software testing is still an open issue to address. Although most Computer Science programs offer at least one course on software engineering (SE), there is evidence to show that formal software testing training is deficient. For example, Chan et al. [1] found that less than 20% of software testing teams in Hong Kong had received formal software testing training at university. It is common practice for CS programs to briefly touch upon the topic of software testing as one of the issues addressed in SE courses [2]. According to Wong et al. [3], some US universities offer around 30 software testing courses as part of undergraduate CS programs but no one of the CS programs in the US has as prerequisite a software testing course. Recently Gómez et al. [4] found a relationship between the level of

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_21

exposure to CS knowledge and the quality of test cases generated by different types of students, Gómez et al. [4] found that students who have had less exposure to CS knowledge generate poorer quality test cases.

With the aim of encourage by alternative means the adoption of software testing education in CS programs, we developed a collaborative virtual environment (CVE) which can be used as a tool support for training in this area.

The rest of this document is organized as following: Section 2 presents an overview of related work. Section 3 shows the main features of the CVE developed. Section 4 presents the experiment conducted. Section 5 presents de results. Section 6 exposes the findings. Finally the conclusions are presented in Section 7.

2 Related Work

Learning through Collaborative Virtual Environments (CVEs) has been an area of increasing interest within the community of Information and Communications Technology (ICT), particularly for those researchers interested in the development of educational software based on the paradigm of Computer-Supported Collaborative Learning (CSCL) [5]. The most important advantage of these applications lies in the communication, coordination and cooperation achieved through teamwork. The aforementioned type of interaction through virtual environments derives from an area of research known as computer-supported cooperative work (CSCW) [6].

CVEs can be understood as the result of the convergence of research in the fields of virtual environments (VEs) and computer-supported cooperative work (CSCW). Such software systems are developed with the purpose of promoting communication, coordination, cooperation and, in the ideal case, collaboration during teamwork [7]. Table 1 describes some characteristics of some existing CVEs.

Table 1. Characteristics of CVEs related to this study.

CVE	Application Environment	Characteristics
HabiPro [8]	Desktop	Text-based virtual environment, designed to solve programming problems in small work groups, developed in Java.
Jazz [9]	Desktop	Research project from IBM that focused on the incorporation of collaborative capabilities into an application development environment, developed in Java using a plugin that is incorporated into Eclipse.
SÁBATO [10]	Web	Tool used in teaching algorithms and programming in engineering, oriented more toward a teaching environment.
EclipseGavab [11]	Desktop	Tool with collaborative characteristics that enables the implementation of project-based learning, developed in Java using a plugin that is incorporated into Eclipse, with support for Pascal, C and Java.
VPL [12]	Web	Tool that is integrated with Moodle, it is not mentioned the programming language which was developed.

The CVEs shown in Table 1 were developed individually, they provide experience with and the use of a software tool, and most are desktop applications. With regard to their evaluation, HabiPro [8] included an experiment in which students were required to solve problems. With regard to Jazz [9], its various benefits were discussed from a contextual collaborative perspective. SÁBATO [10] was used as part of an algorithmic and programming undergraduate course where students solved algorithmic related problems through the use of this CVE. EclipseGavab [11] was compared with other tools, such as Turbo Pascal. Currently it is a project from the Universidad Rey Juan Carlos located in Spain. VPL [12] was used in an algorithms and programming course, and at the end of the course, the results of using and not using the tool were evaluated. In general, it is observed that these proposals are used for coding activities under a collaborative approach and not for testing purpose.

3 Characteristics of the developed CVE

A general structure of the CVE developed [13] and used in this study is described below. We incorporated expert system capabilities into the CVE to support students in detecting defects in a series of instrumented programs. This architecture can be understood as a layered architecture [14] for a system whose primary users are students. The students can use the CVE to view pieces of code from programs implemented under a selected programming paradigm (for example, the structured paradigm), and they are directed to identify defects in a series of instrumented programs. The students can perform this activity in working groups, in consultation with the professor or with an expert system (ES) integrated into the CVE that supports the identification of common defect types. At the moment, the developed CVE is intended for code inspection, we are planning to incorporate other kinds of testing methods such as black-box and white-box into the CVE.

Because of its nature and intended platform, the remainder of the design of the CVE was performed based on web engineering paradigm. For the modeling of the CVE, the UML-based Web Engineering (UWE) method was implemented; this method involves building separate models for the analysis of requirements, content, navigation, presentation and process [15]. An example of the user interface of this CVE is shown in Fig. 1.

In general, the user (student) can view instrumented codes with which he or she wishes to work. In the central part of the interface, the user can access a virtual whiteboard where he or she can share pieces of code with other students. The interface also includes a chat area for collaboration among other colleagues.

4. Context of the experiment

To evaluate the effectiveness of defect detection using the developed CVE, a controlled experiment was conducted using undergraduate students as participants, who worked with CVE (assisted by CVE) and without the CVE (in the same place and at the same time, this being the traditional way of working in software testing).

for the task of detecting defects in a series of instrumented programs. For this evaluation, the following research question was considered.

RQ1: Is the effectiveness in the detection of defects affected by the use of a CVE?

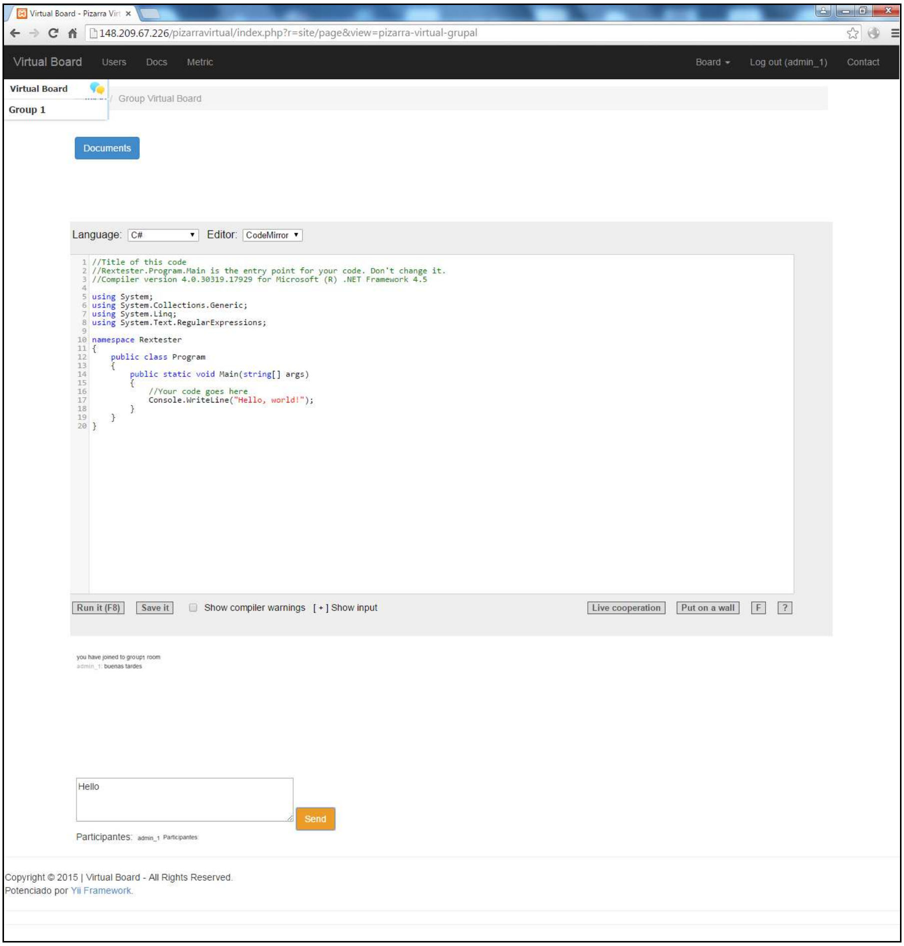


Fig. 1. Screenshot of the CVE user interface.

The above research question was translated into the following working hypotheses:

H₀. The effectiveness, measured as the percentage of observed defects, is equal for those participants who use the CVE and for those who work without the CVE (in a traditional manner).

H₁. The effectiveness, measured as the percentage of observed defects, is different for those participants who use the CVE and for those who work without the CVE (in a traditional manner).

The hypotheses presented above are tested against the measurements collected during the execution of the study. In essence, these measurements are associated with two treatment groups: the participants who worked with the CVE and those who worked without the CVE (in a traditional manner).

A crossover design was used in this study to ensure a greater number of measurements (observations) [16]. In a crossover design, the experimental units (in this case, the groups of participants, which comprised either two or three students) each receive two or more treatments, and the order of application of those treatments is determined by the design structure [16]. In this particular study, a 2 × 2 crossover design was used, involving two treatments (the use of the CVE and in a tradition manner) in two different sessions. With respect of this structure, in the first session, half of the groups of participants work with the CVE, while the other half of the groups work without the CVE (in a traditional manner); in the second session, each group of participants receives the treatment it did not receive in the first.

One shortcoming of this type of design is that the effects of certain treatments can extend beyond the period of application (carryover effect). A possible strategy for mitigating this drawback is to avoid applying the treatments in consecutive periods; in our case, a four-day break between the two sessions was planned. Table 2 summarizes the structure of the applied experimental design.

Table 2. 2 x 2 crossover design used in this study.

	Sequence 1	**Sequence 2**
Session 1 (Program 1)	With CVE	Traditional Manner
Session 2 (Program 2)	Traditional Manner	With CVE

4.1 Execution of the study

Before performing the study, the selected participants received a two-hour session of training on the use and operation of the CVE and on the types of defects that are common in programming. We also explained to participants a basic software inspection process. It was generally explained to the participants that the training session and the remaining sessions were part of a study on the CVE; verbal consent was received from the students to participate in this study.

The experiment here reported took place at the end of 2014 at the Faculty of Mathematics of the Autonomous University of Yucatan (UADY). A total of 46 students participated in this study. The participants were at the end of the first semester of their first year of the Software Engineering undergraduate program, all the participants have knowledge about key aspects of the C programming language.

Concerning the allocation of the experimental units (groups of participants) to the treatments, teams of two or three participants were formed, and these teams (18 total experimental units) were randomly allotted to one of the two sequences of treatments (with CVE - traditional manner or traditional manner - with CVE). For the recording of information regarding the observed defects, the participants used an automated coding instrument, that is, a form integrated into the CVE in which they registered the observed defects.

In two separate sessions of two hours each, the study was conducted in three rooms of the computer center of the Faculty of Mathematics at the UADY. Both the first and second sessions of the study began at the planned time. The students were given the printed specifications and the code of the program (instrumented with defects) to be inspected. Program codes were written in the C programming language. Moreover, the program source codes were available in the CVE. Each program was instrumented with seven defects. The knowledge source for the defects was obtained from [17], and the classification of defects was used from [18]. Each program had an approximate length of 150 LOC.

In the first session, one group of participant teams (teams with id 1, 2, 3, 8, 9, 13, 16 and 19) worked on the detection of the defects present in program 1 using the CVE, while the other group (teams with id 4, 5, 6, 7, 10, 12, 15 and 17) worked without using the CVE (traditional manner). In the second session, teams with id 1, 2, 3, 8, 9, 13, 16 and 19 worked on detecting the defects present in program 2 without using the CVE (traditional manner), while teams with id 4, 5, 6, 7, 10, 12, 15 and 17 worked using the CVE, that is, the treatments were allocated in the opposite manner. Participants assigned to the CVE worked each one in one computer, interacting through the CVE.

Regarding the specification of instrumented programs, program 1 (matrix.c) takes as input a square integer matrix of order *m* and performs four types of calculations on it; the following defects were introduced into this program according to the classification in [18]: two cosmetic, two control, one initialization, one omission and one computation. Program 2 (students.c) specification receives and performs basic functions over the scores of 10 students. Some of these functions are: Finding a student, update a score, compute the average of the scores, among other functions. This program contains the following defects: two computing, one cosmetic, one interface, one initialization, one data and one omission.

In each session, once the participants had completed their inspection activities of the assigned code, in addition to recording their results on printed sheets, they were also asked to register their defects observed in the CVE (for those who worked with the CVE).

5. Analysis and results

This section presents both the descriptive statistical analysis of the measurements collected and the inferential statistical analysis. Table 3 shows the effectiveness means (in terms of percentage of observed defects), standard deviations, minimum and maximum values for both treatments (with CVE and traditional manner). As shown in Table 3 the effectiveness in both treatments was similar.

Table 3. Means and standard deviations for the treatments.

Treatment	Effectiveness (%)	SD	Min.	Max
With CVE	43.45	20.03	14.28	95.28
Traditional Manner	45.33	17.78	14.28	75.71

Table 4 shows the effectiveness means, standard deviations, minimum and maximum values with respect to the instrumented programs used in the two sessions of the study. As shown in Table 4, on average, the participants seem to identify more defects in program 2.

Table 4. Means and standard deviations per session (program).

Period	Effectiveness (%)	SD	Min.	Max
Session 1 (Program 1)	37.88	16.55	14.28	75.71
Session 2 (Program 2)	50.90	18.84	14.28	95.28

Table 5 shows the effectiveness means, standard deviations, minimum and maximum values with respect to the two sequences used in this experimental design. Given the characteristics of this experimental design, there is a possibility that a carryover effect could arise and confound the real effects of the treatments. As observed in Table 5, however, the means for the two sequences appear to exhibit no substantial differences; thus, the absence of carryover effects in the treatments can be inferred.

Table 5. Means with respect to the sequences of application of the treatments.

Sequence	Effectiveness (%)	SD	Min.	Max.
Whit CVE-ad hoc	41.07	16.82	14.28	71.42
Ad hoc-CVE	47.71	20.32	14.28	95.28

Once the measurements are collected, it is possible to test the hypotheses posed above. The statistical model associated with the crossover design used is described in Equation (1).

$$y_{ijk}= \mu+\alpha_i+b_{ij}+\gamma_k+\tau_d+\lambda_c+\varepsilon_{ijk} \quad , \tag{1}$$

where μ is the general mean; α_i is the effect of the ith treatment sequence; b_{ij} is the random effect with variance σ_b^2 for the jth participant of the ith treatment sequence, γ_k is the period effect (session); τ_d is the direct effect of the treatment; λ_c is the carryover effect of the treatment administered in period k-1 of sequence group i; and ε_{ijk} is the independent random error, with an mean of zero and a variance of σ^2 for the participant in period k. In this model, the analysis of variance (ANOVA) is used to evaluate the various components of the model.

Table 6 presents the results of the ANOVA with respect to effectiveness. Effectiveness is measured as the percentage of observed defects for all the groups of participants. As observed in Table 6, only the program component (period or session effect) yielded a significant difference at an alpha level of 0.05.

The calculation of the separation measure between the treatment and the carryover effect produced a value close to 30% (29.2893%). This measure indicates the degree of orthogonality between the two effects. When a carryover effect is present in this type of design, this orthogonality is lost. A low percentage value of the separation measure suggests that there may be problems in the interpretation of the results [16]. For example, a significant difference between the treatments may be deduced when in reality, the difference is caused by the presence of a carryover effect. A 2 x 2 crossover design, such as the one used in this study, is prone to manifesting low percentages in the separation measurement. This design is recommended when the absence of carryover effects is inferred. Because no significant carryover effect was observed, the results of the ANOVA can be considered reliable. The lack of a carryover effect was reinforced by the similar effectiveness rates observed in both sequences (see Table 5).

Table 6. Analysis of variance with respect to effectiveness.

Source of Variation	Partial SS	df	MS	F	Prob > F
Treatment effect	290.22	1	290.22	0.9	0.3517
Session effect (program)	1546.17	1	1546.17	4.78	0.0373
Carryover effect	353.02	1	353.02	1.09	0.3051
Residuals	9058.20	28	323.51		

Finally, the ANOVA should satisfy the assumption of normality for reliable inferences to be obtained. We used the Shapiro-Wilk statistical test [19] in which the null hypothesis is that a given sample of data originates from a normally distributed population. Table 7 shows the results of this test with respect to the effectiveness.

Table 7. Results of the Shapiro-Wilk test.

Aspect	n	W	V	Z	Prob > z
Effectiveness	32	0.9652	1.160	0.307	0.37928

As observed in Table 7, the null hypothesis is accepted in favor of the normality of the data, that is, the measurements of each metric (aspect) are concluded to be drawn from a normal distribution; thus, our results can be considered reliable.

6. Discussion

According to the findings reported here, the results suggest an equivalence in the effectiveness of defect detection for both the groups of participants who worked in a virtual collaborative manner (with the CVE) and those who worked in a traditional manner (**RQ1**). For this type of activity (detection of defects), the obtained results suggest that the use of the CVE is equally as effective as working collaboratively following a traditional approach (ad hoc). However the kind of program used for the collaborative testing may yield different percentages of observed defects, each program to be tested may have different characteristics that as we observed may affect on the effectiveness.

Empirical studies may be exposed to threats to validity, next we discuss how we tackled them. With regard to threats to internal validity, the sessions were conducted as planned, without any disruption (history effect). Because a rest period was planned between the administration of the different treatments, a minimal effect of maturation can be assumed. The selection bias effect was reduced by randomization mechanism used in the experimental design. Although in both sessions only two experimental units (groups of participants) failed to complete the sequence of treatments (they only participated in one of the session), the presence of a low mortality effect can be considered as irrelevant. During the sessions of the study, interest in performing the activities was observed on the part of the participants; thus, a minimal effect of demoralization can be assumed. With regard to threats to the external validity of the study, the results presented here are partially generalizable to students with characteristics similar to those of the students who participated in the reported study. However, further replications of this study will alleviate this threat. We plan to conduct further replications with the aim of serving in the process of generation and consolidation of knowledge on this topic [20].

7. Conclusions

With the aim of foster software testing education by alternative means, a CVE was developed. In this work we presented the results of the assessment of such as CVE. Particularly the experiment focused on the assessment of the effectiveness of defect detection of two instrumented programs with and without the use of a CVE. Results suggest equivalent levels of effectiveness for those participants working with the CVE and for those working in a traditional way.

Some advantages of the use of the CVE are the following: 1) it is not necessary for the students to be located in the same physical space; 2) the CVE offers a virtual whiteboard feature, a component that includes an integrated IDE to write, inspect and compile pieces of code in a collaborative programming language, which is lacking in the traditional approach to learning programming and testing; and 3) the intelligent component (provided through an expert system in this proposal), which is a feature developed to facilitate the interaction between students during programming and testing training activities. Finally, the CVE assessed offers an effective means of assisting teams of students in the identification of defects in a series of instrumented programs.

As concluding remark, when collaborative work cannot be done in a traditional way, the use a CVE is an alternative approach equally effective, in this case for collaborative virtual working on software testing activities.

Acknowledgments

Authors from the Autonomous University of Yucatan (UADY) acknowledge the financial support received from the Mexican Secretary of Public Education (SEP) through project P/PFC-2016-31MSU0098J-13. We also acknowledge the support provided by UADY students participating in this study.

References

1. Chan, F., Tse, T.: Software testing education and training in hong kong. In Quality Software, 2005. (QSIC 2005). Fifth International Conference on, pp. 313-316 (2005)
2. Wong, W.E.: Improving the state of undergraduate software testing education. In 2012 ASEE Annual Conference, San Antonio, Texas (2012)
3. Wong, W., Bertolino A., Debroy V., Mathur A., Offutt J., Vouk, M.: Teaching software testing: Experiences, lessons learned and the path forward. In Software Engineering Education and Training (CSEET), 2011 24th IEEE-CS Conference on, pp. 530-534 (2011)
4. Gómez, O.S., Vegas S., Juristo, N.: Impact of CS programs on the quality of test cases generation: An empirical study. In 38th International Conference on Software Engineering ICSE'2016 Austin, Texas USA, pp. 374-383 (2016)
5. Koschmann, T.: Dewey's contribution to the foundations of CSCL research, Proc. of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community (CSCL '02), Gerry Stahl (Ed.). Int. Soc. of the Learning Sci., pp. 17-22 (2002)
6. Wilson, P.: Computer supported cooperative work: An introduction, Springer (1991)
7. Aguilar, R., Gómez, G., Moo, F., Uicab, R., Tuyub, R.: A Model Based on Intelligent Virtual Environments to Assist the Learning of Programming and Mathematics, International Conference on Education and New Learning Technologies (Edulearn 2010), Barcelona Spain, L. Gómez, D. Martí and I. Cande (Eds.), pp. 2564-2571 (2010)
8. Vizcaino, A.: Enhancing Collaborative Learning Using a Simulated Student Agent. Ph.D. Thesis, Universidad de Castilla-La Mancha, Spain (2002)
9. Hupfer, S., Cheng, L.T., Ross, S., Patterson, J.: Introducing Collaboration into an Application Development Environment, Proceedings of the 2004 ACM conference on Computer supported cooperative work, pp. 21-24 (2004)
10. Jiménez, J., Pavony, M., Álvares, A.: Entorno de integración de PBL y CSCL para la enseñanza de algoritmos y programación en ingeniería. Revista Avances en Sistemas e Informática de la Universidad Nacional de Colombia 5, pp. 189-194 (2008)
11. Gallego, M., Gortázar, F.: EclipseGavab, un entorno de desarrollo para la docencia online de la programación. XV Jornadas de Enseñanza Universitaria de la Informática (JENUI XV), Barcelona, pp. 501-508 (2009)
12. Rodríguez-del-Pino, J.C., Rubio, E., Hernández, Z.: A Virtual Programming Lab for Moodle with automatic assessment and anti-plagiarism features. Internacional Conference on e-Learning, e-Business, Entreprise Information Systems & e-Government (2012)
13. Ucán-Pech, J.P.: Aprendizaje de la Programación Asistido con Entornos Virtuales Colaborativos Inteligentes. Ph.D. Thesis. Dirección de Posgrado e Investigación de la Universidad del Sur, Campus Mérida, México (2015)
14. Bass. L., Clements, P., Kazman, R.: Software Architecture in Practice (3rd Ed.), Addison-Wesley Professional (2012)
15. Rossi, G., Pastor, O., Schwabe, D., Olsina L.: Web Engineering: Modelling and Implementing Web Applications, Springer London (2007)
16. Kuehl, R.O.: Design of experiments: statistical principles of research design and analysis, Duxbury/Thomson Learning (2000)
17. Ko, A.J.: Asking and answering questions about the causes of software behavior, Ph.D. Thesis, Carnegie Mellon University, USA (2008)
18. Basili, V.R., Selby, R.W.: Comparing the effectiveness of software testing strategies, IEEE Transactions on Software Engineering 12, pp. 1278-1296 (1987)
19. Shapiro, S.S., Wilk, M.B.: An analysis of Variance Test for Normality (Complete Samples), Biometrika 52, pp. 591-611 (1965)
20. Juristo, N., Gómez, O.S.: Replication of software engineering experiments. In Bertrand Meyer and Martin Nordio (Eds.), Empirical Software Engineering and Verification, 7007 of Lecture Notes in Computer Science, Springer-Verlag Berlin, pp. 60-88 (2011)

3D objects' shape relevance for saliency measure

Graciela Lara[1], Angélica De Antonio[2], Adriana Peña[1], Mirna Muñoz[3] and Edwin Becerra[1]

[1]CUCEI of the Universidad de Guadalajara, Av. Revolución 1500, Col. Olímpica, 44430 Guadalajara (Jalisco), México.
[2] Escuela Técnica Superior de Ingenieros Informático of the Universidad Politécnica de Madrid,
Campus de Montegancedo, 28660 (Boadilla del Monte), España.
[3]Centro de Investigación en Matemáticas, Avda. Universidad no. 222, 98068 Zacatecas, Mexico

graciela.lara@academicos.udg.mx, angelica@fi.upm.es, adriana.pena@cucei.udg.mx, mirna.munoz@cimat.mx, edwin.becerra@cucei.udg.mx

Abstract. The shape of an object is one of the features that attracts viewers' attention, making it salient from other objects in a scenario. However, the shape of an object might be linked to the viewer's personal experiences. In order to verify if the shape of an object is a feature not linked to the meaning that the object has for the viewer, in this paper we replicate an empirical study, using abstract object. 23 male and 17 female ordered objects according to the most attractive shapes. Using a computational measurement of saliency of the shape of 3D objects in Virtual Reality, based on the proportion of empty and full space within its bounding box, we found that our metric matches, when abstract or non-abstract objects are evaluated.

Keywords: 3D objects, shape saliency, voxelization

1 Introduction

The visual attraction of an object is largely determined by its basic features such as color, size, shape, orientation, position, luminance or texture. The shape of a 3D object represents its exterior geometrical figure; when described, we can use some of these characteristics, such as: area, perimeter, diameter, minimum and maximum distance from the center of mass, axes, angles, and its surrounding space (i.e. bounding box). The visual attraction of these characteristics is linked to the concept of saliency.

Lahera et al. [1] explain that perceptual saliency is the automatic and subliminal process of bottom-up visual discrimination, whereby certain stimuli stand out from the perceptual field and attract attention. In other words, it is the higher-order mental process by means of which certain perceived or mentally represented objects attract the focus of attention, including thinking and behavior. In this same sense Kapur [2] considered that visual saliency is the process of the association of objects and their representation that attracts attention and captures people's thinking and behavior.

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_22

The study of the saliency has been applied by psychologists and computer vision researchers mainly to facilitate the computer recognition of objects and to locate their salient areas in images, having achieved very significant results (e.g. [3], [4], [5], [6], [7], [8]). In recent years the study of the visual saliency of the shape of an object has also attracted the researchers' attention:

Cheng et al. [9] proposed a method for locating and segmenting salient objects from the analysis of an image group. Their system is supported by meta-data, visual saliency, and shape similarity to identifying the regions of salient objects without the influence from background clutter.

Kim et al. [10] presented an algorithm that allows automatic predictions of the static pose that a person would adopt when using a given object like a chair, gym equipment, carts, bicycles, and bipedal devices. Their algorithm considered the kinematic parameters beetween a human body and the shape's surface. This method identifies salient regions, by discovering informative viewpoints and retrieving functionally-similar shapes.

Song et al. [11] developed a system that detects mesh saliency, a perceptually-based measure of the importance of a local region on a 3D surface mesh. They used global aspects by making use of spectral attributes of the mesh in a saliency map, which captures this information, unlike most existing methods which are typically based on local geometric cues. Their approach contemplates the properties of the log-Laplacian spectrum of the mesh.

Su et al. [12] proposed an architecture to recognize the shapes' rendered views independently of each other. The architecture uses the information from multiple views of a 3D shape to get a single and compact shape descriptor, offering even better recognition performance than that of a high-performance descriptor. The method uses multiple 2D projections to detect the 3D shapes contained in the scenario being analyzed at high accuracy using sketches and leverage the implicit knowledge of 3D shapes contained in their 2D views. The architecture generates a saliency map derived by 2D views of the shape.

Leifman et al. [13] presented an algorithm that locates regions of interest on 3D surfaces. Their method studies 3D shape recovering regions from local and global perspective, allowing to detect the saliency in point clouds. The surface of a 3D object is treated from its triangular mesh, which is formed by vertices and faces. The vertices that characterize the geometry in the vicinity of a vertex are stored within a vertex descriptor. The method also includes an algorithm that detects the surface extremities, which vary.

These approaches to measure saliency were designed and evaluated from different perspectives, achieving quite good results. However, none of them are applied directly to the 3D object; in all aforementioned cases the 3D model is decomposed into a set of 2D projections, maps, views or point clouds.

In contrast, we proposed in Lara et al. [14] a computational measure of saliency of the shape of 3D objects, without the need of transforming the 3D model in any way. Furthermore, our measure approach can be used in 3D objects regardless of whether they have a regular or irregular form or their position, scale or orientation. And because it does not require a complex preprocessing, its computational cost is minimal. This measure is characterized by its promptness, simplicity and efficiency.

2 A measure for saliency by shape

In [14] we proposed a measure of saliency for 3D shape in which we explored the theory of that *the flatter is the surface of an object, the less salient it is, and inversely, objects with high pointedness tend to be perceptually more salient*. With this in mind, the measure of saliency by shape is based on the central idea of the volume of a pre-voxelized object.

The proportion of empty space and full space in voxels in the bounding box of each object is calculated. For this, first the volume (size) of the bounding box in voxels is obtained through the length, width and height of the box and the size of the voxel. Then the volume in voxels of the object is subtracted from this number. The result of this operation is called "Empty space". The volume of the object in voxels corresponds to the "Full space" inside the bounding box. Figure 1 illustrates the empty space of a 3D object within its bounding box.

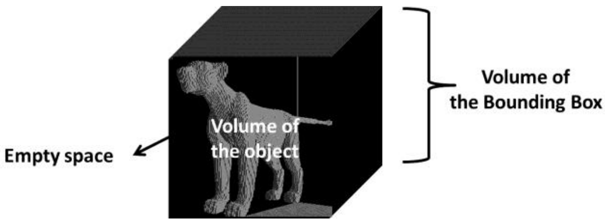


Fig. 1. Empty and full space of the bounding box of a 3D object

We proposed the following equation for the measure of shape saliency:

$$Ssh = Es \,/\, siBBox \qquad (1)$$

Where Ssh, represent the value of saliency by shape, calculated by dividing "Empty space" (Es) and the volume (size) of the bounding box in voxels (siBBox). It is the empty space proportion, which will be used as measure of shape saliency.

With the above calculation of equation, we offered a direct estimation of the saliency of an object's shape. The saliency by shape takes values between [0 - 1], being a proportion.

In [14] to validate this theory we performed empirically an experiment with one hundred 3D objects extracted randomly from the Princeton Shape Benchmark (PSB), which contains 1,814 3D models and is available to the public in the World Wide Web [15]; they included objects such as chairs, tables, vases, bottles, chandeliers, planes, and toys, among others. In that first experiment we found that our metric is comparable with the perceptual saliency of human beings.

We thought that one possibility that could have influenced the results would be the use of shapes of non-abstract objects, which can be linked to the meaning that the object has for the viewer. In order to test if there was some influence related to the type of object that captured the attraction and resulted on the viewer's value of saliency, we decided to make a replica of this experiment but with abstract objects.

3 Experimental evaluation of the shape saliency metric

In order to verify the validity of our shape's saliency measurement and to discard the possibility that non-abstract objects would have had some influence for its meaning, or for the feeling or emotion caused on the participants, we present in this work a replica of our experiment described in [14], but using abstract objects. This allowed us to see that the values given by our metric have a high degree of coincidence with the measures given by the participants, both in abstract and no-abstract objects.

3.1 Method

Participants. Forty undergraduate students of the Escuela Técnica Superior de Ingenieros Informáticos of the Universidad Politécnica de Madrid, twenty three male and seventeen female, with ages in the range of 18 to 42 years, voluntarily participated.

Materials, devices and situation. The experiment was carried out in a laboratory with suitable lighting conditions. Each participant's session was run on a SONY laptop computer, model VGN-CS270T, with a processor Intel ® Core (TM)2 Duo CPU P8600, 2.40 GHz, 4.00 GB memory, using a mouse and a keyboard. A computational application was developed to implement and test the metric using the Unity 3DTM platform, with some scripts created in C# programming language. Results were automatically recorded in a .csv (comma-separated values) file and afterwards statistically analyzed.

The voxelization process of the 3D objects was made through a master script of voxelization of the Unity 3D™ adjusted to extract the number of voxels and so calculate the volume of each object.

Design and procedure. A set of one hundred abstract objects in 3D were modeled with the Unity 3D™ tool; they were modeled using primitive three-dimensional shapes such as the cylinder, the tube, the sphere and the bucket, among others. Then these primitive shapes were edited and superimposed randomly to obtain more complex geometric shapes, see Figure 2. Likewise the set of objects for this experiment were labeled and adjusted to have the same size scale. Twenty-five scenarios were presented to participants, each with four 3D objects. Each of the objects was voxelized and the numbers of voxels was counted, the time for the voxelization process of each object varied from 0.535 to 1.368 seconds.

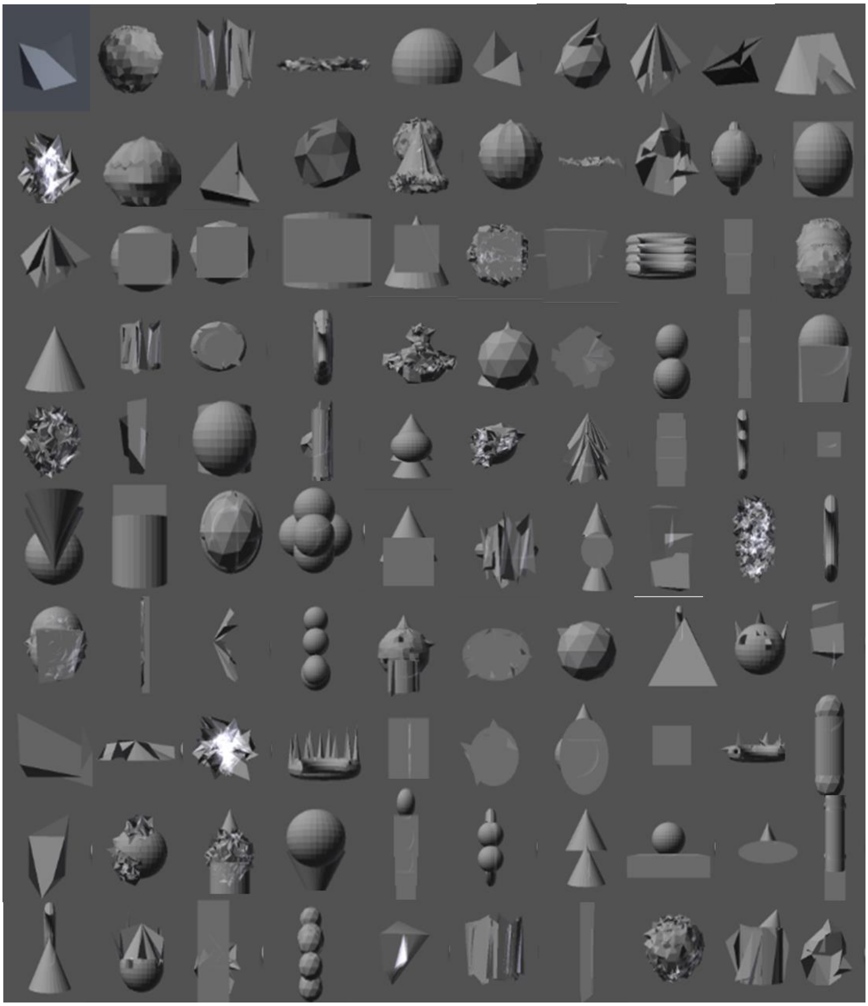


Fig. 2. View of the one hundred abstract 3D objects designed

Participants were informed that their task during the experiment was to place the four objects provided on each trial on an empty platform in front of them (see Figure 3), ordering them from left to right according to their shape saliency, this concept was explained to them as "*the capability of the object's shape to attract their attention*". Figure 3(a), in the left, shows the view of the trial to order the objects. Figure 3(b), in the right, shows the same trial with the objects ordered according to the shape saliency given by participants. Therefore, the most striking object by its shape should be placed to the left. Participants were given a brief demonstration of the system on how to place each object on the platform as shown in Figure 3. Also, we explained to participants that they could make all the necessary place changes, before they confirmed the final order of objects for each trial. Participants were asked to provide basic personal information as their, age and gender, within the system. Each person lasted about 8 minutes to complete the twenty-five trials.

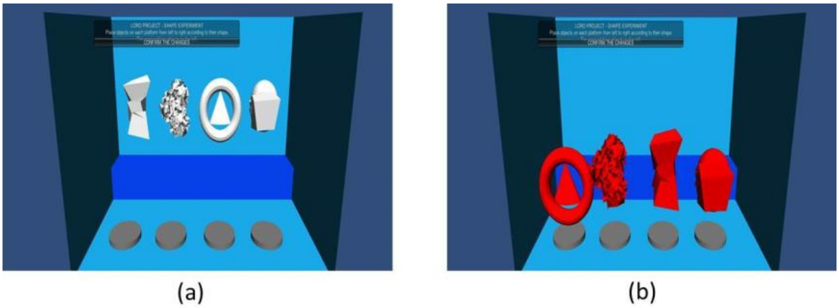


Fig. 3. Views of an experimental system trial

3.2 Statistical Analysis of Results

We performed the same statistical analyzes that in [14], with the aim of verifying in both abstract and non-abstract objects the degree of coincidence of the participants with the salience value given by our metric. We analyzed the following aspects:

1. *The choice of the first or second object as the most salient.* Because the salience of the shape of an object is a very subjective characteristic, and given that there is no precise order with which to compare the given order of our metric and each of the given orders of the participants, the comparison required to be flexible enough to admit variations, but at the same time capable to give information about the performance of the metric. We decided to measure the extent to which the first or second most salient objects, according to our metric of saliency, matched the object placed by the participants in the first position. This condition helped us to identify if the most salient objects to the human visual system corresponds with high-valued objects by our metric.
2. *Comparison of the mode with respect to the metric.* In this second statistical analysis, we compared each trial with the order given by our metric. The mode of the most voted object in each position for each trial was obtained, counting the number of times that an object was placed in each of the four positions within each trial. Then, we computed the distance between each pair of objects (the one given by the metric and the most voted one) in each position of each trial, according to the value assigned to each object by our metric.

4 Results and Discussion

The results of the first statistical analysis show that our metric effectively predicts the objects that humans tend to perceive as the most salient. The number of possible orders for the 4 objects of each trial is 24, but only half of them fit with the restriction established in our first statistical analysis, that is, to include the first or second most salient objects in the first position. Assuming random orders provided by participants for each trial, we would expect that 50% of the orders would fit our condition. However, by the actual order given by each of the 40 participants in each of the 25 trials, we got a mean of 14.3 matches per participant. These 14.3 matches represent a 57% of all possible matches, which slightly exceeds the expected 50% value by random success.

The results in the second statistical analysis showed a degree of 37% agreements between the order based on our metric of saliency and based on the mode (Table 1 '(a)') and (Table 1 '(b)') respectively. It demonstrates that our metric gives a good representation of the saliency perception in a prototypical person.

Table 1. Order of the objects in each trial: (a) based on our metric of saliency and (b) based on the mode. (When the order is the same in both, the cell is highlighted in green).

(a)

Trial	Pos_1	Pos_2	Pos_3	Pos_4
Trial_1	o14	o11	o13	o12
Trial_2	o24	o22	o23	o21
Trial_3	o31	o33	o32	o34
Trial_4	o44	o41	o43	o42
Trial_5	o54	o52	o51	o53
Trial_6	o64	o63	o62	o61
Trial_7	o74	o72	o71	o73
Trial_8	o82	o81	o83	o84
Trial_9	o92	o94	o93	o91
Trial_10	o102	o104	o103	o101
Trial_11	o112	o114	o111	o113
Trial_12	o121	o124	o123	o122
Trial_13	o131	o132	o134	o133
Trial_14	o144	o142	o141	o143
Trial_15	o153	o154	o152	o151
Trial_16	o162	o161	o163	o164
Trial_17	o172	o174	o173	o171
Trial_18	o181	o184	o183	o182
Trial_19	o192	o191	o194	o193
Trial_20	o203	o201	o202	o204
Trial_21	o214	o212	o211	o213
Trial_22	o221	o222	o224	o223
Trial_23	o234	o232	o231	o233
Trial_24	o242	o243	o244	o241
Trial_25	o252	o254	o251	o253

(b)

Trial	Pos_1	Pos_2	Pos_3	Pos_4
Trial_1	o13	o12	o11	o14
Trial_2	o24	o23	o22	o21
Trial_3	o33	o34	o32	o31
Trial_4	o43	o44	o42	o41
Trial_5	o52	o53	o51	o54
Trial_6	o61	o64	o62	o63
Trial_7	o74	o72	o71	o73
Trial_8	o83	o82	o84	o81
Trial_9	o91	o93	o94	o92
Trial_10	o102	o104	o103	o101
Trial_11	o114	o112	o111	o113
Trial_12	o123	o124	o121	o122

Trial_13	o131	o132	o133	o134
Trial_14	o142	o144	o141	o143
Trial_15	o153	o154	o151	o152
Trial_16	o162	o161	o164	o163
Trial_17	o174	o172	o171	o173
Trial_18	o181	o184	o182	o183
Trial_19	o194	o193	o191	o192
Trial_20	o201	o203	o202	o204
Trial_21	o212	o214	o213	o211
Trial_22	o221	o222	o224	o223
Trial_23	o234	o232	o231	o233
Trial_24	o243	o242	o241	o244
Trial_25	o251	o252	o253	o254

Comparing the results of the experiment in [14] and this experiment it can observed that:

a) The result of our first statistical analysis differs 2.2% with respect to the experiment presented in [14].
b) In the other hand, the percentage of the second analysis of this experiment matches exactly with the results obtained in [14].

Table 2 shows the results obtained from the two experiments performed using our salience measure for the shape of 3D objects.

Table 2. Comparisons of experiments 1 and 2

Aspect evaluated	**First experiment Princeton Shape Benchmark (PSB)**	**Second experiment Our own Benchmark**
The choice of the first or second object as the most salient	*16.5*	*14.3*
Comparison of the mode with respect to the metric	*37 %*	*37 %*

Based on these results we can assume that the shape of the non-abstract objects have not a direct influence in the value of salience given by our participants, with respect to our computational measure of saliency. Likewise, the shape has not an influence by the meaning, feeling or emotion that may cause on the participants. The results show that the shapes of abstract objects are neither better nor worse, than the shape of non-abstract objects. This allow us think that the participants' perception focuses on the processes of their attention and not on the significant value that the shape may represent in them.

With this we considered that our measure represents an approximation in the task of measuring the saliency by shape of a 3D object, without decomposition in 2D projections of the object.

The shape of an object is undoubtedly a perceptually salient feature, easily detected and almost always evident to the human being, however it is a challenge for its computational implementation.

5 Conclusions

We presented a replica of an experiment presented in [14], as well as the same statistical analyzes, but using abstract objects, with the measure of computational salience for 3D shape proposed, with the aim to compare that this measure matches in both abstract and non-abstract objects. We showed that the metric can be applied to all kinds of objects with regular and irregular geometry; it is adaptable to all scales of the 3D models, and abstract or non-abstract shapes, without causing any influence on the attention of the participants. This measure has been incorporated into a computational model of individual saliency for 3D objects, which considers the characteristics of color and size within Virtual Environments, Lara [16]. We are currently working on exploring aspects contributing to the visual saliency of 3D objects other than their shape, such as color, texture or size, with the influence of the context in which an object is placed, for a complete model of saliency for 3D objects. Considering the context a virtual scenario in which several objects with different color, size, and shape are found. An advance of this is already published in [17].

Acknowledgements. Graciela Lara holds a PROMEP scholarship in partnership with the UDG (UDG-685), Mexico. We also thank the students Adrián Calle Murillo, Roberto Mendoza Vasquez, Álvaro Iturmendi Muñoz, and Pablo Aceituno Ferro for their help in the implementation of the metric, and the experimental software application and materials.

References

1. Lahera G, Freund N, Sáin-Ruíz J: Asignación de relevancia (salience) y desregulación del sistema dopaminérgico. Elsevier Doyma Revista de Psiquiatría y Salud Mental 2013, 6(1):45 - 51.
2. Kapur S: Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry 2003, 160(1):13 - 23.
3. Hou X, Zhang L: Saliency Detection: A Spectral Residual Approach. In: Computer Vision and Pattern Recognition. Minneapolis, MN: IEEE; 2007: 1 - 8.
4. Itti L: Quantitative Modelling of Perceptual Salience at Human Eye Position. Taylor & Francis Group Psychology Press Visual Cognition 2006, 14(4/5/6/7/8):959 - 984.
5. Itti L, Koch C, Niebur E: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Transactions on Pattern Analysis and Machine Inteligence 1998, 20(11):1254 - 1259.
6. Li C, Hamza B: A multiresolution descriptor for deformable 3D shape retrieval. The Visual Computer 2013, 29(6-8):513 - 514.

7. Raubal M, Winter S: Enriching Wayfinding Instructions with Local Landmarks In: In Proceedings Second International Conference, GIScience: 2002; Boulder, CO, USA: Springer Berlin Heidelberg; 2002: 243 - 259.
8. Sampedro MJ, Blanco M, Ponte D, Leirós LI: Saliencia Perceptiva y Atención. In: La Atención (VI) Un enfoque pluridisciplinar. Edited by Añaños E, Estaún S, Teresa MM. España: Universidad Autonoma de Barcelona; 2010: 91 - 103.
9. Cheng MM, Mitra NJ, Huang X, Hu S-M: SalientShape: group saliency in image collections. The Visual Computer 2014, 30(4):443 - 453.
10. Kim VG, CHaudhuri S, Guibas L, Funkhouser T: Shape2Pose:Human-Centric Shape Analysis. ACM Transactions on Graphics 2014, 33(4):120:121 - 120:112.
11. Song R, Liu Y, Martin RR, Rosin PL: Mesh saliency via spectral processing. ACM Transactions on Graphics 2014, 33(1):1 - 17.
12. Su H, Maji S, Kalogerakis E, Learned-Miller E: Multi-view Convulational Neural Networks for 3D Shape Recognition. In: In Proceedings of the IEEE international conference on computer vision: 2015; 2015.
13. Leifman G, Shtrom E, Tal A: Surface Regions of Internet for Viewpoint Selection. IEEE transactions on pattern analysis and machine intelligence 2016, 38(12):2544 - 2556.
14. Lara G, De Antonio A, Peña A: A computational measure of saliency of the shape of 3D objects. In: 4th International Conference on Software Process Improvement: 2015; Campus of the Faculty of Computer Mazatlan (FIMAZ) of the Autonomous University of Sinaloa (UAS) at Mazatlan, Sinaloa, Mexico: Springer. Advances in Intelligent Systems and Computing.; 2015: 235 - 245.
15. Shilane P, Min P, Kazhdan M, Furkhouser T: The Princeton Shape Benchmark. In: IEEE In Proceedings of the Shape Modeling International: 2004; Washington, DC, USA: IEEE; 2004: 167 - 168.
16. Lara G, De Antonio A, Peña A: (in review) A computational model of perceptual saliency for 3D objects in virtual environment. 2016b.
17. Lara G: Computational model for the generation of directions for object location in virtual environments: spatial and perceptual aspects. Madrid, España: Universidad Politécnica de Madrid; 2016.

Towards Detecting MVC Architectural Smells

Perla Velasco-Elizondo [1], Lucero Castañeda-Calvillo [2],
Alejandro García-Fernandez [3] and Sodel Vazquez-Reyes [1]

[1] Autonomous University of Zacatecas, Zacatecas, ZAC, 98000, Mexico.
[2] Centre for Computing Research, Mexico City, 07738, Mexico.
[3] Centre for Mathematical Research, Zacatecas, ZAC, 98060, Mexico.

{pvelasco, vazquezs}@uaz.edu.mx, b160629@sagitario.cic.ipn.mx, agarciafdz@cimat.mx

Abstract. The term "bad smell" denotes a symptom of poor design or implementation that negatively impacts a software system's properties. The research community has been actively identifying the characteristics of bad smells bad smells as well as developing approaches for detecting and fixing them. However, most of these efforts focus on smells that occur at code level: little consideration is given to smells that occur at higher levels of abstraction. This paper presents an initial effort to fill this gap by contributing to (*i*) the characterization of bad smells that are relevant to the Model-View-Controller architectural style and (*ii*) assessing the feasibility of their automatic detection using text analysis techniques in five systems, implemented with the Yii Framework. The obtained results show that the defined smells exist in practice and give some insight into which of them tend to occur more frequently. Regarding the automatic detection method, results show that it exhibits good performance and accuracy.

Keywords: Software Architecture, Bad Smells, static analysis, text analysis, MVC, Yii.

1 Introduction

In Software Engineering the term *bad smell*, hereafter referred to as "smell", is used to denote a symptom of poor design or implementation that negatively impacts a software system's properties (e.g., maintainability, testability, reusability.) [1]. A smell is usually used to indicate a potential problem with software. Although not universally agreed upon, it is generally accepted that smells can occur at different levels of abstraction going from source code (e.g., long parameter list [2]) to architecture (e.g., connector envy [3]).

Smells are a common factor in the accumulation of *technical debt* [4]. Thus, detecting and fixing them becomes relevant to software system development. In recent years, the research community has been actively characterising smells (e.g. [5]) as well as developing approaches and tools for detecting and fixing them (see [6]). However, most of these efforts focus on smells that occur at lower levels of abstraction and few of them characterize, identify and fix smells at the architectural level. Additionally, in these works little consideration is given to performing these activities within the context of

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_23

architectural styles, which are design structures commonly used for software development.

The work presented in this paper is an initial effort to fill this gap by contributing to (*i*) the characterisation of a set of smells that are relevant to the Model-View-Controller (MVC) architectural style [7], which has been widely adopted for Web systems in major programming frameworks, and (*ii*) assessing the feasibility of the automatic detection of these smells in five systems implemented with Yii Framework [8] by using text analysis techniques. The obtained results show that the defined smells exist in practice and give some insight on which of them tend to occur more. Regarding the automatic detection method, results show that it exhibits good performance and accuracy.

The remainder of this paper is organized as follows. In section 2 the MVC architectural style is explained and, based on its general constraints, a characterization of related smells is defined. Next, in Section 3, an approach to the automatic detection of these smells is explained. A basic evaluation of the approach is presented in Section 4. In Section 5 a discussion of related work is presented. Finally, in Section 6, the conclusions of this work are stated as well as some lines of future work.

2 MVC Architectural Style and Related Smells

Software architecture provides a high-level model of a system in terms of components and, connectors, and properties of both each [9]. While it is possible to specify the architecture of a system using this generic vocabulary, it is better to adopt a more specialized one vocabulary when targeting architectures of a particular application domain. This specialized modelling vocabulary is known as an architectural style [9].

The MVC architectural style has been widely adopted as an architecture for the design and implementation of Web systems. Today, many popular development frameworks allow for the construction of Web systems using this style, e.g., Spring [10], Django [11], Rails [12], Laravel [13], and Yii [8]. Successful use of this style isolates business from presentation logic, which results in a system that is easier to test and maintain. Figure 1 shows a graphical representation the MVC architectural style and Table 1 describes the elements in this representation.

The MVC style defines the following general constraints:

- The Model should not deal directly with processing end-user requests. For example, its implementation should not contain $_GET, $_POST variables.
- The Model should not deal directly with the presentation of data for end-user requests. For example, its implementation should not contain HTML presentational code.

- The View should not deal directly with performing explicit access to system data. For example, its implementation should not contain code for DB queries.
- The View should not deal directly with end-user requests. For example, its implementation should not contain $_GET, $_POST variables.
- The Controller should not deal directly with performing explicit access to system data. For example its implementation should not contain code for DB queries.
- The Controller should not deal directly with the presentation of data for end-user requests. For example, its implementation should not contain HTML presentational code.

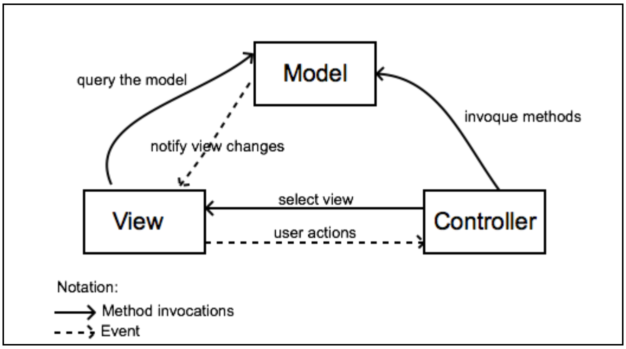


Fig. 1. Graphical representation of the MVC architectural style

2.1 MVC Smells

The concept of *architectural smell* was originally used in [14] to describe an indication of an underlying problem that occurs at a higher level of a system's abstraction than code. Causes of architectural smells include, amongst others, applying a design solution in an inappropriate context, mixing combinations of design abstractions, or applying design abstractions at the wrong level of granularity [14].

Although nearly every Web developer knows the MVC style, properly implementing it still eludes many [7]. Frequently, the general constraints, as defined in the previous section, are not respected, resulting in poor design or implementation decisions that we call *MVC Architectural Smells*. Based on these constraints, Table 2 describes a categorisation of smells relevant to the MVC architectural style.

We should note that this categorization of smells is not our own invention, but is a compilation of elements drawn from several informal sources (e.g. developer blogs, question and answer sites), which we have assembled here in a more comprehensive and consolidated manner.

Table 1. Description of the elements of the MVC architectural style

Element	Description
Model	Represents the system's underlying data and the business rules that govern data access. It notifies the View of any changes made in the data.
View	A representation of the Model in a format desired by the end users. It queries the Model for any changes made in the data.
Controller	An intermediary between the View and the Model. It receives end users' actions, commands requests coming from the View, invokes the required methods in the Model, and changes the View's presentation of the Model when necessary.

Table 2. Categorisation of smells relevant to the MVC architectural style

ID	Name	Description
1.	Model includes View's computations and/or data	Happens when the Model contains presentation of data of end-user requests (e.g. HTML code).
2.	Model includes Controller's computations and/or data	Happens when the Model has direct access to variables that represent the end-user's request (i.e. direct access to $_GET, $_POST variables).
3.	View includes Model's computations and/or data	Happens when the Controller has domain logic (e.g. code of DB queries).
4.	View includes Controller's computations and/or data	Happens when the View has direct access to variables that represent the end-user's request (i.e. direct access to $_GET, $_POST variables.).
5.	Controller includes View's computations and/or data	Happens when the Model contains presentation of data of end-user requests (e.g. HTML code).
6.	Controller includes Model's computations and/or data	Happens when the Controller has domain logic (e.g. code of DB queries).

3 MVC Code Sniffer

PHP_CodeSniffer is a static analysis tool that "sniffs" PHP code files to detect violations of a given set of rules defined in a *coding standard* [15]. It works by tokenising the contents of a code file into building blocks. These are then validated through the use of text analysis to check a variety of aspects against the coding standard in question. In this context, a coding standard can be seen as a set of conventions regulating how code must be written. These conventions often include formatting, naming, and common idioms. Multiple coding standards can be used within PHP_CodeSniffer. After the analysis process, PHP_CodeSniffer outputs a list of violations found, with corresponding error messages and line numbers.

3.1 MVC Sniff files

A coding standard in PHP_CodeSniffer consists of a collection of open-source *sniff files*. Each sniff file checks one convention of the coding standard and can be coded in PHP, JavaScript, or CSS. Thus, it is possible to create new coding standards by reusing, extending, and building new sniff files.

To detect MVC Architectural Smells, we built a *Yii MVC code standard*. The sniff files in this code standard are PHP classes, which are used in six smell detection algorithms that allow sniffing for the smells defined in Table 2.[1] Figure 2 shows an excerpt from the code of a sniff file. This sniff file, in combination with others, is utilised to detect smell number 6, which is *Controller includes Model's computations and/or data*. As previously explained, this smell is related to an issue with a Controller performing the responsibilities of a Model. Thus, in terms of code, this smell results in a Model including code to read, write, or update data or data stores, typically in a database.

As shown in Figure 2, a sniff class must implement the **`PHP_CodeSniffer_Sniff`** interface. This interface declares two functions that are needed for code analysis: the **`register`** and **`process`** functions. The **`register`** function allows a sniff to retrieve the types of token that it will process, in this case strings. Once these tokens are available, the **`process`** function is called with a representation of the code file being checked and the position in the stack where the token was found: the **`PHP_CodeSniffer_File`** and **`$stackPtr`** parameters, respectively.

Information about a token can be found through a call to the **`getTokens`** method on the code file being checked (line 13). This method then returns an array of tokens, which is indexed by the position of the token in the token stack. Tokens have a **`content`** index in the array, consisting of the content of the token as it appears in the code.

The implemented analysis detects the **`delete`** statement (line 14) in the code of a Controller file. This statement allows for the deletion of existing records in a database. If the **`delete`** statement is discovered in the code, the corresponding smell detection message is triggered (lines 17-20). A sniff indicates that an error has occurred by calling the **`addError method,`** which generates the created error message as the first argument, and the position in the stack where the occurrence was detected as the second

[1] The code standard can be downloaded from https://github.com/Lucerin/Yii.

argument, including code to uniquely identify the type of error within this sniff and an array of data used inside the error message.

```
    ...
class Yii_Sniffs_SmellsArchitectureMVC_SA6_DeleteSniff implements
PHP_CodeSniffer_Sniff {
   public function register() {
      return array(T_STRING);
   }

   public function process(PHP_CodeSniffer_File $phpcsFile, $stackPtr){
      $name = $phpcsFile->getFilename();
      $tokens = $phpcsFile->getTokens();
      if ($tokens[$stackPtr]['content'] === 'delete') {
         ...
         if (strpos($name, "controllers")) {
            $error = '6. Controller includes Model's computations and/or
                      data --> Executes a SQL statement.;
                      Delete found';
            $data  = array(trim($tokens[$stackPtr]['content']));
            $phpcsFile->addError($error, $stackPtr, 'Found', $data);
         }
      }
  }
  ...
```

Fig. 2. Excerpt of a sniff file for smell number 6: Controller includes Model's computations and/or data

When PHP_CodeSniffer is run on a Yii project using the Yii MVC code standard, there are two possible report types that can be obtained: detailed or summary. Figure 3 shows an example of a detailed report that produces a list of smells found, including corresponding error messages and line numbers. This figure shows the report produced from the analysis of a single code file: **`ConfigurationController.php`** Figure 4 illustrates a sample summary report, which gives the total of all detected smells arranged by type.

4 Evaluation

In order to assess whether the definition of the Architectural Smells was appropriate, we wanted to compare them to a "*gold standard*" (as in Information Retrieval Systems). As far as we know, there is not a gold standard for determining the number of smells introduced in a MVC system. For this reason we ran a set of basic experiments on existing systems. Specifically, five systems created by third parties and implemented with the Yii Framework were analysed to detect MVC Architectural Smells using the PHP_CodeSniffer with the defined Yii MVC code standard. The systems are public and open-source, and can be downloaded from github.

Table 3 shows the results obtained from running the experiments. Each of the analysed systems had smells. It is notable that the architectural smell that occurred most frequently was smell number 6: *Controller includes Model's computations and/or data.* The only smell with no occurrences in all analysed systems was smell number 4: *View includes Controller's computations and/or data*.

```
FILE ... protected/Controllers/ConfigurationController.php
--------------------------------------------------------------------------------
FOUND 6 ERRORS AFFECTING 6 LINES
--------------------------------------------------------------------------------
84 | ERROR | 6. Controller includes Model's computations and/or data --> Executes a
   |       | SQL statement.;
             Delete found
154| ERROR | 6. Controller includes Model's computations and/or data --> Executes a
   |       | SQL statement.;
             Count found
222| ERROR | 6. Controller includes Model's computations and/or data --> Executes a
   |       | SQL statement.;
             Delete found
244| ERROR | 6. Controller includes Model's computations and/or data --> Executes a
   |       | SQL statement.;
             Update found
264| ERROR | 6. Controller includes Model's computations and/or data --> Executes a
   |       | SQL statement.;
             Insert found
291| ERROR | 6. Controller includes Model's computations and/or data --> Executes a
   |       | SQL statement.;
             Delete found
```

Fig. 3. An example of the errors reported in the detailed report

```
CODE SMELLS SUMMARY
----------------------------------------------------
STANDARD      SMELL ID      SNIFF              COUNT
----------------------------------------------------
Yii MCV       6             Count found         91
Yii MCV       6             Delete found        50
Yii MCV       1             <div/> found        12
Yii MCV       6             Insert found         9
Yii MCV       6             Update found         6
Yii MCV       6             Select found         5
Yii MCV       6             Set found            2
Yii MCV       6             Reset found          1
----------------------------------------------------
A TOTAL OF 176 SMELLS FOUND
----------------------------------------------------
TIME: 69 MIN, 49 SEC
```

Fig. 4. Example of the summary report showing detected errors

For most of the systems, the time required for analysis was reasonably minimal. The exception was the analysis of system 1, which took 69 min, 49 sec. It should be noted, however, that the number of smells that this system had was significantly higher compared to the other systems and we believe that the time required for automatic detection is nonetheless a significant improvement compared to the complexity and challenge of detecting these smells manually.

Table 3. Results obtained by using PHP_CodeSniffer with the Yii MVC code standard.

System	LOC	Detected Smells	Most Occurring Smell	Smell with no Occurrences	Analysis Time
1. Linkbooks	66, 954	176	6	4	69 min, 49 sec
2. yii play ground	17, 341	20	6	4	10 min, 54 sec
3. Blog Bootstrap	6, 393	15	6	4	5 min, 5 sec
4. yii2-shop	5, 324	14	6	4	4 min, 37 sec
5. yii-jenkis	836	1	6	1, 2, 3, 4 and 5	1.57 sec

Having discussed the results of this basic analysis, in the next section we will cover areas of related work.

5 Related Work

There are two primary categories of related work: (*i*) code smell catalogues and (*ii*) code smell detection approaches. In this section, we relate our work to other literature on these two categories.

Code smell catalogues. Despite the range of works discussing the impact of bad smells in software architecture, very few catalogues of architectural smells have been proposed. One of these is proposed in [3] and includes four architectural smells, namely, *connector envy*, *scattered parasitic functionality*, *ambiguous interfaces*, and *extraneous adjacent connector*. In [16] the authors conducted a systematic review so as to to characterize architectural smells in the context of product lines. The authors reported a set of 14 architectural smells which, in addition to the 4 smells in [3] and SLP-specific versions of those smells (e.g. *connector envy SPL*), includes *component concern overload*, *cyclic dependency*, *overused interface*, *redundant interface*, *unwanted dependencies* and *feature concentration*. Additionally, of the few catalogues that do consider smells at the architectural level, none of these consider smells within the context of an architectural style, in sharp contrast to our work.

Code smell detection tools. In [6] the authors present the findings of a systematic literature review of 84 bad smell detection tools. Among other observations, the authors report that existing tools for analysing code concentrate on 3 main languages, namely, Java, C, and C++. They found only 4 tools for the analysis of systems coded in PHP. With regard to detection strategies, the authors discovered that most of the tools are metric-based. Less frequently used detection strategies include tree-based, text analysis, program dependence graph, machine learning, and logic meta-programming. Finally, the authors found 61 different bad smells detectable by tools, including Fowler's [1] as well as others discussed in sources such as [17], [18], [19] and [20]. In contrast with most of these described tools, the tool presented in this paper can detect smells in systems implemented in PHP. Furthermore, the smells detected by our tool are architectural and relevant to the MVC architectural style.

6 Conclusions and Future Work

In this paper, we presented our progress toward (*i*) characterising a set of smells that are relevant to the MVC architectural style, as well as (*ii*) assessing the feasibility of automatically detecting these smells. The results obtained from our experiments in automatic detection show that most of the characterised smells do exist in practice. Our use of text analysis for the purposes of smell detection is a primary focus of this paper, and experiments implementing this technique demonstrated effective, accurate results.

Although the material presented is this paper is still a work in progress, we believe that our preliminary results are valuable not only to researchers but also to developers, who may wish to begin using PHP_CodeSniffer with the defined Yii MVC code standard.

In our future work, we plan to investigate related aspects, including the identification and categorization of other smells that may occur in MVC architectures, smell detection in other coding languages, and the enhancement of tool support.

We believe that it is possible to add to our initial set of MVC Architectural Smells by incorporating smells that are, as yet, undocumented in literature. We envision a classification for MVC Architectural Smells based on dimension, such as behaviour and structure [5]. All smells detected by the technique detailed in this paper belong to the behavioural dimension.

As described in [6] most detection tools are restricted to detecting smells in specific programming languages, which is a significant limitation of existing smell detection tools. By contrast, machine learning techniques are computational methods that use "experience" to make accurate predictions. We believe that the application of machine learning techniques to the detection of architectural smells can provide performance and accuracy, as well as multi-language support, while only requiring a few sets of training examples.

Finally, with regard to enhancing tool support, we intend to explore the identification of refactoring for identified smells.

References

1. Fowler, M., Beck, K., Brant, J., Opdyke, W., Roberts, D.: Refactoring: Improving the Design of Existing Code. Addison-Wesley (1999).
2. Source Makings, Code Smells, https://sourcemaking.com/refactoring/smells
3. Garcia, J., Popescu, D., Edwards, G., Medvidovic, N.: Toward a catalogue of architectural bad smells. In 5th International Conference on the Quality of Software Architectures, pp. 146—162. Springer-Verlag, Berlin, Heidelberg (2009).
4. Brown, N., Cai, Y., Guo, Y., Kazman, R., Kim, M., Kruchten, P., Lim, E., MacCormack, A., Nord, R. L., Ozkaya, I., Sangwan, R. S., Seaman, C.B., Sullivan, K.J., Zazworka, N.: Managing technical debt in software reliant systems. In Workshop on Future of Software Engineering Research, at the 18th ACM SIGSOFT International Symposium on Foundations of Software Engineering, pp. 47--52. ACM (2010).
5. Ganesh, S.G., Sharma, T., Suryanarayana, G.: Towards a Principle-based Classification of Structural Design Smells. Journal of Object Technology, 12 (2), 1-29 (2013).
6. Fernandes, E., Oliveira, J., Vale, G. Paiva, T., Figueiredo, E.: A review-based comparative study of bad smell detection tools. In 20th International Conference on Evaluation and Assessment in Software Engineering, pp. 109--120. ACM, New York (2016).
7. Best MVC Practices, http://www.yiiframework.com/doc/guide/1.1/en/basics.best-practices
8. yiiFramework, http://www.yiiframework.com
9. Bass, L., Clements, P., Kazman, R.: Software Architecture in Practice. Addison-Wesley Professional (2012).
10. Spring, https://spring.io
11. Django, https://www.djangoproject.com
12. Rails, http://rubyonrails.org
13. Laravel, https://laravel.com
14. Lippert, M., Roock, S.: Refactoring in Large Software Projects: Performing Complex Restructurings Successfully. Wiley (2006).
15. PHP_CodeSniffer, https://pear.php.net/package/PHP_CodeSniffer
16. Vale, G., Figueiredo, E., Abílio, R., Costa, H.: Bad Smells in Software Product Lines: A Systematic Review. In *Eighth Brazilian Symposium on Software Components, Architectures and Reuse, pp.* 84--94. IEEE Computer Society, Washington, DC (2014).
17. Bavota, G., De Lucia, A., Di Penta, M., Oliveto, R., Palomba, F.: An Experimental Investigation on the Innate Relationship between Quality and Refactoring. Journal of Systems and Software, 107, 1--14 (2015).
18. Khomh, F., Vaucher, S., Guéhéneuc, Y., Sahraoui, H.: BDTEX: A GQM-based Bayesian Approach for the Detection of Antipatterns. Journal of Systems and Software, 84, 559--572 (2011).
19. Maiga, A., Ali, N., Bhattacharya, N., Sabané, A., Guéhéneuc, Y. G., Antoniol, G., Aimeur, E.: Support Vector Machines for Anti-pattern Detection. In 27th International Conference on Automated Software Engineering, pp. 278--281. IEEE Press, New York (2012).
20. Vidal, S., Marcos, C., Díaz-Pace, J.: An Approach to Prioritize Code Smells for Refactoring. Automated Software Engineering, 23, 501--532 (2014).

Information and Communication Technologies

A Brief Review on the Use of Sentiment Analysis Approaches in Social Networks

Francisco Javier Ramírez-Tinoco[1], Giner Alor-Hernández[1], José Luis Sánchez-Cervantes[2], Beatriz Alejandra Olivares-Zepahua[1], Lisbeth Rodríguez-Mazahua[1]

[1] Division of Research and Postgraduate Studies, Instituto Tecnológico de Orizaba, México
javier.rtinoco@hotmail.com, {galor, lrodriguez}@itorizaba.edu.mx, bolivares@ito-depi.edu.mx
[2] CONACYT, Instituto Tecnológico de Orizaba, México
jlsanchez@conacyt.mx

Abstract. Sentiment analysis is the study of subjective information, for example, opinions, sentiments and beliefs that people express about different topics. In recent years, its importance has grown because a large amount of this type of information has been generated daily in social networks, which can be used to obtain various benefits. There are several research works about sentiment analysis, but very few of them compare the use of sentiment analysis approaches and methods among various social networks. Therefore, the objective of this document is to provide a brief review of the most relevant works related to sentiment analysis and social networks, which shows the main findings regarding the tendencies of using the main sentiment analysis approaches, methods and aspects detected in the different social networks. The review can provide a guide for researchers to know the approaches that exist and how they were used specifically in social networks.

Keywords: Facebook, Sentiment Analysis, Sentiment Analysis Approaches, Social Networks, Twitter.

1 Introduction

Nowadays an enormous amount of subjective information is generated in different social networks such as Facebook®, Twitter®, Sina®, among some others, because they are the means by which people express their opinions, sentiments or beliefs daily. Moreover, this information is of great interest to companies, organizations or individuals because it can generate great benefits, such as: a) knowing the level of satisfaction of a customer with a product or service and taking actions if it is not good, b) finding out how the public opinion is related to some political party and planning better campaigns or strategies based on this, c) adjusting the content of a website regarding the interests that its users demonstrate, to mention but a few.

A field of study that has taken great relevance to analyze subjective information is sentiment analysis (SA). According to Ravi and Ravi [1] the sentiment analysis or opinion mining is the study of opinions, sentiments, beliefs and attitudes that people express about different topics. This analysis involves tasks as detection, extraction and classification of sentiments from different resources in various formats, such as

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_24

discussion forums, blogs, social networks and news [2]. In particular, this paper is focused on the sentiment analysis in social networks texts because most of the subjective information is generated in these media and it is of textual type.

It is important to mention that there are different sentiment analysis APIs (Application Programming Interface) such as PreCeive®, Toneapi®, Semantria®, among others [3–5], which facilitate the development of applications and systems. Furthermore, there are different sentiment analysis approaches, such as machine learning, lexicon-based and hybrid approach. Machine learning is based on a set of data to construct models that try to correctly classify new or unknown data [6]. The lexicon-based approach uses dictionaries that contain a broad collection of terms, phrases and expressions to infer the sentiment conveyed in a text [7]. Finally, the hybrid approach combines the lexicon-based approach techniques with the machine learning techniques in order to take advantage of the characteristics of each one [6].

Moreover, there are different works consisting of researches, systematic reviews and comparative analyzes related to sentiment analysis that are focused on existing applications, algorithms, approaches and tools [1, 6–14]. However, very few of them compare the use of sentiment analysis approaches and methods among various social networks. Therefore, the purpose of this paper is to provide a brief review of sentiment analysis works that are focused in social networks, with the aim of finding tendencies related to the use of the main sentiment analysis approaches, methods to determine the sentiment and the most important aspects detected in the different social networks. This review can guide researchers to select the most appropriate approach according to the work that they intend to do.

The structure of this document consists of six sections. The second Section describes the research methodology that was followed in this work. The third Section shows a classification of the research works according to the social network in which they are focused. The fourth Section describes the main findings discovered after the classification process. The fifth Section discusses the conclusions and future work.

2 Research Methodology

The research methodology that was followed in this work consists of three stages. The first stage is focused on an investigation of works that are related to sentiment analysis in social networks. This research was carried out in digital scientific databases, such as: 1) ACM Digital Library; 2) IEEE Xplore Digital Library; 3) ScienceDirect (Elsevier); 4) Springer-Link; 5) Wiley Library, and 6) Taylor & Francis, which were considered in this research because they are the largest databases of electronic journals. In the second stage, a classification of these works was performed based on the social network in which they are focused. Finally, the third stage shows a report of the review of the most relevant works that identifies the findings regarding the use of sentiment analysis in social networks, such as: the main sentiment analysis approaches, the algorithms associated with these and the sentiment aspects detected.

It is important to mention that this research did not pretend to be exhaustive and was based on the use of keywords and coincidences by title, which had the purpose of limit the results and select the most relevant works. Keywords used in this study were:

1) Sentiment analysis social networks; 2) Sentiment analysis twitter, and 3) Sentiment analysis Facebook. The first keyword was selected with the aim that the search was not limited to any particular social network; however, a search related to Facebook® and Twitter® was performed given their importance as the most popular social networks currently. Besides, only the works published by academic journals and the proceedings of international congresses were considered. Research works that were not directly related to sentiment analysis in social networks, that were not published, that were not written in English, that were textbooks, master or doctoral dissertations, or whose date of publication was not between 2010 and 2017 were discarded.

The selection process resulted in 229 papers that were carefully reviewed and classified. The classification process was carried out manually, taking as classification criteria the social network on which the sentiment analysis was applied. In addition, in order to perform the detailed review of the works during the third stage, the most relevant works were chosen considering as selection criteria the number of works that have cited them and the social network in which they are focused. Fig. 1 shows the different stages of the research methodology described above.

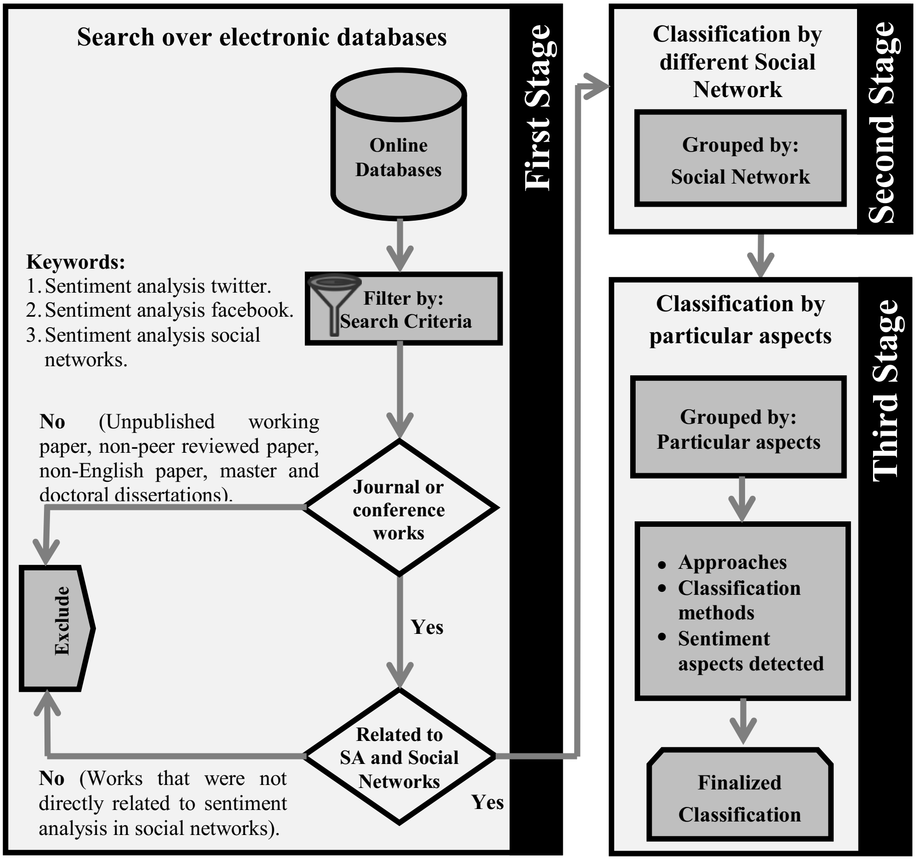


Fig. 1. Stages of the research methodology.

3 Classification of Research Papers

This section describes the classification of documents based on the methodology previously proposed. The classification performed shows the most used social networks for sentiment analysis and aims to provide to other researchers with the guidelines for future investigations related to sentiment analysis in social networks, as well as to serve as a guide for professionals interested in the development of tools, algorithms, approaches, frameworks, among others, within this area.

Documents found after the search process that was carried out, were distributed among the different digital scientific databases. According to this distribution, the highest number of articles published using the search parameters established belongs to IEEE, which is distinguished by containing works related to Computer Science, Electrical and Electronics Engineering. Also, it is important to mention that in the different digital scientific databases, most of the works related to this research were found through the keywords: *sentiment analysis twitter*.

There are several works on sentiment analysis that pretend to obtain different benefits in diverse fields, such as: education, marketing, health, sports, politics, among some others. Social networks are increasingly important for sentiment analysis because they are the medium in which millions of people express their sentiments, emotions and opinions. Besides, there are several social networks around the world and while some of them disappear, others begin to emerge. Table 1 shows the distribution of research works by social network and editorial. Twitter® is the dominant social network in sentiment analysis for the characteristics that it has, which are described below, as well as some of the most relevant works in each social network.

Table 1. Distribution of research works by Social Network and Editorial

Social Network	*Editorial*					*Total*
	Elsevier	*IEEE*	*ACM*	*Springer*	*Others*	
Twitter®	13	101	32	45	5	196
Facebook®	3	14	5	3	-	25
YouTube®	-	-	1	1	-	2
Sina®	-	-	1	1	-	2
Myspace®	-	-	2	-	-	2
Digg®	-	-	2	-	-	2
Total	16	115	43	50	5	**229**

3.1 Twitter®

Twitter® is a social network that allows its users to send and read short messages that are known as tweets, which do not exceed a length of more than 140 characters. In addition, it provides an API that allows access to its functionalities in a simple way [15]. Some authors argue that the sentiment analysis on Twitter® is complex because of the short length of its messages, since it makes difficult to identify the context and causes an abundance of abbreviations and spelling mistakes [16]. However, most consider that these challenges are solved through a pre-processing of the text and that the short messages allow identifying with more precision the sentiment because it is more likely that a smaller number of sentences and entities exists.

Twitter® was used by Wang et al. [17], who proposed a system for analyzing the sentiment of the public about the 2012 United States presidential candidates in real time. The system uses machine learning and offers to the public, the media, politicians and other stakeholders a new and timely perspective on the dynamics of the electoral process and public opinion. Moreover, a new approach that adds semantics as an additional feature to perform sentiment analysis on Twitter® was introduced by Saif et al. [18]. The approach consisted of adding the semantic concept of each entity extracted by each tweet. After evaluating this approach with a machine learning algorithm on three datasets, an average top accuracy was obtained compared to other known approaches. Salas-Zárate et al. [19] proposed a method to detect satirical and non-satirical tweets. The method pre-process tweets in order to clean and correct them, extracts psychological and linguistic features and trains machine learning algorithms such as BayesNet and Decision Trees. It was tested using a corpus of tweets in Spanish and it obtained good results, with an F-measure of up to 85.5%.

Wang et al. [20] proposed a new graph model to automatically generate the polarity of a hashtag based on: 1) The sentimental polarity of the tweet that contains it; 2) The co-relationship between hashtags, and 3) Its literal meaning. It was shown that the performance of some machine learning algorithms improves using a classification adjustment that employs the literal meaning of hashtags as semi-supervised information. In another work, Ghiassi et al. [21] developed a specific lexicon for sentiment analysis on Twitter® and demonstrated through sentiment classification models that the proposed lexicon was significantly more effective than traditional lexicons in terms of memory and accuracy. Finally, Tan et al. [22] showed that the information about relationships between users improves sentiment analysis. This work considered that the users related to each other are more likely to have similar opinions, so some induced models of Twitter® of the network of followers and the network formed by the users through mentions were proposed. The results revealed that the incorporation of this information leads to a higher accuracy in the classification of sentiments compared to other known approaches based on SVM (Support Vector Machine).

3.2 Facebook®

Facebook® is a social network that allows its users to create a profile that represents them, upload photos and videos. In addition, users have the possibility to view the profile of other users, add them as friends, comment and exchange text messages. It also offers an API that provides functionalities for writing and reading data of this social network [15].

Ortigosa et al. [23] developed a method for sentiment analysis on Facebook® that extracts information about the sentimental polarity of users' messages and detects important emotional changes. This method was launched in a Facebook® application that classifies messages according to their polarity using a hybrid approach and displays the results through an interactive interface with an accuracy of 83.27%. In addition, it was used in an e-learning context, so it is useful to recommend appropriate activities for each student according to their emotional state and feedback to teachers about what students feel about their course. In another work, Ngoc and Yoo [24] proposed a content-based classification method that considers the polarity of comments and the user participation in Facebook®. A lexicon-based approach was used and it was concluded that the method is more accurate than other existing methods thanks to it considers the user's perspective, which refers to the comments polarity.

Terrana et al. [25] analyzed the home pages of users and groups of Facebook® in order to detect through machine learning the sentimental polarity of the messages shared. Also, it was possible to create graphs that emphasized the sentimental concord between the publications and the comments. Finally, Troussas et al. [26] described how the sentiment analysis helps in language learning by stimulating the educational process, for which they conducted experiments that consisted of finding the sentimental polarity of Facebook® publications through machine learning. It was concluded that the classifier Naïve Bayes was the one that obtained the best results.

3.3 Sina®, YouTube®, Myspace® and Digg®

Sina® is one of the most influential social networks in China. It allows its users to share information with other people through text, images and other forms. The length of the text is limited to 140 characters [27]. YouTube® [28] is the leading video hosting service. It allows its users to share and manage their own videos, as well as search, view and comment on the videos of other users. Digg® is one of the most popular websites to share and discuss news and ideas [29]. Myspace® is a social network that allows its users to create a profile, share photos, incorporate music or videos to their profiles, among other things. In 2008 it was very popular; however, it has been overcome by other social networks and its use has been decreasing year by year.

Zhao et al. [27] proposed a lexicon-based sentiment analysis algorithm on the content of Sina®. The algorithm achieved a very good accuracy but its efficiency was poor, so an optimization of the algorithm was performed and the problem was overcome. In addition, a real time sentiment analysis platform was developed based on this algorithm. In another work, Hammad and Al-awadi [28] described a work where machine learning techniques such as SVM, neural networks, Naïve Bayes and decision trees were applied in order to find a light approach for sentiment analysis in YouTube® reviews and comments written in Arabic. The results showed that SVM achieved the highest accuracy (96.06%) compared to the other classifiers.

Furthermore, Paltoglou and Thelwall [29] proposed an intuitive, not supervised and lexicon-based approach that estimates the level of emotional intensity contained in text from Twitter®, Myspace® and Digg®, with the purpose of making predictions. The approach allows detecting polarity and subjectivity, and after comparing it with known supervised approaches, the proposed approach overcame to the others in most cases. Finally, Gonçalves et al. [30] conducted a study comparing eight popular sentiment analysis methods on texts from YouTube®, Myspace® and Digg®. In addition, they developed a hybrid method that combines different approaches and a free Web service called iFeel that provides an API to access and compare the results between different sentiment analysis methods for a given text.

4 Findings

One finding regarding sentiment analysis APIs is that although they can facilitate and expedite the use of sentiment analysis, they are rarely used. Of the different papers analyzed, only 9.68% used an API, while 90.32% chose other alternatives such as using some existing classification algorithms through data mining tools, to develop own algorithms, to use known applications, among others, which was done mainly to find new alternatives that obtain greater precision to detect sentiments.

Moreover, another important finding is that within the sentiment analysis approaches, the most used is machine learning. This approach was used in 58.87% of the cases, since it is argued that in practice it tends to achieve higher classification accuracy and does not require lexical resources that are sometimes difficult to obtain or build. Then, the lexicon-based approach is found with 29.84% of use. Those who used this approach believe that it can achieve good results and overcome challenges of machine learning such as the dependency on the domain and labeled data. More lagging, with 9.68% of utilization the hybrid approach is found, which has achieved good results but it is little used because it implies taking on the challenges of machine learning and lexicon-based approach. Finally, the ontology-based approach reached 1.61% of utilization. This approach has showed that it allows obtaining good results; however, it is the least used because it involves a greater effort than in other cases. The distribution of the use of sentiment analysis approaches is shown in Fig. 2.

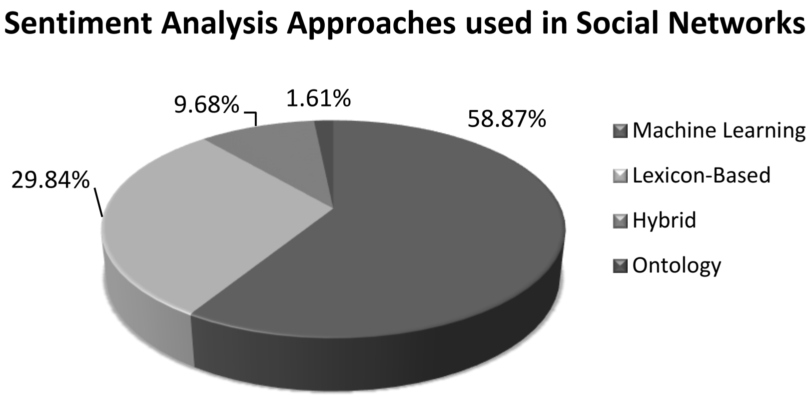


Fig. 2. Distribution of Sentiment Analysis approaches used in social networks.

Moreover, since machine learning is the most used approach, it is important to mention that the works concerning to this approach were based on the use of different methods of classification, regression and clustering to determine the sentiments of numerous texts. In addition, the texts in social networks have peculiar characteristics compared to other texts (abundance of abbreviations, spelling mistakes, and character limitations, among others) that affect the accuracy of such methods. Thus, many of the papers try to find the methods to determine the sentiment with more precision. Therefore, the following localized finding is shown in Fig. 3, which is related to the percentage of use of the main methods to determine the sentiment of texts from social networks among the analyzed works.

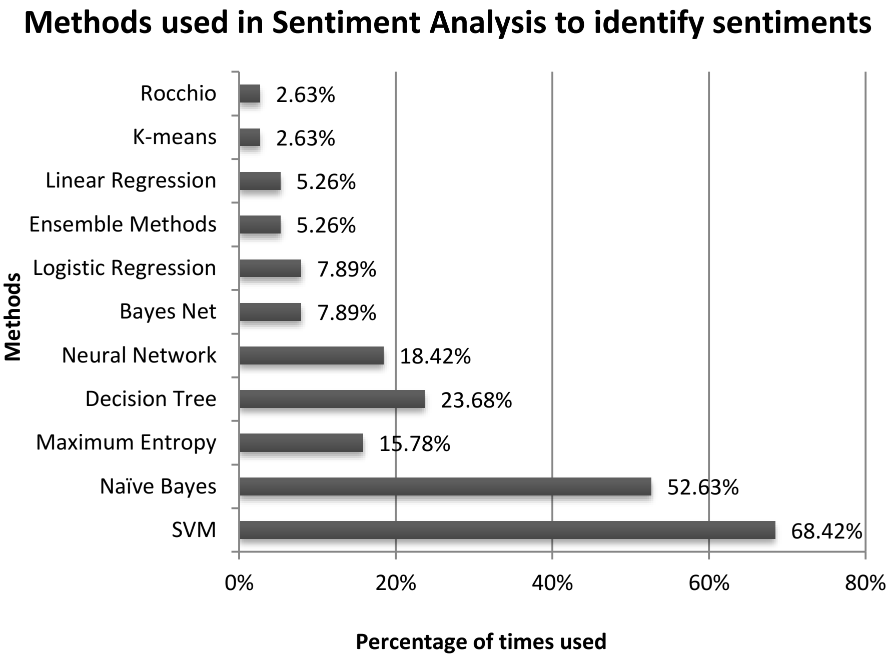


Fig. 3. Machine Learning methods used in sentiment Analysis in Social Networks.

Classification methods generally overcame in the percentage of use to the regression and clustering methods, since the sentiment analysis is considered by many researches as a problem of classification of texts, nevertheless, it is interesting to know within the classification methods, which are the preferred by researchers considering the results obtained with them. According to Fig. 5, these methods are: SVM (68.42%), Naïve Bayes (52.63%), Decision Trees (23.68%) and Neural Network (18.42%), since most authors agree that through them, the best results are obtained.

Besides, the analyzed works considered different aspects that reveal information to determine the sentiment, such as: the subjectivity of the text, the sentimental polarity and intensity, the detection of emotions, sarcasm and negation. However, not all aspects were investigated with the same regularity and importance. Therefore, another finding obtained in this research is that the works tend to focus only on the detection of polarity, since this was considered in the 96.77% of the works, while the other aspects were considered in very few cases. Fig. 4 shows the finding in greater detail.

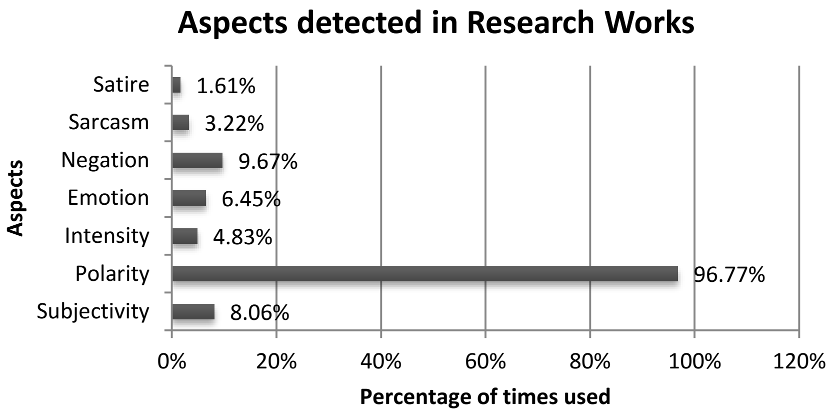


Fig. 4. Main aspects detected in research works.

5 Conclusions and future work

This paper provides a review of the most relevant works related to the use of the sentiment analysis approaches in social networks. After investigating and analyzing these works, some important findings were gleaned. The first one is that the percentage of use of the sentiment analysis APIs is very low compared to the use of other alternatives such as methods or algorithms used independently; however, it would be interesting compare the results obtained by these alternatives and by APIs. Also, it was identified that machine learning is the most used approach given the high accuracy that it usually obtains. However, it is domain dependent and it needs labeled datasets, which implies a lot of effort in an environment such as social networks, in which many domains are involved. Thus, it is necessary to explore other approaches that achieve good results with less effort and resources.

Moreover, it was observed that the works tend to focus on the sentimental polarity, while leaving aside other aspects such as intensity, negation or emotions. It is important consider these aspects, because this would achieve a deeper and detailed analysis that allows to make better decisions, to create better business strategies, among others benefits. In a similar way, the most used methods to determine the sentiment of texts in the context of social networks were identified. SVM and Naïve Bayes were the most used; however, it is worth exploring with greater depth other alternatives in rise, such as Ensemble methods, which have gotten encouraging results.

As future work, we are considering to expand and improve the research from three perspectives. First, it is important to extend the number of papers analyzed including other types of documents such as master and doctoral thesis. In second place, we will evaluate aspects such as: 1) Lexical resources used in the lexicon-based approach, and 2) Feature extraction methods used in machine learning, such as bag of words, n-grams, among others. Finally, we pretend to make a classification based specifically on aspects that the works detect with greater importance and frequency, such as sarcasm, irony, satire, among others.

Acknowledgments. This work was supported by Tecnológico Nacional de Mexico (TecNM) and sponsored by the National Council of Science and Technology (CONACYT) and the Secretariat of Public Education (SEP) through PRODEP (Programa para el Desarrollo Profesional Docente).

6 References

1. Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Syst. 89, 14–46 (2015).
2. Toxtli-Hernández, C.: Computación sentimental, https://sg.com.mx/sgvirtual/sesion/computacion-sentimental#.WJLk5RvhA2w
3. TheySay: PreCeive, http://www.theysay.io/product/preceive/
4. Adoreboard: Toneapi, http://toneapi.com/
5. Lexalytics: Semantria, https://www.lexalytics.com/semantria
6. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: A survey. Ain Shams Eng. J. 5, 1093–1113 (2014).
7. Serrano-Guerrero, J., Olivas, J.A., Romero, F.P., Herrera-Viedma, E.: Sentiment analysis: A review and comparative analysis of web services. Inf. Sci. (Ny). 311, 18–38 (2015).
8. Hussein, D.M.E.-D.M.: A survey on sentiment analysis challenges. J. King Saud Univ. - Eng. Sci. (2016).
9. Schouten, K., Frasincar, F.: Survey on Aspect-Level Sentiment Analysis. IEEE Trans. Knowl. Data Eng. 28, 813–830 (2016).
10. Silva, N.F.F. Da, Coletta, L.F.S., Hruschka, E.R.: A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning. ACM Comput. Surv. 49, 15:1--15:26 (2016).
11. Giachanou, A., Crestani, F.: Like It or Not: A Survey of Twitter Sentiment Analysis Methods. ACM Comput. Surv. 49, 28:1--28:41 (2016).
12. Korayem, M., Aljadda, K., Crandall, D.: Sentiment/subjectivity analysis survey for languages other than English. Soc. Netw. Anal. Min. 6, 75 (2016).
13. Singh, J., Singh, G., Singh, R.: A review of sentiment analysis techniques for opinionated web text. CSI Trans. ICT. 4, 241–247 (2016).
14. del Pilar Salas-Zárate, M., López-López, E., Valencia-García, R., Aussenac-Gilles, N., Almela, Á., Alor-Hernández, G.: A study on LIWC categories for opinion mining in Spanish reviews. J. Inf. Sci. 40, 749–760 (2014).
15. Van Anh, T.T., Dau, H.X.: A Crossed-domain Sentiment Analysis System for the Discovery of Current Careers from Social Networks. In: Proceedings of the Fifth Symposium on Information and Communication Technology. pp. 226–231. ACM, New York, NY, USA (2014)
16. Neethu, M.S., Rajasree, R.: Sentiment analysis in twitter using machine learning techniques. In: 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). pp. 1–5 (2013)
17. Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S.: A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. In: Proceedings of the ACL 2012 System Demonstrations. pp. 115–120. Association for Computational Linguistics, Stroudsburg, PA, USA (2012)

18. Saif, H., He, Y., Alani, H.: Semantic Sentiment Analysis of Twitter. In: Proceedings of the 11th International Conference on The Semantic Web - Volume Part I. pp. 508–524. Springer-Verlag, Berlin, Heidelberg (2012)
19. Salas-Zárate, M. del P., Paredes-Valverde, M.A., Rodriguez-García, M.Á., Valencia-García, R., Alor-Hernández, G.: Automatic detection of satire in Twitter: A psycholinguistic-based approach. Knowledge-Based Syst. 128, 20–33 (2017).
20. Wang, X., Wei, F., Liu, X., Zhou, M., Zhang, M.: Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management. pp. 1031–1040. ACM, New York, NY, USA (2011)
21. Ghiassi, M., Skinner, J., Zimbra, D.: Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. Expert Syst. Appl. 40, 6266–6282 (2013).
22. Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., Li, P.: User-level Sentiment Analysis Incorporating Social Networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1397–1405. ACM, New York, NY, USA (2011)
23. Ortigosa, A., Martín, J.M., Carro, R.M.: Sentiment analysis in Facebook and its application to e-learning. Comput. Human Behav. 31, 527–541 (2014).
24. Ngoc, P.T., Yoo, M.: The lexicon-based sentiment analysis for fan page ranking in Facebook. In: The International Conference on Information Networking 2014 (ICOIN2014). pp. 444–448 (2014)
25. Terrana, D., Augello, A., Pilato, G.: Facebook Users Relationships Analysis Based on Sentiment Classification. In: 2014 IEEE International Conference on Semantic Computing. pp. 290–296 (2014)
26. Troussas, C., Virvou, M., Espinosa, K.J., Llaguno, K., Caro, J.: Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. In: IISA 2013. pp. 1–6 (2013)
27. Zhao, Y., Niu, K., He, Z., Lin, J., Wang, X.: Text Sentiment Analysis Algorithm Optimization and Platform Development in Social Network. In: Proceedings of the 2013 Sixth International Symposium on Computational Intelligence and Design - Volume 01. pp. 410–413. IEEE Computer Society, Washington, DC, USA (2013)
28. Hammad, M., Al-awadi, M.: Sentiment Analysis for Arabic Reviews in Social Networks Using Machine Learning. In: Latifi, S. (ed.) Information Technology: New Generations: 13th International Conference on Information Technology. pp. 131–139. Springer International Publishing, Cham (2016)
29. Paltoglou, G., Thelwall, M.: Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media. ACM Trans. Intell. Syst. Technol. 3, 1–19 (2012).
30. Gonçalves, P., Araújo, M., Benevenuto, F., Cha, M.: Comparing and Combining Sentiment Analysis Methods. In: Proceedings of the First ACM Conference on Online Social Networks. pp. 27–38. ACM, New York, NY, USA (2013)

Impact of organizational and user factors on the acceptance and use of project management software in the medium-sized company in Lima

[1]César Aguilera, [2]María Teresa Villalobos and [1]Abraham Dávila

[1]Departamento de Ingeniería, Pontificia Universidad Católica del Perú, Lima 32, Lima, Perú
[2]Departamento de Ciencias, Pontificia Universidad Católica del Perú, Lima 32, Lima, Perú
{cesar.aguilera, mtvillalobosa, abraham.davila}@pucp.edu.pe

Abstract. The use of software applications usually contribute in most times the contexts where they are applied, however their adoption is not always effective. In particular, concerning the project management, there are some difficulti -es in the context of medium-sized companies in Lima to use project manage -ment tools for the projects they manage. In this research our objective is to study the factors related to the organization and users for the acceptance of management tools. A survey was designed and performed based on a previ -ous and similar study. A response was obtained from 77 managers of med -ium-sized projects with emphasis on the factors analyzed. Factors such as functionality, organizational size, project complexity and soft -ware use are the most representative. The study also found: (i) a strong and significant relationship between the use of software and the performan -ce perception of the project manager and (ii) factors related to experience, train -ing and education level have no effect.

Keywords: project management software, information technology acceptation, technology acceptance model.

1 Introduction

Organizations are relying more and more on project management to achieve initiatives [1]. Economic figures gathered internationally indicate a significant and growing use of project management [2]. Huemann, Turner, and Keegan in 2004 reported that we have become a project-oriented society, and Anantatmula [3] in 2008, citing other studies, estimated annual spending on projects to be in the billions of dollars in the global economy.

In the nature of the projects can be identified different characteristics such as duration or complexity, and due to these factors; some of them influence the determination of using project management software [3]. The project management software has been incorporated during the last years, much functionality that facilitates the work of a project manager or project management team. Some tools or techniques such as WBS (work breakdown structure), CPM (critical path method) and PERT (Project Evaluation and Review Techniques) are common features in any project management software [4]. In addition to other features such as its functionality

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_25

through the use of internet, collaborative work and documentation attached to the project [5].

The involvement of Information Technology (IT) areas has been important in incorporating these tools into project management applications which has enabled project managers to maximize the use of project management methods [3]. However, the main problem is that organizational and user factors that influence the acceptance of project management software are not known; this main problem is the effect of little or no understanding of the process of adopting new technology and that the common characteristics and sociological factors of the users are also unknown.

The structure of the document is organized in such a way that the following section: Section 2, mentions the models that have traditionally been used to explain the adoption of IT at the individual level; in Section 3, research protocol is described; in Section 4 survey results is presented; in Section 5, hypotheses are tested; in Section 6, a final consideration, conclusion and future work are established.

2 Background

The adoption of IT has been extensively studied to find theoretical foundations, factors, roles and organizational structures. All this research has been necessary because the adoption of technology is a complex and social process involving contextual, emotional and cognitive factors [5]. Next, the literature is described on some of the theories that have allowed studying factors of adoption of technology:

- The Behavior theory [5] is used to predict whether a person intends to do something, researchers need to know: (i) whether the person is in favor of doing it ('attitude'); (ii) how much the person feels social pressure to do it ('subjective norm'); and (iii) whether the person feels in control of the action in question ('perceived behavioral control').
- The Theory of Reasoned Action (TRA) [6, 7, 8] is a general model of prediction of human behavior originally introduced by Fishbein in 1967 and then improved, developed and tested by Fishbein and Ajzen in 1975, which defines the relationships between belief, attitude, norm, intention, and behavior.
- The Theory of Planned Behavior (TPB) [9] is a model very similar to the TRA, except that an additional construct is taken into account: perceived behavioral control (PBC), which refers to perception control over the performance of a given behavior.
- Technological Acceptance Model (TAM) [7] was developed by Fred Davis in 1985, where he proposed that the use of a system is a response that can be explained by the motivation of the user, which in turn is directly influenced by an external stimulus on the part of the characteristics and capacities of the real system

The most widely used models, from the perspective where individuals are highly rational and make their decisions to maximize value or utility [10], and therefore become the main source of innovation and organizational changes, are: The TAM, TPB, and the Unified Theory of Acceptance and Use of Technology (UTAUT) [11].

Since the adoption of information technologies presupposes a new use of these technologies and systems or the introduction of them into an organization, IT adoption models are based on innovation models, in the sense that the latter implies a perceived novelty within the organization, such as an idea, artifact, or practice [12].

Project management software was initially developed in the 1960s and 1970s to run on large computers. More than 500 project management software packages were developed in the 1990s, with a wide variety of prices and capabilities [4]. Currently most of the development of software is done on cloud basis, the web based software development trend grows up very quickly [13].

The study by Ali, Anbari, and Money [14] represents a significant advance towards the understanding of the factors that affect the use of information technology in project management.

This research on the impact of organizational and user factors on the acceptance and use of project management software in the medium-sized company in Lima seeks to replicate and expand the study conducted by Ali, Anbari, and Money [9] in the Peruvian context. The factors selected in [9] were drawn from the literature of information systems and project management performance impact and level of use, project management software characteristics, computer self-efficacy, task characteristic and management support.

3 Research protocol

In this Section we present the research protocol based on hypotheses and methodology followed for this study.

3.1 Methodology

In this study, we used a survey as research protocol. The survey design's follow phases were established in [15]: plan (pln), realization (rlz) and report (rpt). Then the activities were developed: (pln-1) set the survey objectives, (pln-2) design the survey, (pln-3) develop the questionnaire, (pln-4) evaluate and validate the questionnaire, (rlz-1) get the survey data, (rlz-2) analyze the data obtained and (rpt-1) report the results.

The instrument used for evaluation was translated from English into Spanish and the questions reviewed to facilitate its compression. Subsequently some items were modified to increase their clarity. The final instrument was constructed using a 5-point Likert scale (1 Almost never, 2 Sometimes, 3 Normally, 4 Almost always, 5 Always). A web survey was used as means of data collection.

Due to the exploratory nature of the study, a non-probabilistic sampling was chosen for convenience, having as reference the project managers of medium-sized IT areas in the city of Metropolitan Lima.

3.1 Hypotheses

In Fig. 1, it is presented the research model examined in this study. The research model assesses the extent of project management software usage by project professionals and the impact of its use on their perceived performance [14].

The model posits that project management software acceptance is a function of perceived information quality, software functionality, ease of use, project complexity, project size, organization size, user training level, education level, and experience. The model also proposes that the use of the project management software has a direct and positive impact on the user's perceived performance [14]. The research model proposes the following hypotheses:

- H1: The perceived ease of use has a positive relationship with the use of project management software
- H2: The perceived functionality has a positive relationship with the use of project management software
- H3: The perceived information quality has a positive relationship with the use of project management software
- H4: Organization size has a positive relationship with the use of Project management software.
- H5: Project size has a positive relationship with the use of project management software.
- H6: Project complexity has a positive relationship with the use of project management software
- H7: Level of project manager training has a positive relationship with project management software usage
- H8: Level of project manager experience has a positive relationship with project management software usage
- H9: Level of project manager education has a positive relationship with project management software usage
- H10: The use of project management software has a positive relationship with project manager performance

4 Survey Results

The data were collected from an independent sample composed of project managers obtained from the researcher's contacts, from the contact database in Peru of the company Dharma Consulting and from a contact list provided by the Section Computer Engineering of the Pontifical Catholic University of Peru.

108 project management professionals participated in the survey (completed) obtaining 77 responses that indicated that they used some type of project management software which represents 70% of the sample the remaining 30% did not use any tools.

Table 1 shows the level of education of the research sample, where more than 76.9% have a professional degree, a master's or doctoral postgraduate degree, and the rest of the participants had a bachelor's degree, had completed university or had

Technical instruction. In addition, 49.4% of the participants have more than five years in the field of project management and 48.0% of them have more than five years of experience using project management software. Table 2 shows the data regarding the age of the participants, where 95.4% of them are in the range of 20 to 50 years of age.

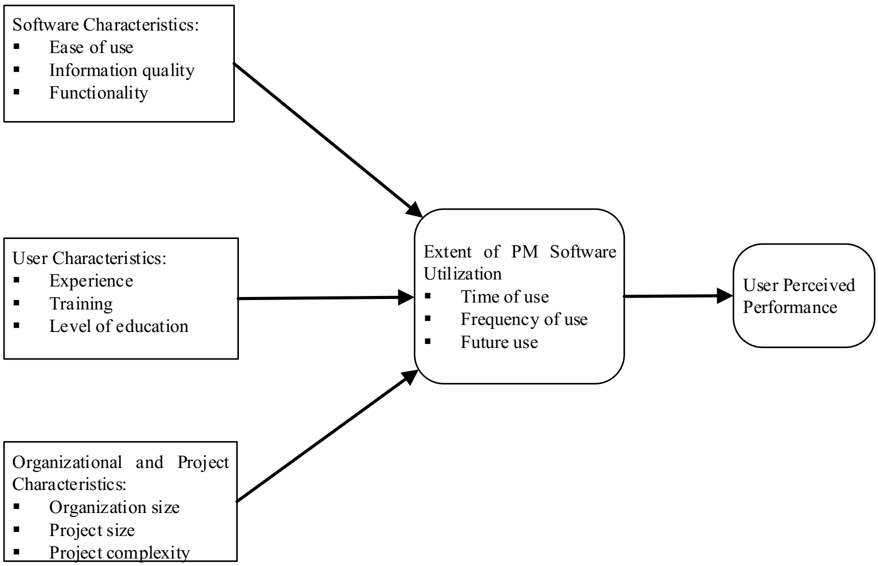


Fig 1. Research model

Table 1. Level of education

		Frequency	Percent	Cumulative percentage
Valid	Doctorate	3	2,8	2,8
	Magister	39	36,1	38,9
	Titled	41	38,0	76,9
	Bachelor	18	16,7	93,6
	College graduate	4	3,7	97,3
	Technical instruction	3	2,8	100,0
	Total	108	100,0	

Table 2. Age of participants

		Frequency	Percent	Cumulative percentage
Valid	20-30	28	25,9	25,9
	31-40	53	49,1	75,0
	41-50	22	20,4	95,4
	51-60	5	4,6	100,0
	Total	108	100,0	

The reliability analysis of the measuring instrument was omitted because it was taken from a previous investigation. However, the high values of Cronbach's alpha obtained in each of the sections indicate and confirm the consistency of the questions. As a first step a Factor Analysis was performed, which included a Varimax rotation for each of the sections of the research questions questionnaire. Then, the factors for each of the sections of the survey were identified: performance, use, software features, self-efficiency, task characteristics and management support. To determine the number of factors to be retained, the Kaiser criterion was applied, that is, to retain those factors with an eigenvalue greater than 1. Once the factors were obtained, the regression analysis of the factors identified versus the variables of use was carried out. It was done as in the original research.

5 Testing the Research Hypotheses

The obtained factors are shown in the Table 3. By analyzing variances of ANOVA, we intend to validate the model and establish the values of the β factors that compose it. The detailed analysis of each of the results obtained for each of the hypotheses is summarized in the Table 4. Five of the ten research hypotheses were supported, and another five were not supported.

Table 3. Summary of factors

	Factor	**Variable**
Performance	F1	Performance
	F2	Performance of the management
Degree of Use	F1	Knowledge area (Extent of use)
	F2	Complex Projects [-]
	F3	Process Groups [-] (Extent of use)
	F4	Small Projects
	F5	Use
Software Features	F1	Quality of information
	F2	Integration
	F3	Functionality
	F4	Easy to use [-]
Self-efficacy	F1	With support
	F2	Unsupported [-]
Characteristics of the task or Complexity of the project	F1	Routine tasks
	F2	Tasks established
	F3	Interdependent tasks [-]
	F4	Equivocality
Management support	F1	Support from Top Management

Compared to the original research study, we have the following results:

- Hypothesis 1 (H1) was not supported indicating that a statistically significant relationship (β = 3.4%, $p < 0.772$, $R2 = 0.12$) was not found. In addition the variable ease of use was in the opposite direction. This is an unexpected result;

the project managers participating in this study do not consider that the ease of use of the project management software is a variable in the decision to use it, This finding is contradicted by a survey conducted by The Access Group in 2013 which states that the most sought after features in project management software are reliability, ease of integration and ease of use [16].

Table 4. Result of the model hypotheses

Hypothesis	Resulted	p-valor	R2
H1	Not Supported	0.772	0.120
H2	Supported	0.000	16.50
H3	Supported	0.001	15.44
H4	Partially Supported	0.058	5.05
H5	Not Supported	--	--
H6	Supported	0.004	11.14
H7	Not Supported	--	--
H8	Not Supported	--	--
H9	Not Supported	--	--
H10	Supported	0.000	24.57

- Hypothesis 2 (H2), can confirm that for the participants of this study it is of more relevance that the software fulfills the requested functionality before the satisfaction in the use. The 2007 PwC survey indicates that "the most commonly used features are resources and milestones" [17]. Then in 2015, Capterra's "Project Management User Research Report" survey indicates that the most important factor in choosing the purchase of project management software is functionality (40%) followed by ease of use (24%) [18].
- Hypothesis 3 (H3) is strongly supported. However, in a survey called "Insights and Trends: Current Program and Project Management Practices" [17] mentions that project management reports are not generated using project management software. In fact, the data suggest that the most frequently reported project management area, cost reporting, is mainly done outside of project management software. Only 23% of cost reports are generated using project management software. However, Capterra's survey in 2015 points out that "one of the major business challenges that drives people to use project management software is to capture project time and costs (62%)" [18]. As a consequence, it can be interpreted as the need to be able to have information on these two aspects in the projects tools.
- Hypothesis 4 (H4) is partially supported by the research. According to this statement we have that PwC in a survey called "Insights and Trends: Current Program and Project Management Practices" [17], indicates that 77% of the participating companies use project management software. From this set, those organizations that have a higher maturity level are significantly more likely to use project management software: 95% of organizations of the highest maturity category use project management software, compared to only 55% within the category of lower maturity level. In an earlier report by PwC itself [19], it is mentioned that: "If the organization's maturity level is low, software installation will create problems and influence the performance of your project. Once the

organization reaches a certain level of maturity, where project management processes are institutionalized, the use of software will significantly increase the overall performance of the project".

- Hypothesis 5 (H5) is not supported. In the survey conducted by PwC [17], it is mentioned that: "We found that software use is frequent - 77% of companies use project management software, while 23% does not. Software tools are used more frequently to manage individual projects compared to multiple projects or programs". No reference to project size.
- Hypothesis 6 (H6) is supported. The participants in this study identified that the complexity of the project is directly related to the use of project management software.
- Hypothesis 7 (H7) is not supported by the present research work. In a survey conducted by PwC [17], they found a correlation between project management certification (PMI and Prince 2 certifications) and project performance. Contrary to what we found in our research. However, in the same PwC study, it was found that the use of project management software is linked to the high performance of the project. 77% of the participating companies use project management software. In a previous report made by the same company, in 2004, the same correlation was found with a level of use of 87%.
- Hypotheses 8 (H8) and 9 (H9) are not supported. No correlation was found regarding the experience and level of education of the project manager and the use of project management software.
- Hypotheses 10 (H10). About H10, it was found that the use of project management software has a positive and statistically significant relationship with the performance perceived by a project manager, due to the values obtained (β = 50.6%, $p < 0.0001$, $R2 = 24.57$). In other words, the use of project management software accounts for 24.57% of the variation in performance perceived by a project manager.

  Based on the regression analysis obtained, the impact of software use on the performance perception can be calculated using the following regression equation (1):

$$\text{Perception of performance} = 0.019 + 0.506 * \text{Use of SW of PM} + \varepsilon. \qquad \textbf{(1)}$$

For each unit of increase in the use of project management software, the perception of performance of the users increases in 50.6% of unit.

In Table 5 we can find the results found for each of the hypotheses in both researches.

According to a study by Straub [20] it is argued that the predictions of TAM will not necessarily be maintained between cultures. This indicates that countries differ in cultural terms. Hofstede's research [21] on cultural dimensions provides a theoretical basis for exploring the impact of cultural differences in the adoption and diffusion of IT-based innovations such as e-mail.

Hofstede [21] describes four dimensions that can be used to distinguish between different cultures: power-distance, uncertainty avoidance, masculinity, and individualism. Table 6 lists the four dimensions and a brief description of each.

Table 5. Hypotheses contrasted with original research

Hypothesis	Original study result	Research result
H1	Strongly Supported	Not Supported
H2	Strongly Supported	Supported
H3	Strongly Supported	Supported
H4	Moderately Supported	Partially Supported
H5	Partially Supported	Not Supported
H6	Strongly Supported	Supported
H7	Not Supported	Not Supported
H8	Partially Supported	Not Supported
H9	Moderately Supported	Not Supported
H10	Strongly Supported	Supported

6 Discussion, conclusions and future work

Since this was an initial research in the application of the TAM model in medium-sized enterprise project management software in the city of Lima, there were several limitations. It is very likely that the results and conclusions of this research will be limited to the sample, variables and time frame represented by the research design.

Table 6. The four cultural dimensions of Hofstede

Hofstede Dimension	Abbr.	Description
Power- distance	PDI	Degree of inequality among people which the population of a culture considers normal
Uncertainty avoidance	UAI	Degree to which people in a culture feel uncomfortable with uncertainty and ambiguity
Individualism	IDV	Degree to which people in a culture prefer to act as individuals rather than as members of groups
Masculinity	MAS	Degree to which values like assertiveness, performance, success, and competition prevail among people of a culture over gentler values like the quality of life, maintaining warm personal relationships, service, care for the weak, etc.

The conclusions should be considered attempts until other further investigations can confirm or reject similar finds. In addition, these findings are limited to the heads of medium-sized enterprise projects in metropolitan Lima.

The characteristics of the sample make it clear that the respondents of this research cannot be seen as a reflection of the population in general. Therefore, the results of this study cannot be extrapolated since the sample may not be strongly representative of the national population of those who manage projects.

For future studies, it is proposed to extend the proposed model and look for other variables that are considered important, related to the characteristics of the software as with the characteristics of the user. Another aspect to study would be to measure the correlation between the variables studied by a bivariate correlation test.

This research enriches the understanding of the factors affecting the use of information technologies in project management. In the study, the use of project management software explains 24.57% of the variation in performance perceived by a project manager.

However, there are some marked differences in the results of our research with respect to the initial study. It was not possible to validate that the following variables have a direct relation with the use of project management software, among them we have: the size of the project, the training in project management, the level of experience in project management, the level of academic education and the perception of software ease of use.

From all these unconfirmed factors, three are specific to the project manager (training in project management, level of experience and level of academic training). The non-confirmation of these three factors may be due to cultural terms, as Straub [20] puts it in his study called "Testing the technology acceptance model across cultures: A three country study", where it is pointed out that cultural dimensions provide a foundation to explore the impact of cultural differences in the adoption and diffusion of IT-based innovations [21].

It was also confirmed with the H4 hypothesis that the size of the organization has a direct relationship with the use of the software. This is supported by other literature that mentions that organizations with a higher level of maturity are significantly more likely to use project management software [17].

On the other hand, the results obtained indicate that the sociological factors do not influence the acceptance of the project management software. In the opinion of the project managers, it seems clear that professional experience and training cannot be replaced, but software can help improve organizational skills, refine the scheduling of tasks and gain a better understanding of how and why things go well or wrong in projects.

Acknowledgements

This work has been carried out within the ProCalProSer project funded by Innóvate Perú under Contract 210 - FINCYT-IA-2013 and by the Department of Engineering and the Software Engineering Research and Development Group (GIDIS) of the Pontifical Catholic University of Peru.

References

[1] T. Creasy and V. S. Anantatmula, "From Every Direction How Personality Traits and Dimensions of Project Managers Can Conceptually Affect Project Success," *Project Management Journal,* vol. 44, no. 6, pp. 36-51, 2013.

[2] S. Davis, "Investigating the impact of project managers' emotional intelligence on their interpersonal competence," *Project Management Journal,* vol. 42, no. 4, pp. 37-57, 2011.

[3] V. S. Anantatmula, "The Role of Technology in the Project Manager Performance Model," *Project Management Journal,* pp. 34-48, 2008.

[4] J. R. Meredith and S. J. Mantel, Project management: A managerial approach, New York: Wiley., 2006.

[5] E. T. Straub, "Understanding Technology Adoption: Theory and Future Directions for Informal Learning," *Review of Educational Research,* pp. 625-649, 2009.

[6] I. Ajzen and M. Fishbein, Understanding Attitude and Predicting Social Behavior, New Jersey: Prentice-Hall, Inc., 1980.

[7] F. D. Davis, *A Technology Acceptance Model for Empiricall Testing New End-user Information Systems: Theory ans Results.,* Cambriage, MA: MIT Sloan School of Management, 1985.

[8] A. Dillon and M. G. Morris, "User Acceptance of information technology: Theories and Models," *Annual Review of Information Science and Technology (ARIST),* pp. 31-32, 1996.

[9] I. Ajzen, "From intentions to actions: A theory of planned behavior.," *Action-control: From cognition to behavior,* pp. 11-39, 1985.

[10] C. Slappendel, "Perspectives On Innovation In Organizations," *Organization Studies,* vol. 17, no. 1, pp. 107-129, 1996.

[11] T. Olivera and M. Fraga Martins, "Literature Review of Information Technology Adoption Models at Firm Level," *The Electronic Journal Information Systems Evaluation,* vol. 14, no. 1, pp. 110-121, 2011.

[12] E. Rogers, Diffusion of innovations, 5th ed., New York: Simon and Schuster Inc, 2003.

[13] M. Sajad, M. Sadiq, K. Naveed and M. Iqbal, "Software Project Management: Tools assessment, Comparison and suggestions for future development," *International Journal of Computer Science and Network,* vol. 16, no. 1, pp. 31-42, 2016.

[14] A. S. B. Ali, F. T. Anbari and W. H. Money, "Impact of Organizational and Project Factors on Acceptance and Usage of Project Management Software and Perceived Project Success," *Project Management Journal,* pp. 5-33, 2008.

[15] B. A. Kitchenham and S. L. Pfleeger, Guide to Advanced Empirical Software Engineering, J. S. D. I. K. S. Forrest Shull, Ed., Springer, 2008.

[16] The Access Group, "Inbox Insight Survey 2013," 2013.

[17] PricewaterhouseCoopers, "Insights and Trends: Current Programme and Project Management Practices," 2007.

[18] Capterra, "Project Management User Research Report," 2015.

[19] PricewaterhouseCoopers, "Boosting Business Performance through Programme and Project Management," 2004.

[20] D. Straub, "Testing the Technology Acceptance Model Across Culture: A Three Country Study," *Information & Management,* vol. 33, no. 1, pp. 1-11, 1997.

[21] G. Hofstede, Culture's Consequences: International Differences in Work-Related Values, 1980.

Appendix

https://drive.google.com/open?id=0B_K_llz5juqPaEZOb0xoYUVqWDA

A Lean mind-set on the Information Technologies sector: Targeting and addressing waste for an increased performance

Jorge F. Guedes

Business Research Unit – BRU-IUL
ISCTE - Instituto Universitário de Lisboa
Lisbon, Portugal
Jorge_Fernando_Guedes@iscte-iul.pt

Abstract. Performance and efficiency are two topics raking high on the agendas of information technologies (IT) top managers across the globe. In order to address these challenges some lessons learned from other sectors started to be incorporated in IT, namely a Leaner mind-set imported from manufacturing. This paper aims to better understand if the adoption of a Lean philosophy would benefit IT in regards to its performance as well as to validate some of the currently available Lean IT recommendations. The author analyzed opportunities for improvement identified in a case study by conducting unstructured interviews and by leaving some comments based on his observations. These finding are then applied to Lean IT recommendations available in the literature, clustering them against the adapted concept of waste. Finally, the author suggests that a Lean mind-set could offer positive outcomes and recommends further research that could add extensive value to practitioners and academics.

Keywords: Lean IT; Lean Manufacturing; Lean Transformation; Information Technologies; Project Management.

1 Introduction

Although the Information Technology (IT) sector has improved considerably over the past decades, researches and practitioners across the globe still struggle in order to increase its efficiency and performance. To be noted that according to the CHAOS report conducted by the Standish Group [1] and considering data until 2015, we can perceive that only 29% of IT projects succeed in being completed within time and budget, with 52% of them being considerably challenged and 19% of them simply failing. These figures suggest the urgent need to adapt new practices or even entirely new mind-sets to address these challenges – In this sense, the adaption of a Lean philosophy could be an interesting path for IT organizations across the globe, since it already proved to be valuable in several other sectors. Although Lean is better known in manufacturing, the same benefits can be harvested in the IT world. It is actually suggested by some authors that manufacturing is similar to other sectors such as

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_26

banks and insurance companies. It is therefore not surprising all are now looking to manufacturing to learn about Lean thinking [2]. It is the author's strong belief that Information Technologies should follow the same suggestion.

The following paper aims to better understand if a Lean mind-set would benefit IT organizations, as well as to provide some preliminary validation of the overlooked Lean IT concepts, which can still be considered a gap in the literature. In the form of a case study, the author analyses the methodologies and processes in use in an IT department of a large European Bank, with a special interest in opportunities for improvement. These findings are then cross-checked with Lean IT concepts in an effort to understand their applicability by practitioners. Finally, the author leaves some final considerations on the topic and suggests some paths for further research

2 From manufacturing to IT – an introduction to Lean IT

Lean concepts have been on managers' lexicons since the early 90's, after Womack, Jones and Roos [3], published the bestselling book "The Machine that Changed the World". This publication generalized Lean, showcasing several concepts and techniques that promised very interesting performance improvements for organizations, tying together "many of the seemingly disparate principles that had caught the attention of researchers and practitioners alike" [4]. This introduction of a corporate philosophy centered on Lean principles showed great potential, especially on manufacturing industries – it is not surprising to acknowledge that the same principles started to be applied on several other sectors, with equally promising results. Yet, giving a short definition of Lean is not an easy task.

We can first define Lean as being a mind-set that discards any action that does not increase value from a customer perspective. Yet, it can also be seen as a management style or practice that constantly tries to understand the "why", making an effort to always involve all employees in a very people centric way. It can also be considered an approach that inspires reengineering of processes and promotes continuous improvement, targeting a better overall performance. Finally, we can even see it as a tool that allows and encourages the visibility of performance. In a nutshell, Lean Manufacturing can be seen as a philosophy that decreases time from order to completion to the end customer, eliminating any potential sources of waste as suggested by Liker [5]. Additionally, and recognizing the importance that waste definition has on the Lean mind-set, we can describe waste as something that represents any excess interruption, misalignment, unnecessary work, or ingrained redundancies that add no value [6]. Actually, the idea of constantly detecting and eliminating waste can be considered as one of Lean's mantras, being that several authors [7] summarize waste in 7 types, also known as the 7 "MUDA": Excessive production, Waiting, Transport, Over-Processing, Inventory, Reprocessing and Defects. Although waste detection and elimination is a very important topic for Lean thinkers, as well as the central piece of this paper, it is important to highlight that the real benefit of Lean is not only eliminating waste. In fact, it can be considered to be the overall strengthening of the system as suggested by Meier and Forrester [8].

Additionally, Lean should be seen as a path, rather than an end point as suggested by Karlson and Ashlstrom [4].

This path can also be taken by IT organizations, importing some of these concepts and philosophies from manufacturing and other sectors. It might seem like a difficult road to cross, but let's have in mind that IT processes and procedures can be mapped, measured and managed – meaning that they can be optimized. Riding on this idea, some efforts to adjust Lean philosophies to IT started to grow all across the globe, being the pioneer book of Mary and Tom Poppendieck [9] "Lean Software Development: An Agile Toolkit for Software Development Managers" a very good example of these efforts. Still, the topic was greatly overlooked by the academia, generating more interest on practitioners than researchers. A good example of this gap in the literature is the adaption of the waste concept to IT, a topic not so easy to find properly validated on specialized academic literature, but already discussed by some practitioners as DOWNTIME. In the IT case, the concept of DOWNTIME can be seen as a simple adaption of the 7 MUDA previously mentioned on this paper, as suggested by Peter Waterhouse [10]. These 8 sources of waste on IT are summarized on Table 1, below:

Table 1. Waste factors in IT (DOWNTIME)

Waste Factor	Example of waste	Example of impacts
D - Defects	Technical errors or miss alignment with requirements	Lack of customer focus (considering internal and external customers in organizations), increase in costs and in time to completion
O - Overproduction	Developments that will not be used, that will not be fully used or that will be used in a non-optimal way	Increase in costs, increase in complexity
W - Waiting	Long waiting times between activities, for instance between defect correction, transport and re-testing	Low performance, lack of customer focus, increase in costs, increase in employee frustration, increase in time to completion
N - Non-Value added processing	Applications that do not deliver relevant value to the organization or to the final customers	Miss-communication, increase in business complexity and increase in costs
T - Transportation	Transports for problem solving, non-optimal environment and client strategy	Increase in costs, lower efficiency, increase in time to completion and increase in employee frustration
I - Inventory	Non adjusted IT sizing, non-needed licensing	Increase in costs and lack of efficiency
M - Motion	Re-active way of working, instead of a pro-active and planned way – also known by practitioners as "firefighting"	Increase in costs, lower productivity, increase in employee frustration
E - Employee knowledge	Inability to record knowledge, non-effective or non-existent	Lowers employee engagement and satisfaction, risk of losing

	knowledge transfer or learning activities	knowledge on the organization

This simple mapping of waste sources, adapted to IT, can be used as a quick guide to target and address performance pain points on an IT organization, being very valuable for practitioners in continuous improvement initiatives. Still, the actual value it can deliver to organizations could be further validated from an academic perspective, adding value to the community and increasing the confidence on the topic. The following research paper aims to take one step further in this validation, sustaining the findings with the research methodology detailed on the following chapter.

3 Research Methodology

In order to better understand if the adoption of a philosophy centred in Lean concepts would benefit an IT organization in regards to its performance as well as to validate some of the currently available Lean IT recommendations, the author analysed 5 months of IT services delivered by a large consultancy and outsourcing firm based in Lisbon, Portugal, to the local headquarters of a large European Bank, in Lisbon, Portugal. These 23 weeks under analysis consider more than 5500 hours of work delivered by a total of 26 employees, all working at the bank premises. To be noted that not all the consultants were working full time, being that some of them were just brought in for very specific activities, being rolled off once completed. This data was collected in 2010/2011 and unpublished so far due to confidentiality agreements, being the author's belief that the findings are still relevant for academics and practitioners.

The author starts by analysing the efforts consumed per project phase of all the projects that took place in this period, cross-checking the results with available literature. Additionally, the author conducted 3 unstructured interviews to main stakeholders from the vendor side participating on the projects in order to understand their main lessons learned from this account. These stakeholders were carefully chosen considering their visibility of the overall positioning of the IT department on the organization, their knowledge of the methodologies and processes in place, their deep understanding of the activities being completed by the teams as well as their understanding of the challenges that the teams faced while delivering the services. Based on the findings, on a second interview with the selected stakeholders, the author will group the results with the main sources of waste identified on Lean IT literature, creating a clustering between sources of waste and opportunities for improvement, leaving some comments related to its applicability and recognizing its relevance to academics and practitioners. Finally, the author draws some conclusions based on the findings and highlights some threats to the validity of the study.

4 Opportunities for improvement from the case at study

Based on the author's empirical knowledge and observations, the organization at study revealed a low to medium maturity level despite its efforts to adapt best practices and its healthy and constant eagerness to improve. The first observation was that the methodology being used on the projects was not only not clear and communicated to the teams, but also overlooked from a conceptual perspective, especially from a resourcing estimation and effort allocation point of view. To be noted that this is a typical indicator of performance issues, considering that poor effort allocation is one of the main root causes of rework due to insufficiently or badly resourced activities [11]. In order to better understand this opportunity for improvement, the author analyzed all the efforts employed on the projects at study by the vendor, considering 23 weeks of activity – out of the total hours analyzed (more than 5500), the author validated that 37% of the efforts were management activities, 11% of the efforts were dedicated to requirements analysis and design, 41% were related to technical design, development and unit testing, and the remainder 11% were dedicated to SW integration and further testing. For the study at hand the percentage of effort for management activities will not be further analyzed, although it already suggests some lack of management efficiency. Further, and disregarding the management efforts, we can consider that 17.5% of efforts were consumed for requirements analysis and design, 65% were related to technical design, development and unit testing, and the remainder 17.5% were dedicated to SW integration and further testing.

In order to better understand the deviation of this effort consumption from industry best practices, the author compared the collected data with the COCOMO II model [12], which suggests the allocation of effort for new software development as 20% for requirements analysis and design, 57% for detail design, code and unit test, and 23% for SW integration and test. The results of this analysis are on Table 2, below:

Table 2. Case Study efforts against COCOMO II

Project Phase	Efforts from Case Study	COCOMO II	Deviation (Case Study VS COCOMO II)
Requirements analysis and design	17.5%	20.0%	+14.3%
Detail design, code and unit test	65.0%	57.0%	-12.3%
SW integration and test	17.5%	23.0%	+31.4%
Relative Effort	100.0%	100.0%	0%

To be noted that in all phases we have a considerable deviation, being the deviation on the phase of detail design, code and unit test the largest. Considering the projects environment, this deviation can point to the lack of commitment to the blueprinting phase, since it is a clear indication of lack of direction on the developments that needed to be completed, creating the need to take extensive iterations between teams for functionality clarification. This suggests that a better usage of the methodology, considering its processes and quality gates, would be an opportunity for improvement that could have a direct impact on the performance of the teams.

Additionally, in order to further understand opportunities for performance improvement on this organization the author conducted unstructured interviews to 3 main stakeholders from the vendor side (Project Lead, Project Manager and Business Advisor). Based on the outcome from these interviews, the author was able to identify 6 main opportunities for improvement:

1. Improving and enhancing development environments - Currently, the data present in the Development Environment is outdated and incoherent, which leads to several errors during the testing phase. Although the software development and testing process is rigorous and demanding, it is also worth highlighting the absence of a pre-production environment as well as the right usage of a quality environment - the software created under development, if approved in the Change Advisory Board, often goes directly to production, creating a massive risk on operations and over stressing the support organization. This practice can potentiate problems and lack of quality in production, leading to defects and lowering the overall efficiency, being a clear opportunity for improvement

2. Better management of project planning and expectations - Projects only start when the implementation deadline is already close, which is a serious project planning issue. To be noted that the organization still behaves in a very re-active way, suggesting a low maturity level. This re-active way of working, without a solid project planning and coherent processes, leads to several issues on the full life cycle of the project, while creating a much stressed working environment. Additionally, with the printed urgency in every project, is common to detect some miss alignment with the initial requirements, with solutions delivered that are not entirely compliant with the end customer expectations.

3. Improve commitment of middle and top management - This is a recurring theme on several projects and, at the same time, one with high impact on performance. In this particular case the lack of commitment from management translates in the teams taking too many “educated guesses”, meaning that the teams proceed working without formal approvals or gate controls. This practice obviously leads to a lot of re-work as well as to applications that are not fully aligned with the objective of the project (sometimes with less functionalities, some other times with more functionalities than what was actually required)

4. Improve responsibility mapping - Although a conscious effort is made to define responsibilities and stakeholders in all areas, cross-boundary activities often reach a "circular" situation - no one takes responsibility for one activity and forwards it to the next in line. This entropy on the organization leads to a massive amount of re-work activities as well as functionalities that are not fully aligned with requirements.

5. Optimization of the collaboration between different teams/parties – In an environment with activities that can be considered as “cross-work streams”, the collaboration should be improved for a higher efficiency. To be noted that the knowledge sharing between teams is not ideal, which translates into a very low performance on some projects. It is also important to highlight that this issue occurs between every team of the project, but is an especially serious problem when the teams are from different vendors.

6. Optimize processes and deliverables (methodology) - The effort to create solid foundations for quality control of projects is explicit but flawed in some cases since the current process is too heavy and not supported by an appropriate single tool. Additionally, the process is not well controlled, meaning that the adherence to the defined methodology is, in some cases, below expectations. In example, the deliverables matrix is currently extensive and all deliverables are mandatory for each project, which in some cases can be seen as NAV activities (Non-Added Value).

It is first interesting to observe that all the interviewees highlighted the processes and deliverables as an opportunity for improvement, which is fully aligned with the author's analysis of the methodology. Although the author focused his analysis on the methodology usage itself, in order to reveal miss alignments with industry best practices, the 3 interviewees focused on processes and deliverables – This will be considered for the next chapter as the same opportunity for improvement, since the pain points that it causes can be considered as being the same.

Furthermore, this exercise clearly suggest that there are still a few points to be improved in order to achieve a higher level of performance and hence an improved competitive advantage in this IT organization. These 6 main opportunities for improvement already point a clear path for improvement and suggest that the organization should get "Leaner" by reviewing some of its processes and practices, aiming to tackle several very clear pain points that could have a direct overall positive impact.

5 Applying a Lean IT concept of Waste - DOWNTIME

As previously presented on this paper, the Lean IT philosophy suggests 8 main sources of waste (DOWNTIME), that should be carefully analyzed and managed for an improved performance. In order to understand its relevance to practitioners, the author clustered with the help of the stakeholders all the opportunities for improvement identified on the previous chapter against the 8 types of waste as shown on Table 3. A deeper analysis of this clustering exercise reveals that the first identified opportunity for improvement, 1. Improving and enhancing development environments, is the source of 4 different types of waste (Defects, Waiting, Transportation and Motion). Further, the second opportunity for improvement, 2. Better management of project planning and expectations, translated directly in 4 different sources of waste (Defects, Overproduction, Non value added processing and Motion). The following 2 opportunities for improvement – 3. Improve commitment of middle and top management and 4. Improve responsibility mapping - were interpreted as the root cause for the same 4 entries of DOWNTIME (Overproduction, Non value added processing, Inventory, Motion). Furthermore, the fifth opportunity for improvement of this case study, 5. Optimization of the collaboration between different teams/parties, has a direct relation with 3 sources of waste (Defects, Motion and Employee Knowledge). Finally, the last opportunity for improvement, 6. Optimize processes and deliverables (methodology), that comprises the findings from the methodology analysis and the findings from the unstructured interviews, revealed to be the source of 3 types of waste (Non value added processing, Motion and Employee knowledge).

Table 3. Table view of Opportunities for improvement against DOWNTIME

Opportunities for Improvement	DOWNTIME	Number of relations
Improving and enhancing development environments	Defects, Waiting, Transportation, Motion	4
Better management of project planning and expectations	Defects, Overproduction, Non-Value added processing, Motion	4
Improve commitment of middle and top management	Overproduction, Non-Value added processing, Inventory, Motion	4
Improve responsibility mapping	Overproduction, Non-Value added processing, Inventory, Motion	4
Optimization of the collaboration between different teams/parties	Defects, Motion, Employee knowledge	3
Optimize processes and deliverables (methodology)	Non-value added processing, Motion, Employee knowledge	3

It is noteworthy that all the opportunities for improvement identified on this case study have a direct relation with at least 3 sources of waste from DOWNTIME. Some of them actually translate into several sources of waste, indicating a large potential for improvement. It is also revealing to understand that all the 8 types of waste suggested by Lean IT literature were present on the analyzed projects, reinforcing the idea that the organization can take several steps forwards in regards to its performance. It is also interesting to note that the source of waste that is more present on the analysis is "Motion", a very typical root cause for low performance in low to medium maturity organizations, since it reflects a very "re-active" way of working rather than pro-active, validating the authors initial comment regarding the organization's maturity level. This direct relation between the identified opportunities for improvement and the waste sources from DOWNTIME suggests that a Lean mind-set would help on improving the overall performance of the organization, considering that one of the most important mind-sets created by a Lean philosophy is to consciously and constantly detect the origins of poor performance or waste – and it is this same mind-set that could improve the performance of IT organizations and departments across the globe, unlocking their full potential.

6 Conclusions

Lean IT is still in a very early stage and the adoption of its principles by practitioners of the IT sector is not yet generalized. However, it is the author's belief that the concepts of Lean will also be widely adopted by Information Technology and the results will be equally promising, as suggested by this paper. It is also necessary to highlight that Lean needs to be seen as a journey [7] and as such, as a mean and not an end. But the mindset can be adopted by most practitioners, especially by the Leadership teams, in order to start with some simple but meaningful changes as a starting point for a holistic Lean Transformation. In fact, it is the constant challenging of the methodologies, methods and techniques used on a daily basis, as well as a constant curiosity on how it could be done better that can unlock the full potential of an organization. This concept of constantly pursuing better performance is actually a fundamental principle of the Lean mind-set, although not easy to achieve since there is no “cookbook” or detailed roadmap for a Lean transformation [7]. The present study starts suggesting that a Lean mind-set can help IT organizations in taking some initial steps in that journey, especially in detecting and correcting waste sources. To be noted that all opportunities for improvement identified on this case study have a direct relation with DOWNTIME wastes, suggesting that an initiative that targets these sources would be able to considerably improve the efficiency and maturity of an organization, ultimately increasing its competitive advantage.

7 Threats to Validity

From the author’s perspective, the goal of this paper of better understanding if the adoption of a philosophy centered in Lean concepts would benefit an IT organization in regards to its performance as well as to validate some of the currently available Lean IT recommendations was achieved, yet with some limitations, especially when considering that this study might have some risks to its validity - from an external validity point of view, data representativeness risks should be considered since other organizations with different maturity levels might provide different findings. On the other hand, from an internal validity point of view, the generalization of the empirical findings poses a risk since other projects in other regions with different conditions might provide other findings. These risks also indicate that further research on the topic, with data collected from different projects in different organizations and regions could greatly benefit the academic community.

8 Opportunities for future research

Finally, the reader should note that there is already a considerable amount of literature on Lean philosophy, but literature on Lean IT topics is still limited on the academic sphere. It would be interesting to invest further efforts in understanding the benefits of a Lean mindset on IT organizations in order to close the current literature gap, especially if those studies focus on methodologies and agile methods, which seem to be a very good starting point for a holistic IT Lean transformation. On another note, to focus on pit falls and main risks of implementing a Lean mindset on IT organizations would also be an interesting path, considering that the literature has not been focused on success factors over the past years. Also, and reinforcing the idea that the major benefit from a Lean transformation point of view is the consolidation of the full processes and life-cycle of products, some further research focused on other areas often related to IT projects or business transformation projects, such as organizational change management or end user training, would add significant value for researchers and practitioners.

References

1. The Standish Group 2015, The CHAOS Report. Available from: The Standish Group International [15 December 2016]
2. Jones, D. 2004, 'Lean Beyond Manufacturing', LeanUk Blog, blog post, 01 January. Available from: http://www.Leanuk.org [31 March 2016]
3. Womack, J, Jones, D & Roos, D 1990, The machine that changed the world: Based on the Massachusetts Institute of Technology 5-million dollar 5-year study on the future of the automobile, Rawson Associates, New York
4. Karlson, C & Ashlstrom, P 1996, 'Assessing changes towards Lean production', International Journal of Operations & Production Management, vol. 16, no. 2, pp. 24 – 41.
5. Liker, J 1996, Becoming Lean: Inside Stories of US Manufacturers, Productivity Press, New York
6. Kleiner, A 2005, Leaning toward utopia. Available from: http://www.strategy-business.com [31 March 2016]
7. Bhasin, S & Burcher, P 2006, 'Lean viewed as a philosophy', Journal of Manufacturing Technology Management, vol. 17, no. 1, pp. 56-72.
8. Soriano-Meier, H & Forrester, P 2002, 'A model for evaluating the degree of Leanness of manufacturing firms', Integrated Manufacturing Systems, vol. 13, no. 2, pp. 104-109.
9. Poppendieck, M & Poppendieck, T 2003, Lean Software Development: An Agile Toolkit, Addison-Wesley, Crawfordsville
10. Waterhouse, P. 2008, Improving IT economics: Thinking 'Lean'. Available from: CA Technologies [31 March 2016]
11. Yang, Y, He, M, Li, M, Wang, Q & Boehm, B 2008, 'Phase distribution of software development effort', International Symposium on Empirical Software Engineering and Measurement, Kaiserslautern, pp. 61 – 69.
12. Boehm, B et al. 1996, Software Cost Estimation with COCOMO II, Prentice Hall, Upper Saddle River

Simulation and path planning for quadcopter obstacle avoidance in indoor environments using the ROS framework

Yadira Quiñonez[1] , Fernando Barrera[2], Ian Bugueño[2], Juan Bekios-Calfa[2],

[1] Universidad Autónoma de Sinaloa, Facultad de Informática Mazatlán,
Av. Universidad y Leonismo Internacional S/N, México
[2] Universidad Católica del Norte, Departamento de Ingeniería de Sistemas y Computación,
Avda. Angamos 0610, Antofagsta, Chile

yadiraqui@uas.edu.mx, fernando.barrera.pincheira@gmail.com, ibv001@alumnos.ucn.cl,
juan.bekios@ucn.cl

Abstract. This work focuses on the analysis of different algorithms dedicated to the planning of trajectories in a quadcopter. The times and distances of the paths from one point to another have been evaluated autonomously, and the evaluation of the reactive algorithms Bug1, Bug2 and DistBug have been considered to carry out the planning of the quadcopter paths. From the experiments and metrics defined, the efficiency and robustness of the algorithms have been determined. To carry out the implementation of the experiments, we have chosen and evaluated an environment based on the Robot Operating System (ROS) and Gazebo development platform.

Keywords: Bug Algorithm, Unmanned Aerial Vehicle, Robotic Software Framework, ROS.

1 Introduction

In recent years, Unmanned Aerial Vehicle (UAV) has been object of study for different researchers in the scientific community in order to develop tools that facilitate the tasks of repetitive, complex or dangerous work. For example, autonomous exploration in urban environments [1], exploration and generation of three-dimensional maps by autonomous devices of underground mines [2], support to rescue teams to avoid losses in case of landslides or when searching in inaccessible places [3], assistance in decision-making when high-risk situations occur [4, 5], among others.

UAV is a vehicle capable of carrying out a mission without needing of a crew, although it does not necessarily exclude an operator on the ground. It is possible to classify UAVs in different ways, one of the most useful is based on their take-off, dividing vertical take-offs, also known as VTOL (Vertical Take-Off and Landing), which is the fastest growing sector [6], as non-vertical take-offs, known as CTOL (Conventional Take-Off and Landing).

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5_27

In this sense, calculating the appropriate movements for a machine to reach a specific point in space is a great challenge, and involves the execution of complex algorithms to achieve this end, the study of the trajectory planning has been widely developed and an endless number of methods are known to give a suitable solution in time and computational cost, which motivates and justifies to perform a comparative analysis specifically for an aerial mobile robot, in order to demonstrate the viability and performance of the algorithms. Some works focus on the bug algorithm using visual topological maps in unmanned ground vehicle and unmanned aerial vehicle [7, 8].

This work focuses on the analysis of different algorithms dedicated to the planning of trajectories in a quadcopter, where the times and distances traveled from one point to another are evaluated autonomously avoiding obstacles to define the effectiveness and efficiency of the algorithms. The analysis has been done through tests in a simulated environment, which seeks to emulate the real behavior of robots in various environments. The simulator also allows not depending on parts or sensors that may imply a cost in time or money for its maintenance. The algorithms have been implemented in the ROS development platform, an open-source robotic operating system. ROS is oriented to the specific needs of robotic platforms using reusable and modular components.

2 Robotic Software Framework

One of the basic principles of Robotic Software Framework (RSF) is to run many programs or processes in parallel that must be able to exchange data synchronously or asynchronously. For example, a RSF needs to consult sensors of the robot at an established frequency (ultrasounds, temperature sensor, gyroscope, accelerometer, cameras, among others), to recover this data, to debug them, and to direct them to the processes that are dedicated to its processing (Speech processing, computer vision, SLAM). Finally, to act on the motors or other electromechanical devices. All this process is performed continuously and in parallel. In addition, the robotic operating system needs to contain an administrator to ensure efficient access to the robot resources. Currently in the RSF there is no standard, although one of the most used for prototyping in the robotic field is ROS.

2.1 Robot Operating System

ROS is an operating system for robots created by the research laboratory *Willow Garage* in collaboration with Stanford University, is an open-source initiative for the development of robotics [9]. ROS has different tools used during programming, simulation or execution of the robot tasks, such as: Stage, Gazebo, Rviz, TF Library, OpenCV, PointCloud Library, among others.

3 Navigation algorithm

Reactive algorithms do not know the environment, their function is simply limited to the robot reaching the destination point, regardless that path found is optimal or not, an advantage of this type of algorithm is that it decreases the cost of complexity. In this sense, Bug algorithms are navigation methods for mobile robot using local planning. The Bug algorithms make three assumptions about the robot. First, the robot is considered a point in space. Second, the robot has the perfect location capability. And third, the robot has perfect sensors. Although these assumptions are not realistic in a real environment, Bug algorithms are considered as the simplest algorithms and the first step towards a solution to the navigation problem. Even often are used as a top-level monitoring element of a system. The most used and referenced in path planning in mobile robots are Bug1, Bug2, and DistBug algorithms.

3.1 Bug1 algorithm

Bug1 algorithm is the oldest algorithm for obstacle avoidance, it is very simple, it uses a minimum of memory and it does not undergo local minimums [10]. The execution procedure is simple, first, the robot moves along a line until it hits the i-th obstacle, defining a point H_i. Second, it begins to follow the edge of the i-th obstacle, in search of the goal in order to achieve the minimum distance. That is, it simultaneously calculates the distance from the current position to the destination, and finally, stores the point that has the minimum distance.

3.2 Bug2 algorithm

Bug2 algorithm is an improved version of Bug1; it is characterized because it can be run at any point on a continuous path. In this sense, the algorithm starts by generating a slope from initial position to the destination; the robot begins to follow it until it is interrupted by an obstacle. Once it is interrupted, it begins to follow the edge of the obstacle and recalculates the new slope from the new position until the new slope becomes equal to the original slope. Finally, when reaching the point that has the same slope as the previous one, it begins to move to the destination following the previous route generated.

3.3 DistBug algorithm

DistBug algorithm is an improved version of the Bug family, based on distance. That is, the robot travels the shortest distance to get better navigation and reach the destination in less time. Different behaviors are used to avoid obstacles, when the robot encounters an obstacle in the road, begins to follow it and simultaneously calculates the distance from current position to the destination.

4 Experiments and results

One of the main problems of quadcopters is flight instability, which leads to loss or damage of device, especially in the first experimental tests. For this reason, the architecture used to carry out the experiments is necessary which covers the realistic flight dynamics, vision and range sensors, besides an easy integration as the robotic middleware ROS, Gazebo has been selected as a simulator since it has a variety of robots, both generic and commercial.

4.1 Simulation environment

The criteria considered for selecting the simulator were as follow:

- **License:** Authorization that the author(s) grant to third to use his system, there are different types of licenses, commercial, free and gratuitous. Those lower-cost licenses are preferred.
- **Multi-platform** Capacity that a project can be used in different operating systems. The fact that a system is multi-platform, gives it an additional quality, which makes it more versatile.
- **Complexity** Indicates the difficulty that presents when using a simulator, either by its installation, programming, or use. The more complex a system becomes, greater number of people who avoid its use, so it receives less interest.
- **Information** Simulator support found in books, web pages, forums, papers, etc. When there is more information (quality), there is a better handling of the subject, which also implies greater ability to solve problems and errors.
- **Simulated Robots** Ability to simulate different types of robots. The availability of a greater number of robots or of those necessary for a given project, avoids the time invested in simulating a specific robot.
- **Supported languages** Programming languages with which it is possible to work in a system. Having a greater number of languages facilitates the task to the programmers.
- **Physical motor** Software capable of performing simulations of certain physical systems such as rigid-body dynamics, the movement of a fluid, etc. The better the physical engine of a simulator, the simulation will be more real and therefore better results can be obtained.

Table 1 presents a summary of the evaluation obtained between the different simulators considered: Gazebo[1], V-REP[2] y MORSE[3]. Each of the criteria scored between 1 and 7, where 1 is the worst score and 7 is the best.

---

[1] gazebosim.org
[2] coppeliarobotics.com
[3] openrobots.org/wiki/morse

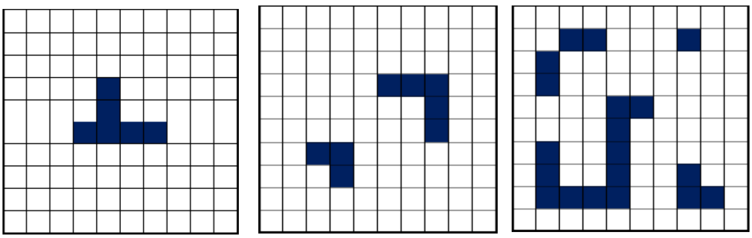
Finally, we have selected Gazebo above V-REP (both obtained the same score) because we have considered that the licensing conditions are marginally better than V-REP.

Table 1. Evaluation of the chosen simulators.

	Gazebo	**V-REP**	**MORSE**
License	3	2	2
Multi-platform	1	3	1
Complexity	3	2	1
Information (access	3	2	1
Simulated robots	3	3	2
Supported languages	2	3	1
Physical motor	2	2	1
TOTAL	**17**	**17**	**9**

4.2 Design of the experiment

The experiment consists of a set of different flight simulations for a quadcopter. This has to go from an initial point to an objective avoiding different obstacles. These simulations were run on three different maps Fig. 1), a number of times with different path planning algorithms. The algorithms selected were the reactive algorithms (Bug Family). Each map is defined as a grid that is divided into squares of 1 x 1 meter, this map format is used to facilitate the positioning of obstacles and the initial position of the quadcopter, in addition, it is the default format of the Gazebo simulator. The execution of the algorithms will have the execution time and the distance covered as metrics.



(a) Map A (b) Map B (c) Map C

Fig. 1. Maps proposed

4.3 Simulations

The experiments were implemented on the ROS platform using the Gazebo simulator and encoded in the Python programming language. As mentioned, in tests six times each of the algorithms per map were executed. The objectives for each map that were defined can be seen in Table 2. In addition, the results obtained for the time metric are shown when executing the algorithms Bug1, Bug2, and DistBug. This metric is based on the **run-time** and it is defined as the time it takes the quadcopter to reach the established goal. For the displacement metric the **distance traveled** is calculated, and is defined as the distance that quadcopter travels until to reach the goal. Table 2 shows the results obtained. It is clear that the **run-time** of Map C is higher because it has more obstacles.

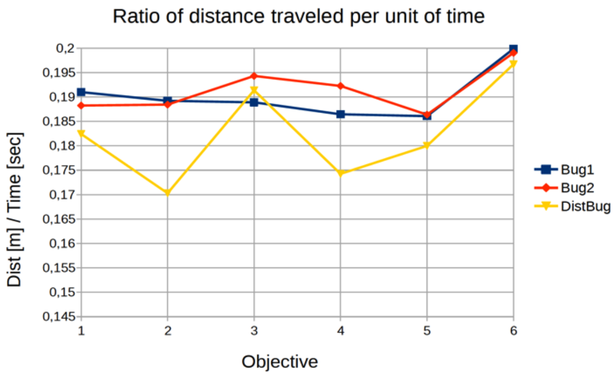
Table 2. Results obtained from the run-time and distance metrics for maps A, B, and C respectively.

Map A	**Position**	**Bug1**		**Bug2**		**DistBug**	
Objectives	**(x, y)**	**Time**	**Distance**	**Time**	**Distance**	**Time**	**Distance**
1	(8.0,8.0)	173,351	33,109	86,476	16,279	75,254	13,728
2	(7.0,9.0)	195,058	36,908	91,783	17,295	88,844	15,129
3	(8.0,6.0)	163,367	30,860	62,166	12,079	58,862	11,264
4	(6.0,6.0)	70,061	13,063	68,601	13,189	73,949	12,886
5	(5.0,9.0)	201,64	37,522	101,031	18,830	92,888	16,718
6	(3.0,7.0)	37,957	7,586	38,663	7,695	38,318	7,537

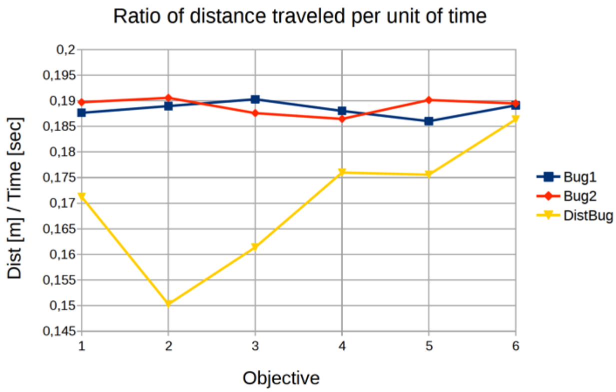
Map B	**Position**	**Bug1**		**Bug2**		**DistBug**	
Objectives	**(x, y)**	**Time**	**Distance**	**Time**	**Distance**	**Time**	**Distance**
1	(3.0,6.0)	140,203	26,307	75,593	14,340	72,119	12,35
2	(6.0,5.0)	124,198	23,468	49,868	9,503	59,003	8,867
3	(9.0,5.0)	139,675	26,578	67,502	12,662	68,869	11,114
4	(9.0,9.0)	278,681	52,394	120,756	22,515	102,781	18,087
5	(5.0,9.0)	305,825	56,885	88,307	16,79	139,122	24,424
6	(3.0,9.0)	158,101	29,897	97,544	18,482	82,72	15,416

Map C	**Position**	**Bug1**		**Bug2**		**DistBug**	
Objectives	**(x, y)**	**Time**	**Distance**	**Time**	**Distance**	**Time**	**Distance**
1	(3.0,7.0)	249,597	45,663	147,849	25,735	102,633	16,583
2	(3.0,3.0)	115,958	20,274	112,000	20,128	117,842	19,939
3	(1.0,9.0)	299,435	53,245	217,269	38,27	137,479	22,336
4	(6.0,8.0)	238,068	43,689	88,150	15,989	86,905	14,313
5	(9.0,9.0)	304,334	55,697	115,274	20,559	111,633	18,929
6	(9.0,5.0)	229,391	41,998	65,219	12,448	81,715	14,221

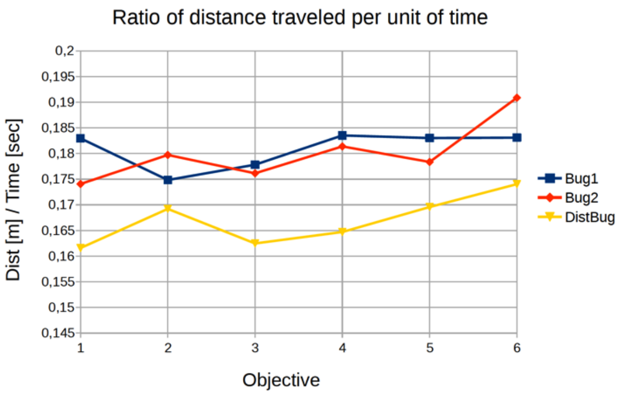
In general, the results are similar for the Bug2 and DistBug algorithms, although, DistBug shows a significant improvement, because its characteristic is to move to the goal once it finds the opportunity to do so. In relation to the Bug1 algorithm the run-time is much greater than the other algorithms because it completely surrounds the obstacle. Fig. 2 presents the results obtained in relation to evaluation of the average speed of displacement (*m/sec*). It can be observed that DistBug is marginally worse than Bug1 and Bug2 algorithms, because it requires a large computational load of both inputs and outputs in each iteration of motion.



(a) Map A



(b) Map B



(c) Map C

Fig. 2. Average speed graphs for maps A, B, and C.

Then, in Fig. 3 and Fig. 4 present the results obtained in Map A and Map C for each of the algorithms, for these simulations in general the results are similar, however, it can be clearly seen that the Bug1 algorithm performs navigation longer than Bug2 and DistBug, because it completely surrounds the obstacle. In relation to the DistBug algorithm, it can be seen that it reaches the goal in less time.

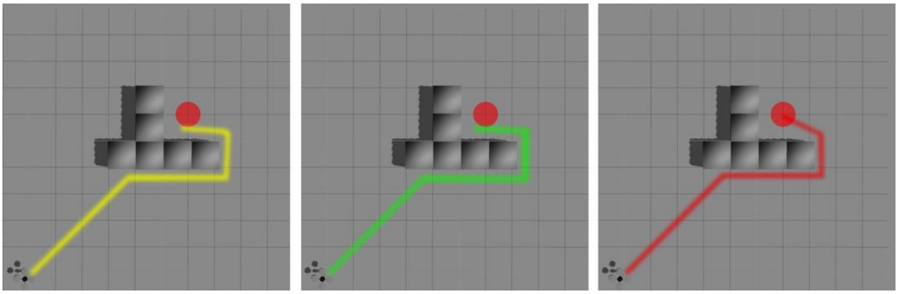


Fig. 3. Comparison of the paths generated by the robot for the Bug1, Bug2 and DistBug algorithms in map A.

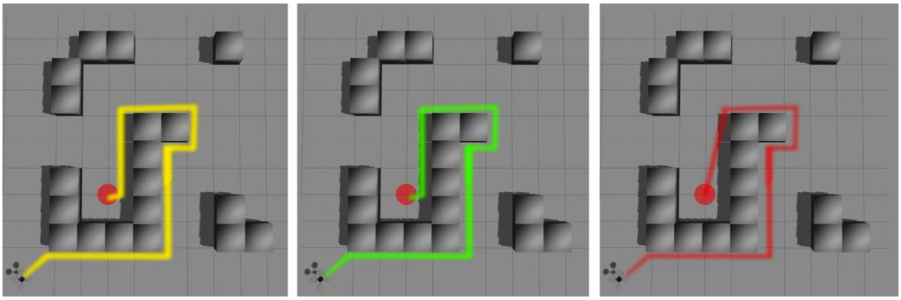


Fig. 4. Comparison of the paths generated by the robot for the Bug1, Bug2 and DistBug algorithms in map C.

Additionally, an experiment has been carried out which includes a modification of the Bug2 algorithm, which in addition to its own performance, the possibility of overflying the obstacle was added. That is, when the robot encounters an obstacle in the path, it rises to find a new point where it is possible to move towards the target. In case it is not possible, the robot begins to follow the edge of the obstacle until to find a new point belonging to the imaginary line. Fig. 5 shows the simulation of quadcopter overflying the obstacle. If the results are compared with the Table 2 it can be noted that for map A and B the improvement is significant in terms of distance traveled (8,596 and 6,548 meters respectively).

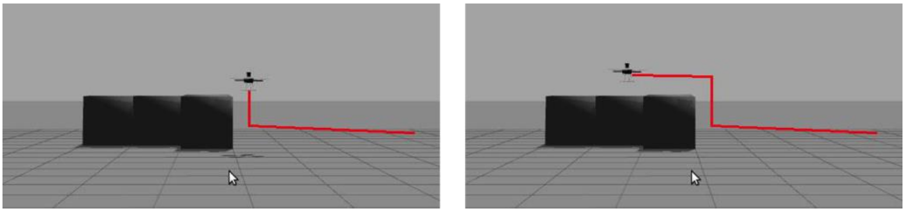


Fig. 5. Simulation of the modified Bug2 algorithm.

On the contrary, in the map C the distance traveled is much greater when the modified Bug2 algorithm is applied. That is, for environments with many obstacles those algorithms tend to fail. Finally, the values obtained in **run-time** are worse than the results obtained in the original algorithm, because the quadcopter tries to evade the wall by making an ascent. This operation has a cost in time especially if the algorithm decides not to continue ascending and return to the original position with the original version of Bug2 algorithm.

Table 3. Results obtained from the run-time and distance metrics for maps A, B, and C respectively.

Map	Goal (x, y)	Distance (m)	Time (s)
A	(5, 9)	10,234	53,690
B	(5, 9)	10,242	54,689
C	(1, 9)	9.309	49,517

5 Conclusions

It has performed an evaluation on three different types of simulators for the development of tests in simulated environments. It has verified that it is possible to perform this type of tests in a simple development environment such as ROS and Gazebo. To achieve this goal, it has executed a set of flight simulations on a quadcopter using different trajectory planning strategies. Three reactive methods of Bug family have been considered as they offer a simple solution to go from an initial point to a target point. In addition, two metrics been defined based on run-time and distance traveled. According to the results obtained, it has been observed that there are obvious differences between the Bug1, Bug2, and DistBug algorithm.

According to the results, Bug1 algorithm obtains the worst results in relation to the run-time and distance traveled metrics, on the other hand, Bug2, and DistBug show similar results. In the case of map C (which has a greater number of obstacles) DistBug has a better performance than the rest. Under these parameters, it can be said that DistBug algorithm performs better than Bug1 and Bug2 algorithms. However, when the average speed of displacement (m/sec) is considered, DistBug is marginally worse than Bug1 and Bug2 algorithms because DistBug is an easy code to implement, but it is difficult to perform for any robot due it requires a large computational load of both inputs and outputs in each iteration of motion. Finally, the modified Bug2 algorithm shows interesting previous results, and although these are inconclusive, it shows an interesting way to explore for the development of new reactive planning algorithms for aerial vehicles.

It would be interesting to study the possibility of varying the speed of the quadcopter and implement simulations with evasion of mobile obstacles, this would imply faster reaction times in both the sampling and the quadcopter. Finally, to contrast the simulations in real environments, since although the simulators are becoming more realistic, they do not consider all the problems that are in a real environment.

References

1. Adler, B., Xiao, J., Zhang, J.: Autonomous exploration of urban environments using unmanned aerial vehicles. Journal of Field Robotics, vol. 31, no. 6, pp. 912–939 (2014)
2. Thrun, S., Thayer, S., Whittaker, W., Baker, C., Burgard, W., Ferguson, D., Hahnel, D., Montemerlo, D., Morris, A., Omohundro, Z., Reverte, C., Whittaker, W.: Autonomous exploration and mapping of abandoned mines. IEEE Robotics & Automation Magazine, vol. 11, no. 4, pp. 79–91 (2004)
3. Kumar, V., Michael, N., Yoshida, K., Naganati, K.: Aerial robots for rapid response: Remote autonomous exploration and mapping. University of Pennsylvania, Philadelphia, USA, (2012)
4. Chen, J. Y.: UAV-guided navigation for ground robot tele-operation in a military reconnaissance environment. Ergonomics, vol. 53, no. 8, pp. 940–950 (2010)
5. Nugroho, G., Satrio, M., Rafsanjani, A. A., Sadew, R. R. T.: Avionic system design Unmanned Aerial Vehicle for disaster area monitoring. In IEEE International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation, pp. 198-20, (2015)
6. Garcia, R., Valavanis, K., Kandel, A.: Autonomous helicopter navigation during a tail rotor failure utilizing fuzzy logic. Mediterranean Conference on Control Automation, pp. 1–6, (2007)
7. Maravall, D., de Lope, J., Fuentes, J.P.: A Vision-Based Dual Anticipatory/Reactive Control Architecture for Indoor Navigation of an Unmanned Aerial Vehicle Using Visual Topological Maps. International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC, vol. 7931, pp. 66-72, (2013)
8. Maravall, D., de Lope, J., Fuentes, J.P.: Visual Bug Algorithm for Simultaneous Robot Homing and Obstacle Avoidance Using Visual Topological Maps in an Unmanned Ground Vehicle. International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC, vol. 9108, pp. 301-310, (2015)
9. The Robot Operating System (ROS), www.ros.org
10. Sezer, V., Gokasan, M.: A Novel Obstacle Avoidance Algorithm: "Follow the Gap Method". Robotics and Autonomous Systems vol. 60(9), pp. 1123-1134 (2012)

Author Index

J. Mejia et al. (eds.), *Trends and Applications in Software Engineering*,
Advances in Intelligent Systems and Computing 688,
https://doi.org/10.1007/978-3-319-69341-5